AN AUTOMATED CLASSIFICATION SYSTEM TO DETERMINE MALIGNANT GRADES OF BRAIN TUMOUR (GLIOMA) IN MAGNETIC RESONANCE IMAGES BASED ON META-TRAINABLE MULTIPLE CLASSIFIER SCHEMES

AHMED NASER HUSSEIN AL-ZURFI

Ph.D. Thesis   **2019**

An Automated Classification System to Determine Malignant Grades of Brain Tumour (Glioma) in Magnetic Resonance Images Based on Meta-Trainable Multiple Classifier Schemes

Ahmed Naser Hussein Al-Zurfi

Thesis Submitted in Partial Fulfilment of the Requirements of the

Degree of Doctor of Philosophy (Ph.D.)

School of Computing, Science and Engineering

University of Salford, Manchester, UK

2019

TABLE OF CONTENTS

V

LIST OF TABLES

IX

X

XI

LIST OF FIGURES

XVIII

XXI

DECLARATION

As the author of this thesis, I hereby confirm that the research work presented in this PhD thesis is original and has not previously been submitted at any other university in support of an application for another degree or qualification.

LIST OF ABBREVIATIONS

2D	Two Dimensions
2DGLCM	Two-Dimensional Grey Level Co-occurrence Matrix
3D	Three Dimensions
3DGLCM	Three-Dimensional Grey Level Co-occurrence Matrix
ADC	Apparent Diffusion Coefficient
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
BNN	Back-Propagation Neural Network
CAD	Computer-aided Detection or Diagnosis
CCR	Correct Classification Rate
CRUO	Cancer Research UK Organization
CSF	Cerebrospinal Fluid
CT	Computerised Tomographic
DINN	Deep Iteration Neural Networks
DNA	Deoxyribonucleic Acid
DNN	Deep Neural Networks
DT	Decision Tree
DTI	Diffusion Tensor Imaging
DWI	Diffusion-weighted Magnetic Resonance Imaging
EBTree	Ensemble Bagged Decision Tree
ECOC	Error Correcting Output Coding
ESDA	Ensemble Subspace Discriminate Analysis
FA	Fractional Anisotropy
FFNN	Feedforward Neural Networks
FLAIR	Fluid-Attenuated Inversion Recovery
FMRI	Functional Magnetic Resonance Imaging
FN	False Negative
FP	False Positive
FTD	Features Associated with Tumour Descriptors
GLCM	Grey Level Co-occurrence Matrix

GLRLM	Grey Level Run Length Matrix
HC	Hierarchical Classification
HFSA	Hybrid Feature Selection Algorithm
HMTMCS	Hierarchical Meta-Trainable Multiple Classifier System
HS	Hierarchical Structure
KNN	K-Nearest Neighbour
KNNCOS	KNN where the number of k-neighbours is 10, distance metric is cosine.
KNNCUB	KNN where the number of k-neighbours is 10, distance metric is cubic.
KNNF	KNN where the number of k-neighbours is 1, distance metric is Euclidean.
KNNM	KNN where the number of k-neighbours is 10, distance metric is Euclidean.
KNNW	KNN where the number of k-neighbours is 10, distance metric is Euclidean, and the distance weight is squared inverse.
LDA	Linear Discriminate Analysis
LOO	Leave-one-out
MCS	Multiple Classifier System
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MTMCS	Meta-trainable Multiple Classifier System
NNs	Neural Networks
Obj	Objective
ODM	Output Decision Matrix
PWI	Perfusion Weighted Imaging
RCBV	Relative Cerebral Blood Vessels
ROI	Region of Interest
SC	Select best Set of Classifiers
SCA	Select best Set of Classifiers Algorithm
SD	Standard Deviation
SVM	Support Vector Machine
SVMCUB	SVM with cubic kernel function.
SVMG	SVM with Gaussian kernel function
SVML	SVM with linear kernel function.
SVMQ	SVM with Quadratic kernel function.

T1c	T1 with contrast enhancement
T1C-W	T1-weighted images with contrast enhancement
T1-W	T1-weighted image
T2-W	T2-weighted image
TD	Tumour Descriptors
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
WHO	World Health Organisation

ACKNOWLEDGMENT

In the name of Allah, the Entirely Merciful, the Especially Merciful.

First and foremost, I would like to express my sincere gratitude and thank ALLAH ALMIGHTY for his generosity towards me, the patience and guidance he provides to complete this research thesis and achieve my dream.

Sincerely, I would like to present my great thanks to **Professor Farid Meziane** for his invaluable guidance, extreme support, and inspiration during this research journey. He is recognised as a kind and supportive person. His open-door policy is always open for me throughout my research journey. Professor Farid Meziane is not only an academic role model for me; but also he gives me the confidence and motivation through my PhD journey. To me, he has been a philosopher, a teacher and a friend. Also, my great thanks must go to **Dr. Rob Aspin** for his guidance, patience, motivation and continuous support with his immense knowledge. He is an intelligent academic professional; his support and guidance were always obtainable to me. This research would not have been possible without the fruitful feedback of both Prof Farid and Dr. Rob generously offered to me. I would also like to take this opportunity to express my sincere thanks to Dr. Andrew England for his assistance and encouragement.

My gratitude must go to **the department of electrical and communication**, Faculty of Engineering, University of Kufa, to offer me the opportunity to complete this PhD research. My gratefulness also goes to The Iraqi government represented by the Ministry of Higher Education and Scientific Research and the **Iraqi cultural attaché in London**, for their continuing support and help during this research journey.

My great and special thanks must go to my family; to the soul of my dearest mother, she prays for me all the time. In my memory, she was praying to me when I had an exam or facing difficulties in my life, she is always caring and supportive of her entire life, may Allah rest her soul in peace. My special appreciation to my lovable father who always prays for me and gives encouragement to progress in my life. He always says, 'Oh son you should work hard, you have to succeed, do not let anything defeat you from winning this PhD journey and come back as soon as you can' and repeats this words to me. To my brothers and sisters, special gratitude especially my dear brother Hadi who handled many issues and helped my family and me during this PhD journey. Additionally, I would like to say thanks to my dear brother Dr. Mahdi who support me and encourage me a lot, 'you should focus in

your PhD, this is your dream, be patient, this PhD is your life now, and the best thing could happen to you in your life' he keeps saying these words to me. My sincere gratitude should go to my sweetheart wife, to my lovely son Mustafa and to my queen's daughters Tuqa, Dhoha, and Fatima. Although it was hard to live alone far away from my family, this motivated me to work harder and give all my possible time to succeed in my PhD journey.

Moreover, I would like to take this opportunity to express my special gratitude for professional people who supported me and encouraged me to finish this research work, names are as follows; Dr. Aminu Abba Yahaya, Dr. Ali Al-Waeli, Asaad F. Qasim, Dr. Santosh More, Sadeq Al-Murshedi, and Ali Mohammed. Last but not the least, my thanks should go to all the faculty and staff of the University of Salford who support me in one way or another through my research journey. Finally, I would like to thank all people who have presented their assistance to me.

DEDICATION

In the memory of my dearest Mother

To my lovable Father

To my sweetheart Wife

To my Brothers and Sisters

To my lovely daughters; Tuqa, Dhoha and Fatima

To my dear son Mustafa

To research societies

XXX

LIST OF PUBLICATIONS

AlZurfi, A., Meziane, F. & Aspin, R. Automated glioma grading based on an efficient ensemble design of a multiple classifier system using deep iteration neural networks matrix. Proceedings of the 24th International Conference on Automation and Computing (ICAC'2018), 2018 Newcastle University, Newcastle upon Tyne, UK. IEEE.

Al-Zurfi, A. N., Meziane, F. & Aspin, R. A Computer-aided Diagnosis System for Glioma Grading using Three Dimensional Texture Analysis and Machine Learning in MRI Brain Tumour. 2019 IEEE 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART), 2019 Paris, France. IEEE.

ABSTRACT

The accurate classification of malignant grades of brain tumours is crucial for therapeutic planning as it impacts on the tumour's prognosis, where the higher the malignancy levels of the brain tumour are, the higher the mortality rate is. It is also essential to provide patients with appropriate clinical management that may prolong survival and improve their quality of life. Determining the malignant grade of a brain tumour is a critical challenge because different malignant grades of brain tumours, in some cases, have inconsistent and mixed morphological characteristics. Consequently, the visual diagnosis using only the naked eye is a very complex and challenging task. The most common type of brain tumour is glioma. According to the World Health Organisation, low-grade glioma, which includes grade I and grade II are the least malignant, slow growing, and respond well to treatment. While, high-grade gliomas, which include grade III and grade IV are extremely malignant, have a poor prognosis and may lead to a high mortality rate. Hence, the motivation to develop an automated classification system to predict the malignant grade of glioma is the aim of this research. To achieve this aim, several novel methods were developed and this includes new methods for the extraction of statistical measures, selection of the dominant predictors, and the fusion of multi-classification models. The integration of these stages generates an accurate and automated decision system to determine the malignant grade of glioma. The feature extraction starts from the viewpoint that the objective measure of the brain tumour descriptors in MR images lead to an accurate classification of malignant brain tumours. This work starts from the standpoint that meta-trainable fusion of multiple classifier models can offer a better classification accuracy to recognise the malignant grade of glioma in MR images. This study developed a novel strategy based on two stages of multiple classifier systems for glioma grades. In the first stage, different machine learning algorithms were used. In the second stage, a systematic trainable combiner was designed based on deep neural networks. This research was validated using four benchmark datasets of MR images, which are publicly available and confirmed with the histopathological diagnosis. The proposed system was also evaluated and compared against different traditional algorithms; the experimental results showed that the proposed system has successfully achieved better and optimal discrimination in glioma grades on all dataset.

CHAPTER 1 : Introduction and Motivation

Overview

This chapter introduces the research developed in this thesis and the background to the traditional visual diagnosis of malignant brain tumours. It is used to justify the need to develop an automated system to classify glioma grades using medical imaging techniques as this will offer a significant help to the clinician during the decision-making process to achieve more objective and accurate diagnosis of the different glioma grades. Furthermore, this chapter elaborates on the aim and objectives of the study, as well as introducing the scope, contributions and direction of the research and exploring the main challenges considered in the study. In addition, it provides an introduction to the methods and tools used to overcome the challenges throughout the development of this study.

1.1 Introduction

The number of newly diagnosed cases of brain tumours is increasing all over the world annually. The incidence rate of brain tumours as presented by the Cancer Research UK Organisation (CRUO) has shown that the number of new cases recorded in 2013 was about 10,624 in the UK, while in Europe the statistics indicated that around 57,100 new cases were reported in 2012, and worldwide more than 256,000 cases. Similarly, CRUO stated that since the late 1970s, the incidence rates of malignant tumours have increased by almost two-fifths (39%) in the UK. Inevitably, brain tumours can grow, become more aggressive and lead to mortality. For instance, the number of deaths caused by brain cancer due to malignant tumours in 2012 was 5,187 cases, and the survival rate was about 14% for ten years or more in the UK[1].

Glioma is the most common type of malignant brain tumour and can be classified into four malignant grades according to the World Health Organisation (WHO). Low-grade gliomas, which include grade I and grade II, grow very slowly with a remarkably better prognosis (Weller, 2011). The best therapeutic process offers an extensive resection of the lesion and delays adjuvant radiotherapy postoperatively until the progression of the tumour is observed

---

[1] http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/brain-other-cns-and-intracranial-tumours#heading-Zero

again (Pouratian and Schiff, 2010). The high-grade gliomas, which include grade III and grade IV, are managed with an essential resection and chemotherapy and radiotherapy (Stupp et al., 2010). The classification of malignant grades of glioma allows the determination of the management and prognosis for the patient (Hadziahmetovic et al., 2011). Therefore, it is crucial to differentiate low-grade gliomas from high-grade gliomas preoperatively, as this effectively impacts the prognosis and treatment of a patient's health (Theeler and Groves, 2011, Siker et al., 2006).

The traditional diagnosis of malignant brain tumours such as glioma relies on the visual assessment of the various attributes of medical images. However, making a correct decision of the malignant grade needs relatively high experience in the neuroradiology field. Furthermore, the inconsistency of different visual characteristics of malignant brain tumours leads to more subjectivity in the diagnosis. This process also may need to test the pathological information of tumour tissues derived from a biopsy or clinical surgery to yield further confirmation of the diagnosis, which is an invasive procedure and the patients may suffer many complications during clinical surgery.

The gold standard approach to determine the correct grade of malignant brain glioma tumour is the histopathological examination. This is achieved by the surgical conditions of the specimen tissue being examined with a microscope. This process is known as the biopsy test. Although biopsy is less invasive and better tolerated by patients, it is still an invasive practice and can cause excessive pain. Nonetheless, brain injury may take place due to the removal of brain tissue. Moreover, eliminating any healthy tissue may affect the normal functioning of the brain. Hence, this approach is considered a costly surgical procedure and can lead to many complications (Lasocki et al., 2015).

Alternatively, computer-aided diagnosis or detection (CAD) has been presented as an efficient and reproducible approach to improve upon radiological diagnosis or detection performance (Herlidou et al., 1999, Kassner and Thornhill, 2010, El-Dahshan et al., 2014, Saad et al., 2015). Computer-aided detection (CAD) has been developed and used successfully to improve the detection performance of experts in the medical field, for example the research carried out by Helbren et al. (2015), in which the effect of CAD on observer's performance has been investigated within clinical environment through a novel methodology based on eye-tracking observers of computerised tomographic colonography. They have found that CAD is able to support observers for fast detection of region of interest and increase of correct polyp identification for both experienced and inexperienced

observers. The usefulness of computer-aided diagnosis (CAD) was also studied in the assessment of malignancy of brain tumours. For example, the research conducted by Hsieh et al. (2017c), in which they have developed a computer-aided diagnosis system for glioma grading to discriminate between grade IV and the lowest glioma grades (II, III). They examined the effect of CAD within the clinical environment on the diagnostic performance of a group of three radiologists that have different experiences. A comparative analysis was conducted between diagnosis with and without the use of CAD. They stated that the diagnostic performance of the radiologists for glioma grading in term of classification accuracy has improved from 72% to 81% when using CAD compared to when performing the glioma grading by only the radiologists.

The malignant grade of a tumour can be estimated by the analysis of several descriptors of a brain tumour. For instance, tumour heterogeneity derived from medical images is a significant indicator for the growth of malignancy of a tumour. Subsequently, it has been shown that image texture analysis reflects tumour heterogeneity (Ryu et al., 2014). Thus, the use of texture analysis as a quantitative measurement of lesion surface patterns is a significant approach to measure the tumour heterogeneity. Texture analysis has been used for the identification and recognition of morphological characteristics of brain tumours (Nielsen et al., 2008, Holli et al., 2010, Bauer et al., 2013, Roy et al., 2013). Medical imaging is a robust technology used for creating visual images of the internal organs or tissues of the human body. Image processing methods have been employed for medical systems and applications and have been used to support clinicians for faster and efficient diagnosis and manage proper treatment within different medical fields. These techniques have an impact on enhancing the diagnostic performance of medical images and provide clinicians with a more objective and efficient second opinion. The accuracy of brain tumour diagnosis has the potential to further improve by automating the classification procedure based on objective analysis to the characteristics of the medical image of a brain tumour. The classification of glioma grades is critical due to different levels of malignancy of brain tumours associated with different treatment and prognosis strategies and is associated with a high mortality rate. Consequently, this creates the enticement to this study to develop an automated classification system of malignant grades glioma, towards objective and accurate classification of glioma grades.

1.2 Background to Brain Tumour

In this section, a brief introduction to brain tumours, definition and diagnosis are presented. It also introduces the definition of malignant brain tumours such as glioma.

1.2.1 Brain Tumour

A typical mechanism of cell division in body tissues is the generation of new cells to take the place of the ones that died or have been damaged. When this process goes wrong for any reason, new growth of undesirable cells is generated. This growth of extra cells often forms a mass of tissue called a lesion or a tumour. Brain tumours are relatively less common than another lesions, such as those of the breast and lung cancers, but are considered extremely significant because the brain is the most vital human part which controls all the functional activities of the body, and malignant brain tumour is associated with a high mortality rate (Tonarelli, 2013). Due to the aggressive nature of malignant brain lesion that causes uncontrolled mass growth that eventually leads to dangerous complications such as pressure to the critical structure inside the brain. Therefore, a malignant brain tumour is a life-threatening condition.

1.2.2 Glioma

Glioma is the most common type of primary brain tumours in adults (Bauer et al., 2013). Therefore, in this research work, the focus is given to glioma. The malignant brain tumours can be subdivided into high-grade tumours (grades III and IV) and low-grade tumours (grades I and II) according to the biological behaviour of brain lesion. They can also be divided into three types depending on the cellular origin type: astrocytomas, oligoastrocytomas, and oligodendrogliomas (Behin et al., 2003). Low-grade gliomas grow slowly and have a good prognosis. While high-grade gliomas have a penchant for invading surrounding tissues, highly-vascular tumours and extensive areas of necrosis are mostly present. High-grade glioma mainly generates a breakdown to the blood-brain barrier in the vicinity of the tumours. Different glioma grades can be visualised using Magnetic Resonance (MR) images (Figure 1.1) to assess the presence of various attributes of brain lesions that assist in characterising different brain neoplasm descriptors.

1.2.3 Medical Imagining Diagnosis

The traditional visual diagnosis starts with providing morphological information about the presence of a tumour using imaging techniques such as magnetic resonance imaging (MRI)

and Computerised Tomographic (CT), which offer plenty of information regarding the malignant growth of a tumour. CT is useful for identifying acute haemorrhage, calcification, and skull lesions. MRI is the most common technique, which can better detect the soft structure of the lesions. Single Photon Emission Tomography and Positron Emission Tomography are generally used for postoperative functions such as distinguishing tumour recurrence from necrosis (Hutter et al., 2003). The final step of the diagnostic procedure, which may apply for subspecies cases, is by using a biopsy, which includes obtaining a tissue sample that belongs to the suspected tumour under surgical intervention and sending it to be examined through the histopathological procedure. Eventually, the final assessment of tissue sample cells is drawn by histopathologist's conclusion (DeAngelis, 2001).

| A | B | C |
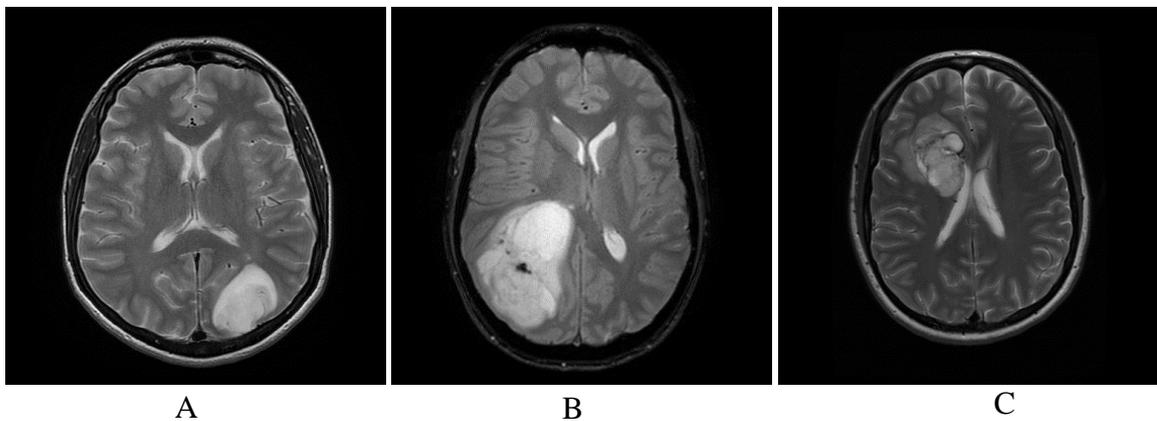
Figure 1.1 Samples of Axial T2-weighted magnetic resonance imaging showing different glioma grades, where A, B and C represent glioma grade II, III, and IV respectively. The hyper-intense regions seen in all these MR slices indicate the dominant unusual growth mass in the brain. In image A, the presence of tumour seems to be of a small size homogenies lesion in the left side of the brain. Image B shows a tumour in the right side of the brain with a bigger tumour in size and seems to have irregular heterogeneous tumour. Image C indicates a tumour in the right side of the brain, which seems to have irregular mass and high heterogeneous lesion.

1.3   The Significance of the Study

The accurate classification of malignant grades of gliomas is crucial to provide patients with the appropriate clinical management that may prolong survival and improve their quality of life (Chao et al., 2006). Developing an automated system to classify glioma grades will offer an objective and accurate decision-making process to determine the malignant level of a brain tumour, which will support a clinician to improve the throughput of diagnosis of

malignant grads of brain tumours (Hsieh et al., 2017c). The developed system will also offer a second opinion for the preoperative diagnosis of glioma grades. This will lead to better prognosis and to manage the most proper treatment for a patient who develops malignant brain cancer.

1.4   Problem Statement and Challenges

Determining the malignant grade of a brain tumour is a significant challenge because different malignant grades of brain tumours, in some cases, have inconsistent and mixed morphological characteristics. Consequently, the visual diagnosis using only the naked eye is subjective due to inter and intra-observer variability (Saad et al., 2015). There is an increased evolution in the incidence rate of a malignant brain tumour over the world, and this raises the challenges in the medical health sector as they are associated with high mortality rate. A patient who develops brain cancer requires an in-depth clinical diagnosis and assesses many health conditions and factors about tumour behaviour to determine the most appropriate treatment that can effectively cure brain cancer. The growth of a brain tumour to a higher level of malignancy is threatening to human life. Incorrect diagnosis of the malignant grade of a glioma leads to inappropriate treatment and raises the risk of unsuitable extraneous treatment. The tradition diagnosis of a brain tumour is accomplished based on visual diagnosis to assess the visual appearances of tumour descriptors that may appear in the medical image of a brain tumour such as the presence of contrast enhancement and necrosis. This examination can also be extended to include a biopsy under the clinical surgery to confirm the diagnosis and determine the accurate malignant level of a brain tumour in cases where there is a probability for a high malignant tumour. However, the biopsy is an invasive approach and the visual diagnosis of a brain tumour depends mainly on the skills, experience and qualification of the expert.

1.5   Research Aim and Objectives

The aim of this study is to develop an automated classification system, which will improve objective discrimination among different malignant grades of glioma, describing it in terms of WHO standardised clinical grading schemas based on objective and recognised predictors extracted from MR images.

The developed system will offer a great assistant to clinicians towards an accurate and objective decision to determine the malignant grade of glioma (Hsieh et al., 2017c). This

will lead to better prognosis and manage proper treatment for a patient who develops brain cancer (Chao et al., 2006).

This work will review and explain how automated methods could be defined to generate fast, accurate and objective assessment for glioma grades using statistical features measured from medical images.

The above aim will be demonstrated by achieving the following research objectives:

Obj.1: Review the literature to understand the problem domain and to identify the research requirements, opportunities and boundary of the research undertaken. Also, it will evaluate and investigate the appropriate methods and techniques, which support the development of automated classification system for glioma grades. The developed system will achieve specific goals of non-invasive (without clinical surgery), automated and objective analysis.

Obj.2: Design a new method that supports the automated classification for glioma grades based on developing the following stages; the first stage is extracting efficient features and the second stage is selecting the most significant feature. The developed methods will contribute to improve the quality of the classification process.

Obj.3: Design a new method within the classification stage, which can support the automated classification system to achieve a better discrimination for glioma grades.

Obj.4: Evaluate the new method experimentally for improving the classification accuracy by measuring the performance of the automated system using common quantitative technique such as the confusion matrix. This seeks to determine that the new method aligns to the recent state-of-the-art.

To achieve the aim and the objectives of this study, it necessary to address the following research questions.

1- Can the objective measures of the brain tumour descriptors incorporated with automated methods enable better grades discrimination in gliomas?

2- How to select the most efficient features to classify glioma grades more correctly?

3- How to develop a new classification method that can achieve further accuracy in the classification of glioma grades?

4- How to improve the classification accuracy for glioma grades?

1.6   Contributions and Novelty of the Study

The prediction of malignant grades of a brain tumour is traditionally performed based on a visual diagnosis of MRI findings in addition to demonstrating a biopsy test to gain the full confirmation of the diagnosis. However, the visual diagnosis is a complex task, and the biopsy is an invasive approach that may harm the patient through a clinical surgical procedure. Many existing works used the combination of different advanced MRI imaging techniques to achieve further improvement in the classification accuracy for glioma grades. However, the advanced MRI techniques are costlier and have limited availability in MRI clinical centres. Therefore, this study is conducted based on conventional MRI techniques, which are readily available in any MRI clinical centre.

The key contribution of this study is the automated classification system and the methodology which is undertaken. In addition to employing several dominant predictors of a brain tumour incorporated with the ensemble of different effective machine learning algorithms, which can improve the accuracy of distinguishing various WHO glioma grades. This will offer support to the clinicians towards an accurate, objective and automated decision for glioma grading.

Four benchmark datasets were used to evaluate the proposed framework. These datasets are publicly available online for academic use and are pre-diagnosed with the confirmation of the histopathological test.

The main contributions and novelties of this study are summarised as follows:

1- New method to extract features from MR images of brain tumours is proposed, based on generating ratio predictors and objective analysis extracted from the presence of different descriptors of a brain tumour, such as contrast enhancement, non-enhancement, necrosis, and edema, which offer an objective analysis of MRI attributes of a brain tumour. It is noted that the existing work relies on the expert domain to analyse these tumour descriptors but this has limitations of inter and intera variabilities in the diagnosis. Other existing studies do not take into account the objective analysis of the relations of these features either with each other or the influence of these relations if integrated with machine learning algorithms on the classification of glioma grades. The proposed method is beneficial as the discrimination ability of these tumour descriptors can be analysed and built to differentiate malignant gliomas. Consequently, more impact from these tumour

descriptors can be gained to improve the classification accuracy of glioma grades. The main advantage of these features is that they can achieve a solid conclusion for the suspected growth of malignant grades of glioma. This is the first study that investigates the impact of these features on glioma grading.

2- A new feature selection method is developed that eliminates redundant features, not only related to maintaining the same level of accuracy but also achieving further improvement in the classification accuracy for glioma grades. This method is based on taking advantage of a fusion between filter and wrapper methods. It is based on the correlation analysis incorporated with several classifiers to update and guide the selection process.

3- A comprehensive analysis of three-dimensional textures feature based on the Grey Level Co-occurrence Matrix (GLCM) is established to support the development of an automated MRI classification system of glioma grades.

4- A novel method is proposed to support the ensemble of different machine learning algorithms that further improve the classification accuracy for glioma grades. This includes the development of the meta-trainable strategy based on deep neural networks (DNN). The existing works mainly concentrate on either using single machine learning algorithm or using one stage of multiple classifier systems. Therefore, better classification accuracy can be achieved using a multiple classifier systems (MCS) based on two stages of learning for glioma grading.

5- A novel method is proposed to optimise the output accuracy of DNN in an effort to provide a systematic trainable design for the MCS that can be beneficial to improving the classification of glioma grades. The existing work applied a few trials, which are randomly selected in attempt to achieve the best design and parameters of the neural networks (NNs). While applying a systematic approach in developing the DNN can play an important role in the optimisation of MCS and thus improve the classification accuracy for glioma grades.

1.7   Scope of the Research Work

This work is concerned with the automating the classification of the malignant grades of gliomas within the medical images processing. This work concentrates specifically on determining different malignant grades of a brain tumour, starting from the point at which segmented images are delivered to the classification system through to the final decision within a valid public standard segmented dataset. Localisation and segmentation of a brain

tumour from medical images are significant tasks, and they are intensively pursued by others (Roy et al., 2013, El-Dahshan et al., 2014, Menze et al., 2015, Al-Waeli, 2017). Consequently, these two tasks are out of the scope of this research.

There are different types and subtypes of brain tumours. This work concentrates on gliomas since they are the most common type of brain tumours. The traditional method of the diagnosis of glioma grades relies on a visual diagnosis of a brain tumour image, which is a complex task, and could be extended to be an invasive approach through clinical surgery. The current work concentrates on a non-invasive methodology and image processing analysis based on extracting objective predictors from brain tumour images to achieve an objective classification of glioma grades.

Within the medical imaging techniques, brain tumour characterisation can be demonstrated based on different imaging techniques such as CT and MRI. This work focuses on MRI to extract the recognised predictors that support the development of an automated classification system for glioma grades. The reasons behind using MRI techniques are as follows. Firstly, MRI is a safer acquisition technique for brain tumour image and is preferable because it has no radiation that harms the patient's body. Secondly, the soft tissue details are also more explicit in this technique. Ultimately, this technique has several modalities that can be used to show different representations of a brain tumour.

Within the tumour descriptors used for the characterisation of the brain tumour image, several descriptors can be extracted from MRI images to reflect different malignant grades of glioma. This study concentrates on the most common and dominant brain tumour descriptors, namely, tumour heterogeneity, contrast enhancement, non-enhancement, necrosis, and edema.

Within the assessment of heterogeneity of a brain tumour, this work focuses on the texture feature to measure the heterogeneity of a brain tumour, because the texture driven from medical images is the most common efficient and objective method to reflect tumour heterogeneity. Other image features such as shape and colour are outside the scope of this work because no significant shapes or colours can be recognised from medical images of glioma grades.

Within the processing of machine learning algorithms, both unsupervised and supervised learning can be used to classify glioma grades. Unsupervised learning is relatively faster than the other learning scheme because there is no requirement to perform a training phase

and mainly relying on statistical measurements to perform the classification. On the other hand, very few works used unsupervised learning because it produces less classification accuracy compared to the supervised learning. However, in critical medical fields, such as the current work, the high priority is given to achieving a better accuracy. Therefore, the work concentrates on the use of supervised machine-learning algorithms, which are widely used for classifying glioma grades.

Within the machine learning, many recent works utilised deep learning approach in several applications. However, to achieve the high accuracy, it is essential to support the training phase of deep learning with huge sample size of datasets and large computation time needed thereby leading to complexities in the design (Papernot et al., 2016). In contrast to the dataset availability in the field of this work, indeed at present, it is a significant challenge to acquire a large image dataset of glioma grades confirmed with the histopathology test. This confirmation is essential for the validation process to any classification or grading system. There is a lack of resources within the boundary of the research environment such as a lack of advanced parallel computing machine. Accordingly, deep learning was not considered in the current research. Instead, the work developed in this thesis is dedicated to developing two stages of learning in a multiple classifier systems, which can address the challenge mentioned above with deep learning.

This development will include the improvement of several techniques of image processing attainment to the targets of this study. To this end, it is necessary to apply objective measurements and combine several recognised features that are used to reflect the malignancy level of a brain tumour. In addition, an objective analysis is required to achieve a reproducible and repeatable experimental framework.

1.8   Research Methodology

The research method utilised in this work is predominantly formative being concerned with the definition of concepts, methods and framework. The research plan within this thesis reflects common approaches in the literature review in the domain of achieving the research aim and objectives. The research methodology of this work has adopted the combination of quantitative and empirical strategies (Kothari, 2004).

This works will pursue to address the ambiguity in the classification of glioma grades by developing methods and techniques that are able to avoid the subjectivity in the determination of the malignant grade of brain tumours. Several image datasets and different

statistical quantitative criteria were used to evaluate the performance of the proposed framework quantitatively and to compare it with other existing algorithms. Accordingly, the approach adopted in this research is in line with the experimental and quantitative research approaches. Consequently, the steps followed during the research work were listed below:

1- Review in depth the state-of-the-art literature to identify the strengths and limitations in existing research works to draw the opportunities and the boundary of the research.

2- Propose, develop and implement a solution that aims to overcome the research problems, and challenges to achieve the aim and objectives of the research.

3- Experimental evaluations using different datasets based on quantitative statistical criteria and compare the performance against existing algorithms.

4- Analysis the results of the experimental implementation and then draw a conclusion and identify the future research trends from the research findings.

1.9 Work Plan and Requirements of the Research Work

This research work within this thesis followed three general phases to explore the research domain, requirements, and research opportunities, which lead to achieve the aim and objectives of this study (Figure 1.2). These phases are detailed below.

Phase 1: Understanding the Challenges and Research Requirements

Within this phase, the scope of the research domain and opportunities are identified. To acquire a deep understanding of the research domain, it is necessary to accomplish a critical review of previous relevant works, considering their strengths and weaknesses. This critical review, consequently, identifies the research requirements to achieve the aim of the research. Understanding a research requirement is essential to address challenges in research. Hence, such understanding enables the identification of the research directions to overcome these limitations, resulting in more opportunities and better improvements in the development process of the proposed solution.

Phase 2: System Design and Development

The traditional process to classify malignant grades of a brain tumour is determined based on the visual diagnosis including the examination of different clinical information, such as the presence of the contrast enhancement, the presence of necrosis, and patient age. The diagnosis process could be extended to the histopathological examination under clinical surgery (biopsy). However, the visual diagnosis is subjective and the biopsy is an invasive process. Therefore, this research follows an alternative approach, which can be determined

by designing a decision maker that is able to achieve an objective, automated, and reproducible classification for the malignant grade of glioma.

Understanding Challenges and Research Requirements

Obj.1

Review Related Work and Backgrounds

Identify Strengths and Limitations of Existing Works

Identify Research Requirements

Data Collection

System Design and Development

Feature Extraction and Selection

Obj.2

Feature Classification

Obj.3

Implementation and Evaluation

Obj.4

Iterative Development
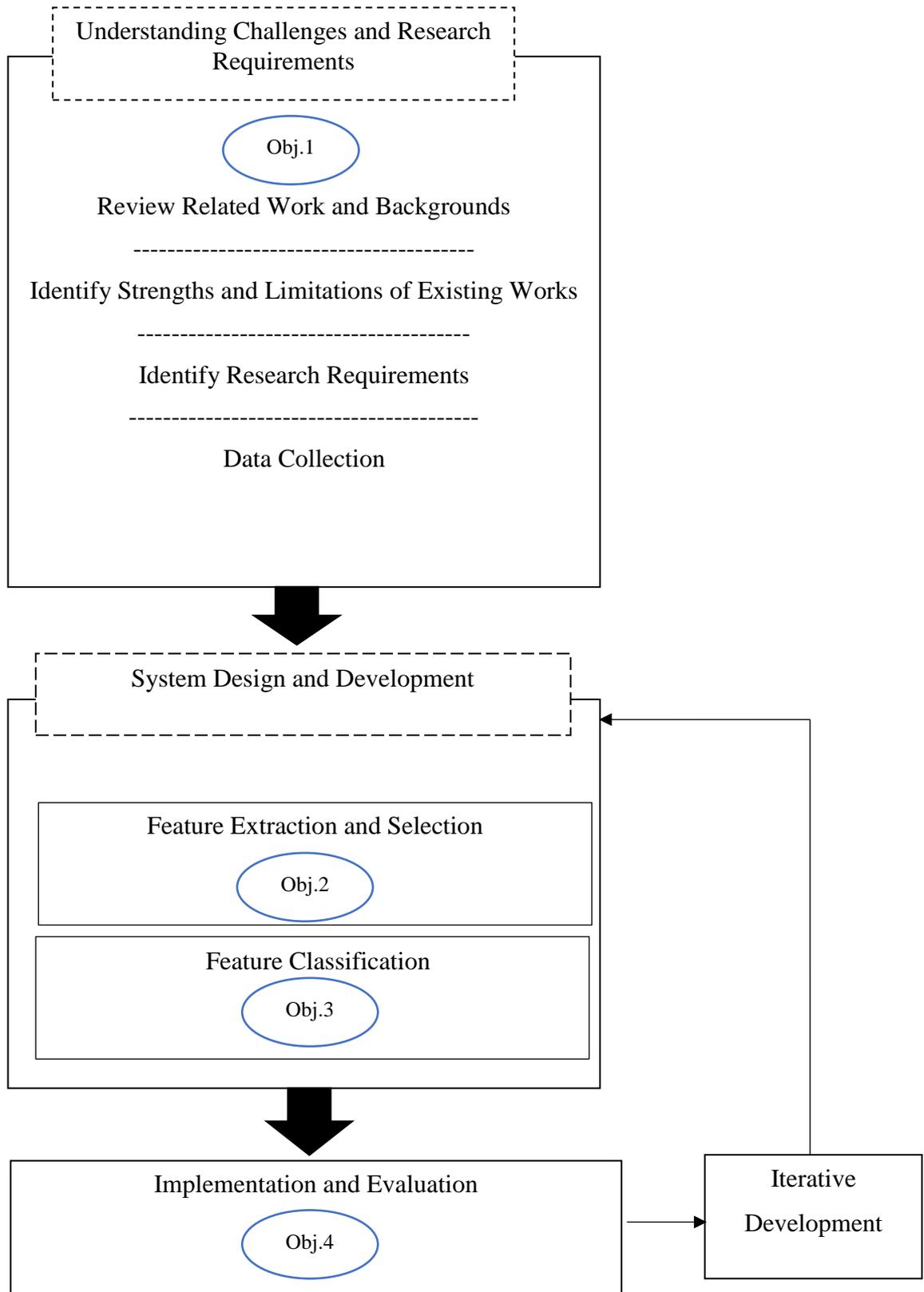
Figure 1.2 Research plan to approach the proposed solution of this study

This solution will also provide significant aid to the clinician in accomplishing the diagnostic procedure for malignant brain tumour. Within the medical image processing community, the development of a classification system requires the following steps.

Step A: Data Collection

Dataset collection is an essential requirement for the development of the classification system; this is necessary to evaluate and validate the proposed system. In the current investigation, four benchmark datasets were used to evaluate the proposed approach. These datasets are anonymous to the user. The use of such public datasets is valuable in the development of the classification system for a reliable and reproducible methodology. To verify the results provided by the classification process, it is required that the acquired dataset has to be diagnosed in advance by the solid confirmation from a histopathological test or a clinical surgery.

 Step B: Identification of Tumour Descriptors

Identification of significant descriptors of brain tumour that are relevant to the research problem is essential and fundamental to support the development of an automated classification system for a brain tumour. Specifically, in this research work, several significant and common descriptors of a malignant brain tumour are driven from MR slices are considered.

Step C: Feature Extraction

There are two approaches to accomplish this step. The first approach is by using the features provided by a qualified expert who uses their experience to identify and track the behaviour of tumour descriptors in medical images. However, this approach is subjective and has the drawbacks of intra and inter-observer variations (Fujita et al., 2008). The second approach is achieved based on using statistical measure correlated with the tumour descriptors, which are driven from medical images of a brain tumour. It is necessary to identify an objective measure for each of these tumour descriptors to avoid the subjectivity in the assessment of malignant brain tumours. Accordingly, to follow a quantitative research methodology and achieve the aim of this study, the second approach is more objective than the first one, therefore, in this work; an objective measure for each of the tumour descriptors is considered and developed.

Step D: Feature Selection

Feature selection is an important task as it selects the most significant features set to enhance the performance of the classification system and eliminate redundant features. In this work, a robust features selection and reduction method will be developed.

Step E: Feature Classification

The final step in the classification system is to classify an unknown sample to one of the classes. Many approaches are used to achieve this step. The visual procedure is the fundamental approach, which is based on using the morphological appearance of many tumour descriptors such as the presence of contrast enhancement to determine the malignant level of brain tumour. The second approach is analysing the features extracted from MR images of a brain tumour and is based on discovering a threshold value from these features to discriminate different malignant grades of brain tumours. The third approach is by taking the merit of applying machine learning algorithm using the features extracted from medical images to establish a model through training and testing of the extracted features. The established model is able to make a prediction on an unknown sample and then classify it to one of the tumour classes. The work within this thesis pursues the third approach to avoid subjectivity, and invasive choices inherent by the first approach, as well as, to produce a more general solution than the one provided by the second approach. Consequently, the present investigation will use machine learning algorithms incorporating active features driven objectively from MR brain tumour images, which would lead to better and objective classification accuracy for glioma grades.

Phase 3: Implementation and Evaluation

The implementation of the proposed solution is the complementary step for the research work. It is essential to implement the proposed system design in order to ensure the quality of the research plan, evaluate and compare the proposed algorithm experimentally against other existing approaches. The implementation is achieved by applying all the processes and steps that are developed in the system design. It is crucial to analyse system performance and the produced results for evaluating the proposed classification system. For the implementation platform, MATLAB R2018 software under academic license of the University of Salford was used to develop the code, functions and simulations to implement the system design. To ensure the generalisation of the classification performance, it is vital to apply the cross-validation method. Several objective metrics are used to evaluate and

validate the classification performance, which are measured using the resultant confusion matrix, for examples, classification accuracy, sensitivity, specificity and precision.

Iterative development and evolution of the methods within a framework is used to enable the developed system in achieving the aim of this research. Evaluating the performance of the developed system is required at each stage to guarantee that the impact of the evolution is entirely valid.

1.10 Summary

This research is dedicated to developing an automated classification system to determine the accurate grades of glioma. The research aim has been defined and explored in terms of the objectives to address the research questions of this study. This chapter also presents the opportunity for the main contributions and novelty in the context of the study as well as indicates the significance of the study. This work seeks to define issues related to the traditional diagnosis of malignant grades of glioma. It also defined the challenges associated with developing a classification system based on machine learning algorithms. The research starts from the standpoint that the traditional visual diagnosis to determine an accurate glioma grade is a subjective, complex task, and time-consuming. An objective analysis incorporating an effective machine-learning algorithm can overcome the limitation in the traditional diagnosis and will enable clinicians towards an accurate and objective decision to determine the malignant grade of glioma. The second standpoint is that each single machine learning algorithm has a shortcoming that tends to reduce the classification accuracy in some cases, as no single classifier is suitable for every dataset, while the fusion of multiple classification models is promising to overcome this limitation and offer further enhancement in the classification accuracy of glioma grades.

1.11 Thesis Structure

This thesis is organised into the following chapters:

Chapter 1:   This chapter includes an introduction to the entire research work, defining the challenges, aim and objectives; elaborating on the research methodology, scope, contributions, novelty, and the implications of this research work.

Chapter 2:   This chapter presents a detailed background related to the medical methods and concepts undertaken in this research work. This chapter also concentrates on defining the state-of-the-art (current approaches) and understanding them within

the problem domain and the approaches were taken to address it. The outcomes identify the research boundaries and opportunities for the novelty in this research with a review of the success criteria against an evaluation framework.

Chapter 3: This chapter provides details of the datasets utilised in this work and the preparation of region of interest-driven from the MR images datasets. It also includes the extraction of texture features with 2D (Two Dimensions) and 3D (Three Dimensions) analysis of grey-level co-occurrence matrix. This also covers the features selection method, which integrates the filter and wrapper approaches. In this integrated method, the merit of using several machine learning algorithms was taken into account. The chapter also covers results evaluation and discussion to apply this method incorporated with the single and ensemble classification method.

Chapter 4: This chapter presents the development of the classification system for glioma grades based on the objective measure of different tumour descriptors including necrosis, edema, non-enhancement, and enhancement. It also covers the objective assessment of these tumour descriptors to discover the discrimination ability of their features to classify glioma grades.

Chapter 5: This chapter demonstrates the development of multiple classifier systems based on two stages of learning, to determine the malignant grades of glioma towards further improving the classification accuracy of glioma grades. It also includes results analysis and comparison with other classification algorithms.

Chapter 6: This chapter presents the development of the multi-class classification of WHO glioma grades. It also provides evaluation and results analysis of the developed system. The proposed method was developed to enhance the classification accuracy of different WHO glioma grades.

Chapter 7: This chapter describes the overall discussion, evaluation, conclusion and future works. It further discusses and compares the overall results obtained by the methods developed and approaches taken and built a comparison against the recently identified developments in the field. It ends with conclusion for the study and suggested trends for future research.

CHAPTER 2 : Background and Literature Review

Overview

This chapter presents a background to brain tumours with an emphasis on glioma. Different characteristics of glioma grades, extracted from MR images, are also explored. Furthermore, conventional and advanced MRI techniques, which are involved in the assessment of malignant brain tumour, are discussed.

This chapter besides presenting a comprehensive survey of existing research works. It also defines the theoretical backgrounds of the processing framework, exploring a wide range of the current approaches and methods aimed to classify malignant grades of a tumour and particularly to glioma. This seeks to address the following main stages: feature extraction and selection, single classifier systems and multiple classifier systems. The common descriptors used to define glioma such as heterogeneity, contrast enhancement and necrosis, etc, are also explored. Furthermore, the significant predictors including texture features are elaborated. This also seeks to establish a scientific background to identify the research requirements and to select the most appropriate and efficient methods to develop an automated classification system to determine malignant grades of glioma.

Ultimately, this chapter is summarised with the research limitations and boundaries, which are found in existing work, and potential opportunity for novelty in the research undertaken. This also covers a highlight of the proposed solutions to address these limitations. This establishes the guidance trends to the research work undertaken in this study.

2.1 Introduction

Medical image analysis for brain tumour studies has gained significant research attention in recent years due to the raising needs for efficient and objective assessment of a large number of medical images of brain tumours (Bauer et al., 2013). The rapid development in medical imaging techniques and computer-aided algorithms has enabled pioneer methods to become more mature. Practically, computer-aided intelligent algorithms are employed in assisting and automating specific radiological tasks such as the detection and classification of tumours (Birry, 2013). Computer-aided diagnosis (CAD) has been developing fast in the last two decades. The significant purpose of CAD is to support radiologists by using developed intelligent algorithms to provide 'second opinions' where the visual analysis used for the

diagnostic procedure is highly domain experience-dependent, which is subjective and a complex task. Therefore, CAD can offer significant help to enhance the diagnostic accuracy of radiologists, improve inter- and intra-reader variabilities and reduce cancer missed due to fatigue, overlooked or data overloaded (Fujita et al., 2008, Marshkole et al., 2011). CAD can improve the diagnostic ability of the radiologist based on the integration between medical image analysis and machine learning techniques (Duda et al., 2012). Accordingly, pattern recognition techniques including machine learning play vital roles in the development of CAD systems (Illán et al., 2011, Graña et al., 2011, Kumar et al., 2013, El-Dahshan et al., 2014, Moradi et al., 2015, Hsieh et al., 2017b, Citak-Er et al., 2018, Gupta et al., 2019, Latif et al., 2019, Gupta et al., 2017). Pattern recognition includes extracting features from the region of interest (e.g., tumour) and represented in raw data and making a decision based on a classifier outcome, such as classifying the input sample into one of the possible classes. Therefore, CAD becomes a significant supportive approach in enhancing diagnostic accuracy and confidence even for those with high experience (Helbren et al., 2015, Hsieh et al., 2017c). MR image analysis has been the subject of many research works for imaging of human organs including the detection and classification of different types and grades of tumours such as brain, lung, liver and breast. The pre-operative diagnosis of glioma grades is a challenging task, associated with high mortality. The accurate classification of glioma grades plays a vital role in survival prediction and for managing appropriate treatments. A particular focus of this research work is to achieve a classification of malignant brain tumours (glioma), to enable clinicians for an objective, accurate and robust decision-making in distinguishing different grades of glioma. This has created the inducement to develop an automated classification system of glioma grades. Developing classification systems of a brain tumour in medical images are primarily motivated by the necessity of achieving maximum possible accuracy. The general stages to develop the classification system for brain tumour images are mainly based on features extraction, selection and features classification.

2.2 Brain Tumours

Brain tumours can be classified into two types: secondary and primary. Secondary brain tumours develop when cancer cells are transferred from other parts of the body to the brain such as the lung or breast, whereas primary brain tumours arise from tissue cells in the brain. The most common group of cancers are primary brain tumours, which are categorised as malignant (cancerous) or benign (non-cancerous) tumours. Benign tumours develop slowly,

19

and their diagnosis can be a difficult task because their cells resemble normal ones (Kitange et al., 2003). Benign tumours can still be life-threatening if they are in the vital parts of the brain, where they put pressure on sensitive nerve tissue or if they increase pressure within the brain. Although some benign brain tumours may present a health risk, including the threat of disability and death, most of these are successfully treated with surgical removal. Radiation can be applied as an alternative therapeutic way especially when life-threatening circumstances are provoked due to the location of the benign mass (Doolittle, 2004). Malignant tumours are extremely threatening to human life leading to mortality because of their invasive and aggressive progress. Furthermore, the uncontrolled mass development of a malignant tumour leads to several difficulties such as pressure to vital brain structures.

The most common type of primary brain tumour in adults is glioma (Schwartzbaum et al., 2006, Sugahara et al., 1999). Diagnosis of glioma grades is a vital decision because of its impact on patient prognosis. Prognosis can be described as the likely outlook of disease on whether it is likely to be cured and the person's life expectancy. Glioma grades can be identified by pathological evaluation of a brain lesion. To reach a precise diagnosis, it is crucial to understand their morphology, which leads to the classification of tumour grades correctly (Barnett, 2007). The common MRI visual characteristics of glioma grade are explained in the next sections.

2.3   MRI Morphological and Clinical Characteristics of Glioma Grades

Gliomas can be classified based on tumour growth into two categories: low and high grades. Low-grade gliomas are benign and have a better prognosis. On the other hand, high-grade gliomas are in the malignant category and carry a poorer prognosis. Malignant brain tumours can be further categorised into four histological grades with the least aggressive type denoted by grade I. Tumour aggressiveness could grow to the second-grade denoted grade II and then further malignancy growth will lead to the third-grade denoted grade III. Finally, the most aggressive tumour is denoted by grade IV (Louis et al., 2007).

The glial cells can be classified into three types: Astrocytomas, Oligodendrogliomas and Oligoastrocytomas. About 75% of glial tumours is accounted by the Astrocytomas (Behin et al., 2003). Astrocytomas according to WHO can be classified further into four histological grades ranging from grade I, also denoted as Pilocytic Astrocytomas, to grade IV which is known as Glioblastoma Multiform (Moore and Kim, 2010).

2.3.1 Malignancy Assessment and Visual Diagnosis

The malignancy assessment of a brain tumour requires a rather complicated characterisation of MR images and it is generally performed by experienced radiologists. This is a critical task, which should be accomplished with a significant degree of precision. Several criteria (Table 2.1) are commonly used to assist the clinician in the visual diagnosis of glioma grades. The table includes tumour descriptors that can be identified from conventional MR images (Figure 2.1), as well as visual guidelines that are commonly used for the assessment of gliomas grade (Moore and Kim, 2010). However, in clinical practice, there are many other suggested descriptors used by some experts and ignored by others. Furthermore, some of these descriptors are not essential to present in MR images.

Table 2.1 Summary of the brain tumour descriptors and their common incidences used for visual assessment of the malignancy diagnosis of brain tumour grades (Moore and Kim, 2010). The incidences of the tumour descriptors are more likely to occur but are not essential.

	Grade I	Grade II	Grade III	Grade IV
Contrast enhancement	Commonly Enhanced	Usually Absent	Enhanced	Enhanced
Edema	None	Rare	Low attenuation	Present
Necrosis	Usually Absent	Usually Absent	Present but may absent	Present but is not essential to occur
Lesion heterogeneity appearance	None	None, but it may present in some cases with enhancement	High	Highest
Median Age at diagnosis (years)	10	34	41	53
Male/female ratio	1:1	1.18:1	1.8:1	1.5:1
Survival (years)	Variable, cures	5 (2–12+)	2 (1–5)	1 (0.25–1.5)

2.3.2 Low Glioma Grades

Low glioma grades include two sub-grades namely grade I and grade II. Brain tumours with glioma grade I are rare in adults and are primarily seen in children, where the survival rate is variable. It has a better prognosis, and they are treated with surgery that usually leads to a successful cure. Also, they usually do not require postoperative chemotherapy or radiation.
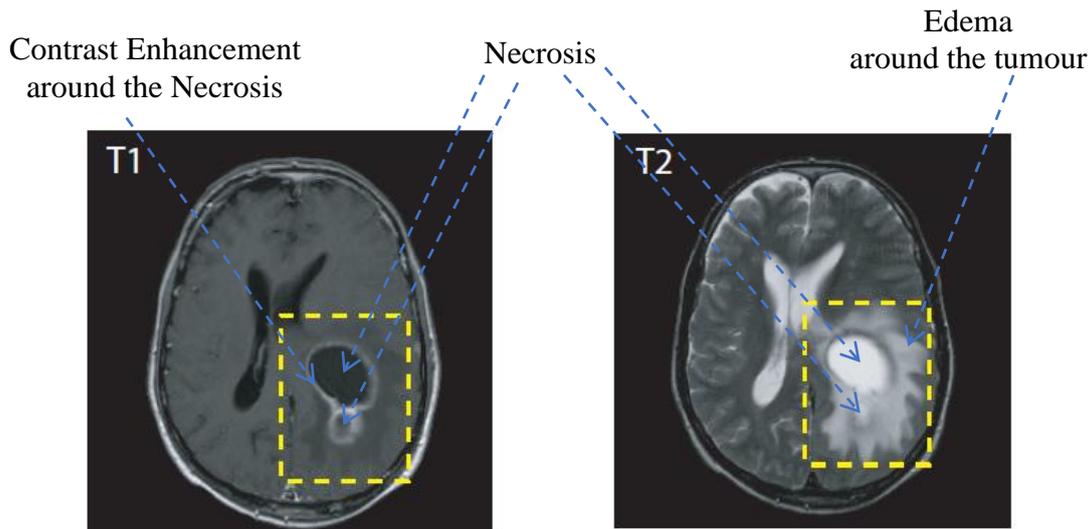
Figure 2.1 Morphology charateristics of brain tumour including the precence of tumour enhancement in T1 after applying the contrast enhacement, necrosis is in the centre of tumour and edema is located around the tumour appeared in T2. These MR images are for grade IV of glioma (Corso et al., 2008).

The morphological characteristics are as follows: first, they are well circumscribed, as they have a well-defined tumour border with homogenise lesions. Calcifications are rare. The majority is presented as a brightly enhancing mural nodule with large cystic lesions on T1 with a contrast agent (Weller, 2011).

Tumours are usually hyperintense (more intense) on T2-weighted images; however, in some cases, the solid component may be hypointense (less intense) in the grey matter, similar to that of cerebrospinal fluid (CSF). Low-grade Astrocytoma is commonly enhanced after the administration of contrast agents. They usually show enhancement which can be nodular or ring-like with a significant cystic component (Grant and Griffin, 2013).

Brain tumours with grade II (Figure 2.2), which are known as Diffuse Astrocytoma, are low-grade tumours of a more invasive nature, which arise typically in the hemispheres of young adults involving white matter cortex. However, focal circumscribed lesions can also occur. The contrast enhancement is usually absent. Grade II tends to grow to a higher histological grade within 3 to 10 years. The survival time is from 2 to 12 years. The MRI findings of this grade are that they are hypointense or isointense on T1-weighted images, and hyperintense on T2-weighted images and Fluid-attenuated Inversion Recovery (FLAIR). The lesion is homogeneous with hypointense mass on T1, and necrosis is usually not appeared in this grade of glioma (Moore and Kim, 2010).

2.3.3 High Glioma Grades

High glioma grades include two sub-grades: grade III and grade IV (Figure 2.2). Brain tumours with glioma grade III , also denoted as Anaplastic Astrocytoma, have clinical characteristics as follows: they are highly heterogeneous in appearance, and indicate more extensive infiltration adjacent tissue than grade II tumours (Wintermark et al., 2005). This leads to a mixed intensity on MRI. Contrast enhancement is commonly perceived, and rapid tumour growth with the development of edema may cause mass shifts. Grade III tumours typically invade white matter zones. Tumour cells can usually originate in the edema zone as well as outside this zone (Moore and Kim, 2010). The prognosis is poor with a median survival of 1–5 years. These tumours have a tendency for progression to Glioblastoma Multiform. The most aggressive type of malignant primary brain tumour is Glioblastoma Multiform, which is commonly found in adults and has the worst prognosis. The survival time is from 0.25 to 1.5 years (Moore and Kim, 2010). These rapidly developing tumours may grow from previous lower grades or occur in older patients. The MRI clinical outcomes of a brain tumour with grade IV are as follows: a thick and irregular rim of enhancement on T1 delineates a lesion. Tumour necrosis usually appears as hypointense areas on T1, frequently surrounded by a ring-like zone of contrast enhancement. While on T2 (Figure 2.2) the lesion is more heterogeneous than other glioma grades, with hyperintense mass and edema (Behin et al., 2003).
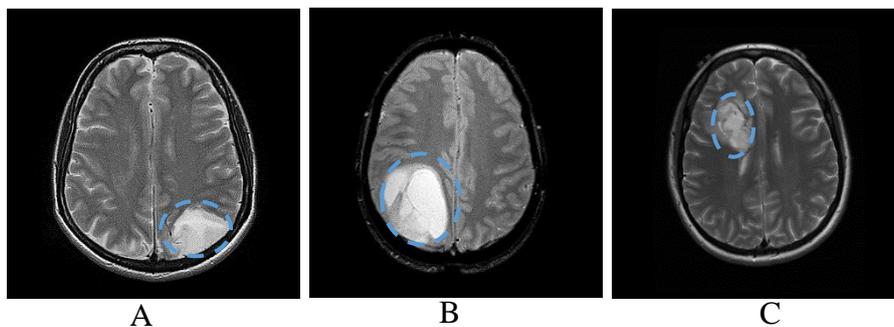
Figure 2.2 Brain lesions highlighted by blue circles in axial T2-weighted MR images where A, B and C represent glioma grade II, III, and IV respectively. Image A shows hyper-intense region with well circumscribed and homogenies lesion. Image B seems to have irregular shape with more heterogeneous tumour. Image C seems to have irregular mass with lower conspicuity and higher heterogeneous lesion.

2.4 Magnetic Resonance Imaging Modalities

Magnetic Resonance Imaging (MRI) is a modern imaging technique, which utilises the fundamental principles of Nuclear Magnetic Resonance. MRI is a non-ionizing technology

developed independently by Edward Mills Purcell and Felix Bloch for precise measurements of frequency and nuclear magnetism of atomic nuclei, which was used mainly for the construction of different materials (Purcell et al., 1946, Bloch, 1946).

The MRI has been used widely in the medical imaging and assisted clinicians to diagnose and manage treatment such as radiation therapy and surgery. This importance is given to MRI technology because of its ability to provide an excellent soft tissue contrast, high signal to noise ratio and high-resolution images (Blink, 2004). Based on the fact that the most parts of biological tissue of the human body consist of water or billions of hydrogen atoms, the MRI technology has gained more advances, which is designed on the basis of sensing the signals that reflects the interaction between an external magnetic field and protons of the hydrogen atoms (Petrou, 2010). Accordingly, the MRI is particularly more appropriate for the imaging of biological tissue such as the brain rather than bones because the latter does not include many hydrogen atoms.

Two relaxation times that could be captured through the relaxation process namely T2 and T1 relaxation times; T2 is also called spin-spin relaxation, which represents the required time to induce the excited net magnetisation to 37% of the original state. During this process, all protons are rotating at slightly different frequencies around the z-axis and exchange energy is started between each other (Dougherty, 2009). T1 is known as the spin-lattice relaxation that represents the required time for relaxing the protons back to recover 63% of the original net magnetisation. The required time for T2 relaxation is always shorter than T1 relaxation time. Different types of tissues can reflect different relaxation times, e.g., the water is de-phased much slower than fat tissue (Blink, 2004).

Fluid-attenuated inversion recovery (FLAIR) sequence is a special sequence that generates adaptive T2-w images by eliminating the signal of the brain edema and other tissue types with high-water content such as CSF. FLAIR is an important tool in tumour delineation with better recognition between tumour and edema as well as tumours that are adjacent to CSF and small hyper-intense tumours (Nabizadeh and Kubat, 2015).

MRI is a very dominant imaging technique and can capture the soft details in tissues; however, some pathological structures have not recognised using only MR relaxation weighting. Considering that some of the tumours produce an abnormal breakdown of the blood-brain barrier, a contrast agent that distributes throughout the extracellular space became a significant choice to enhance MR image contrast. Gadolinium chelates are the

most common agents used to enhance MR images. Gadolinium is used with T1 and the contrast enhancement of T1 images depends on the concentration of Gadolinium injected to acquire a better-enhanced brain tumour (Dougherty, 2009).

2.5   MRI Against CT for Diagnosis of a Brain Tumour

The most common imaging techniques used for the diagnosis of brain tumours are Computerised Tomography (CT) and MRI. CT is one of the largest accessible conventional imaging techniques used in most clinical centres. CT scans provide a contrast-enhanced image of small amounts of calcification within tumours (Ferlay et al., 2015). It is more cost-effective and needs a shorter scanning time. However, even though CT has many benefits, MRI, when compared with CT, is a more desirable technique when diagnosing brain tumours because it has several benefits. Firstly, MRI has a higher soft tissue contrast with high spatial resolution. Therefore, it can provide a better sensitivity of the tumour description. Moreover, MRI provides multiplanes imaging – with transversal (axial), sagittal and coronal planes. Besides, MRI can be repeated for monitoring tumour growth and treatment progress without any dangerous ionising radiation. Finally, MRI is a non-invasive technique, which is painless and it can be established without contrast enhancement (Saad et al., 2015). Accordingly, this thesis concentrates on the MRI technique to extract recognised features required to the classification system for glioma grades.

2.6   Conventional Against Advanced MRI Modalities

MRI imaging is a non-invasive technique used widely in the clinical diagnosis of brain tumours. MRI techniques, in general, can be categorised into two approaches: conventional and advanced MRI techniques, the conventional techniques such as T1, T1 with enhancement (T1c), T2 and FLAIR-weighted images are offered in any MRI clinical centre, while the advanced MRI techniques such as diffusion-weighted magnetic resonance imaging (DWI) and MR spectroscopy are costlier and less available in MRI clinical centres. The use of the advanced MRI methods as opposed to the conventional MRI approach could lead to more accurate results in tissue characterisation of benign against the malignant tumour. A possible example would be the research carried by Roshdy et al. (2010), in which the clinical diagnosis drawn from both the advanced modalities including DWI, MRS and the conventional MRI modalities consisting of T1, T2 and T1c were compared. The obtained results suggest that the advanced imaging methods have shown a better result when used for the clinical diagnosis of brain lesions and outperform other conventional MRI techniques.

Many studies relied on advanced MR modalities in developing a classification system of brain tumours (Citak-Er et al., 2018, Aragao et al., 2014, Geneidi et al., 2015, Zhang et al., 2017). On the other hand conventional MR techniques were also utilised by many other recent research work in different applications including the segmentation and classification of brain tumours and they achieved promising results (Hsieh et al., 2017b, Hsieh et al., 2017a, Mohsen et al., 2018, Hasan and Meziane, 2016, Hasan et al., 2016a, Anaraki et al., 2019, Ye et al., 2017, Khawaldeh et al., 2017). However, the conventional MRI techniques are more desirable due to their common availability and lower cost when compared to the advanced imaging techniques. Therefore, in this thesis, the concentrate is given to conventional MRI techniques.

Several descriptors are used to assess the malignancy level of a tumour and are exploited to extract recognised predictors which are correlated with different malignant grades of a brain tumour. Predominantly, the common descriptors used to determine the malignant grades of glioma are: heterogeneity, contrast enhancement, necrosis, edema, vascularity and cellularity. These tumour descriptors and the methods and features used to measure them are explained below.

2.7   Tumour Heterogeneity

The extensive varieties of genetic, molecular and cellular modifications, which may occur during the progression of tumour growth, are complex and described as heterogeneity. Cancers show various degrees of heterogeneity such as gene expression and a cellular morphology that can be specifically investigated using different imaging methods.

Tumour heterogeneity can be recognised from magnetic resonance images from the significant variations in the image intensity. Heterogeneity is correlated with tumour aggressiveness (Skogen et al., 2016). Particularly, relative growth of tumour grade reflects increasing in tumour heterogeneity. Although appropriate indices of heterogeneity have already shown good predictors of tumour aggressiveness progression, there is still a lack in methods of investigating and evaluating the impact of heterogeneity, due in part to a poor understanding of the molecular mechanisms underlying it. A better and more systematic appreciation of tumour heterogeneity is crucial for drug development as well as for the accurate assessment of response to treatment (Davnall et al., 2012).

Several approaches are used to measure tumour heterogeneity and analyse image biomarkers as malignancy indicators. Significantly, texture analysis reflects an objective analysis for

tumour heterogeneity in different medical images (Tantisatirapong, 2015). Furthermore, texture features of MR images have been used efficiently for monitoring malignancy progression of a pathological lesion and specifically for the recognition and identification of morphological characteristics of brain tumours (Holli et al., 2010, Roy et al., 2013).

The texture feature describes the appearance, arrangement and structure of an object within an image. Texture can provide an abundance of visual information (Materka and Strzelecki, 1998). Image texture analysis can access complex visual patterns comprising sub-patterns or entities. Furthermore, texture can be utilised to distinguish the intensity, brightness and distribution of image information. Texture analysis provides an efficient and reproducible diagnostic tool for MR image analysis (Kassner and Thornhill, 2010). It has been used to assess MR images of biological tissues that contain significant amounts of microscopic details. Texture analysis can better characterise patterns of lesion compared to the human visual perception that is highly subjective and dependent on expertise. Accordingly, tumour patterns can be measured automatically based on texture analysis and hence are independent of clinician expertise. It has been shown that texture analysis can outperform visual examination, in the discrimination between pathological and healthy tissues (Herlidou et al., 1999, Hsieh et al., 2017c, Dennie et al., 2016, Lerski et al., 2015, Chevrefils et al., 2018). Image texture has this advantage because it is highly sensitive to the variation of intensity in image pixels. Significantly, texture analysis has been used for tumour detection, diagnosis, segmentation and classification as well as for distinguishing between malignant and benign lesions.

Methods to measure image texture feature can be categorised into four groups: statistical, transform, model-based and structural processes. The statistical techniques include histograms, co-occurrence matrices and run length matrices. The transform approach includes Gabor (Qian and Chen, 1993), Fourier (Bracewell, 2000) and Wavelets tools (Walnut, 2013), which are efficiently used for different medical applications (Castellano et al., 2004). However, transform-based methods, in general, do not consider the spatial relationship of texture information, while model-based approaches represent image texture based on sophisticated models. The parameters in the model-based method are estimated and then prepared for the image analysis. The estimation of their parameters carries a higher computational cost and is highly sensitive to orientation selectivity (Castellano et al., 2004). Structural-based methods represent a texture as a connected set of pixels with the same properties and employ morphological techniques such as opening binary images. The

structural methods are recommended for textures with a large structure (macrostructure). However, this approach has the drawback of not being able to capture a texture with no structure (Maani et al., 2016). Other features that can be extracted from images are the colour and shape information. Nevertheless, a brain tumour has no colour information and no distinct shape could be used for the assessment of different glioma grades.

The textural extraction methods of the medical images that are particularly employed for the malignancy assessment of tumours are reviewed in the following subsections.

2.7.1   Transform Texture Analysis

Transform texture features are multiscale representations resulting from the decomposition of an image into a set of sub-images revealing image structures and details at multiple orientations and multiple scales. Each sub-image relates to a frequency sub-band. Such decompositions are performed as follows: the image data is filtered using linear filter banks. Afterwards, each filter output is up/down sampled producing several image representations with specific properties such as multiple scales, frequency selectivity and directional orientation (Baaziz et al., 2010). Transform texture analysis has been used to assess tumour malignancy growth. A possible example for this approach would be the research carried out by (Zacharaki et al., 2009), where the Gabor transform was employed to distinguish between metastases and glioma, as well as to discriminate glioma grades into high against low grades. However, due to a high correlation between adjacent pixels of MRI images, the Gabor transform produces many redundant features (Nabizadeh and Kubat, 2015). Transform texture features require less computational time and thus facilitate the analysis of large datasets (Drabycz et al., 2010, Kassner and Thornhill, 2010). Nonetheless, transform texture analysis, in general, has the disadvantage that there is a relative lack of localised frequency content of spatial information (Materka and Strzelecki, 1998, Tantisatirapong, 2015, Kumar and Singh, 2018).

2.7.2   First-Order Texture Feature Based on Histogram Analysis

Texture features based on histogram analysis are a popular method for characterising tumour heterogeneity. Histogram analysis incorporated with statistical measurements was used to analyse distributions of image intensities as a biomarker for tumour heterogeneity. The statistical measures used with histogram analysis are as follows: entropy, standard deviation, mean, mode, kurtosis, skewness, maximum, minimum and percentiles.

Histogram methods can be significant predictors in the assessment of tumour malignancy. Many studies have investigated the use of histogram analysis for the discrimination between low and high grades. Specifically, investigation of these statistical measurements has shown various degrees of correlation with tumour heterogeneity. For example, it was found that the standard deviation shows a correlation with the heterogeneity of brain tumours as its value increases with the growth of tumour heterogeneity (Skogen et al., 2013, Skogen et al., 2016). Entropy and uniformity have also been explored to measure the tumour heterogeneity. It is concluded that the entropy was higher and uniformity is lower for a whole tumour with greater heterogeneity (Ganeshan et al., 2012, Ng et al., 2013). It is argued that entropy is significantly higher in high-grade gliomas than low-grade tumours while skewness, kurtosis and percentiles are less correlated (Ryu et al., 2014).

Similarly, the fifth percentile derived from the apparent diffusion coefficient (ADC) shows promising results for differentiating high and low-grade gliomas (Song et al., 2013, Kang et al., 2011). Likewise, further different percentiles have also been used such as the 10th, 25th, 50th, 75th and 90th to test the correlation between a low glioma grade subtypes and histogram analysis using ADC (Tozer et al., 2007). In the same direction, it was claimed that both 90th and 95th percentiles produce better results than the standard deviation in predicting the histological grade of endometrial cancer (Woo et al., 2014).

Other studies examined the combination of several measurements calculated from the histogram analysis on the classification of the malignancy degrees of a tumour. A possible example would the research conducted by Carter et al. (2013) in which several statistical measures such as standard deviation, skewness and kurtosis incorporated the histogram analysis are combined and evaluated to differentiate benign from malignant ovarian masses. Similarly, the combination of mean, entropy and uniformity was determined using a range of filters applied to medical images to highlight texture for lung cancer (Ganeshan et al., 2010), and that for colon cancer (Ng et al., 2013).

Histogram analysis has been widely used as a prediction tool for tumour heterogeneity. It has the advantage of being easy to implement. Furthermore, it presents a non-invasive assessment of malignancy of brain tumours rather than the clinical approaches which are biopsy-dependent (Just, 2011). Nevertheless, a significant limitation of histogram analysis is that it does not consider the spatial distribution of tumour information. Moreover, histogram features have a high dependency on data distribution and thus may perform poorly if the data has high heterogeneity (Rose et al., 2014). Besides, many texture patterns cannot

be recognised using the first order analysis (Pantelis, 2010, Kassner and Thornhill, 2010). Also, it has limited robustness in distinguishing unique textures in specific applications as this method does not consider the interaction and spatial relationship of the neighbouring pixel value (Florez et al., 2018).

2.7.3   Second Order Texture Analysis

Second order analysis is an efficient tool in the classification of image patterns. It was reported that it is possible to resynthesize textures with the same visual properties if enough second-order statistical information is taken into account (Nielsen et al., 2008). For example, the grey level run length matrix (GLRLM) is a known technique to extract second-order features. However, The major drawback of GLRLM is that it produces features which are highly correlated and that it offers insufficient information about the image texture (Tang, 1998). Hence, it has limited use in practice compared with other texture extraction methods (Tantisatirapong, 2015).

Grey level co-occurrence matrix (GLCM) is a second-order texture method, first proposed by Haralick et al. (1973), and is the most powerful method for extracting second-order features and measuring the spatial relationship among image pixels. It investigates the relationship between a pixel pair in the region of interest of an MRI a brain tumour. Furthermore, GLCM has been used commonly for heterogeneity assessment and predicting the level of tumour malignancy and it achieved remarkable results in the classification of tumours in MRI (Kovalev and Kruggel, 2007, Bonilha et al., 2003, Wibmer et al., 2015, Larroza et al., 2016). For instance, GLCM incorporated with 14 textural measures such as contrast and entropy was performed to improve the discrimination performance between benign and malignant lesions for breast cancer (Gibbs and Turnbull, 2003). Similarly, (Gómez et al., 2012) used the advantage of GLCM in a comprehensive analysis of texture evaluation to discriminate between benign and malignant tumours in breast cancer using a wide range of angles and distances of the GLCM. Further example used the merit of GLCM is the research carried by Subashini et al. (2016), in which five predictors were selected manually; these parameters are contrast, dissimilarity, angular second moment, entropy, maximum probability and inverse difference moment driven from GLCM and employed for the classification of glioma tumours into high against low-grades. Notably, the significant strength of GLCM relies on measuring the joint tumour probability of spatial distributions of pixel pairs that incorporate statistical predictors such as uniformity, homogeneity, energy, entropy and correlation. GLCM has shown a promising result in the detection of the internal

intensities arrangement and directionality of tumour image. However, it has the limitations of adding complexity, large memory storage and computation time requirements (Tantisatirapong, 2015). The combination of several features could lead to potential improvement in classification accuracy. For instance, the combination of histogram features and GLCM was shown to improve discrimination accuracy between glioma grade IV and lower grades (II and III) (Hsieh et al., 2017b). However, the results suggest that the fusion of the two features have slightly better results than the use of GLCM individually. Involving multiple features may lead to improvements in the classification performance nevertheless it is not guaranteed. This approach also adds more complexity and computation time, particularly for large datasets. It could also lead to an increase in the number of redundant features that degrade the classification accuracy.

In conclusion, among the different extraction methods dedicated to image texture features, the GLCM is the most popular spatial method to determine an objective analysis of texture feature of medical images (Maani et al., 2016). GLCM has been proven to outperform other feature extraction methods such as the Gabor, wavelet and Fourier transforms (Materka and Strzelecki, 1998, Kharrat et al., 2010, Larroza et al., 2016, Liu et al., 2018). A GLCM based texture analysis is therefore primarily considered in this thesis.

2.7.4 Three-Dimensional Texture Analysis

Texture analysis based on three dimensions provides complementary information that could lead to an improvement in the classification accuracy. Significantly, this enhancement can be achieved based on information accessible in the Z-dimension of an image which reflects a vital part of the signals obtained from MR images. Texture features based on a 3D representation have produced promising results in different applications such as biometric recognition (Yazdi et al., 2007) and iris identification (Chen et al., 2009).

An example of the 3D analysis of the image texture feature is the research conducted by Chen et al. (2007), where 3D texture analysis based on 3DGLCM was investigated and compared to 2DGLCM. In this regard, the obtained results indicated that 3DGLCM has better classification results than 2D analysis in distinguishing between malignant and benign MRI breast lesions. Similarly, a comparative study between 2D and 3DGLCM was demonstrated for the segmentation of brain tumour sub-regions. The results indicate that 3D textural features could enhance the discrimination between different tumour sub-regions, which is very challenging to distinguish the different parts of a brain tumour with the only

naked eye, and without the need for contrast agents (Mahmoud-Ghoneim et al., 2003). A further example is the research undertaken by Fetit et al. (2015), in which the ability of 3D texture analysis driven from the fusion of multi-slice of conventional MRI (T1, and T2) is evaluated for the classification of paediatric brain tumours. The results obtained suggest that the classification accuracy can be boosted up using the 3D texture analysis in the classification of childhood brain tumours.

Texture analysis based on 3DGLCM has the advantage of considering all the information available in the image signals. It also reflects the complementary spatial details of the tumour tissues. However, it requires large amounts of memory, high complexity and computational cost (Hsieh et al., 2017b). Accordingly, to include and investigate all the possible textural information extracted from the medical image of a brain tumour, this thesis takes into account the 3D analysis of the GLCM in addition to 2DGLCM.

2.8   Brain Tumour Vascularity

Tumour vascularity is the relative amount of blood vessels compared to those in the surrounding areas (white matter) (Moenninghoff et al., 2010) and correlates with the growth of tumour grade (Ruoslahti, 2002). The number of blood vessels in a tumour varies structurally with the growth of tumour grades. The relative cerebral blood vessels (rCBV) have been used for assessing tumour vascularity, and this feature is measured using the perfusion weighted imaging (PWI) technique. The rCBV has been employed successfully in predicting glioma grades. For example, it was shown that the rCBV technique has a better diagnostic performance compared to other methods such as metabolite ratios driven from the proton MR spectroscopy for predicting glioma grades (Law et al., 2003). Similarly, the rCBV is assessed against Ktrans, and it is found that the rCBV revealed a better accuracy compared to the other technique for glioma grading (Law et al., 2004, Law et al., 2006). In the same manner, rCVB, metabolite ratio and apparent diffusion coefficient (ADC) were measured to examine their correlation with glioma grading. It was reported that rCVB, metabolite ratio and ADC are as follows: lower, higher and higher for lower glioma grade, respectively, and vice versa for higher glioma grade (Aragao et al., 2014). Likewise, rCVB and fractional anisotropy (FA) values derived from diffusion tensor imaging (DTI) were evaluated, and their ability to determine different glioma grades is investigated. As a result, the use of DTI was found to play a crucial role in the grading the gliomas. Additionally, a combination of both techniques improved the accuracy of glioma grading (Geneidi et al.,

2015). Ultimately, these approaches have suggested promising results in the classification of glioma grades, which were extracted from advanced imaging techniques such as the rCBV extracted from PWI, the ktrans measured from diffusion-weighted MR image (DWI), and the metabolite ratio obtained from spectroscopy imaging techniques. However, recently, this approach has been costlier requiring advanced MRI techniques, exceptional setting and significant experience, which have limited availability in the clinical radiology practice.

2.9 Brain Tumour Cellularity

Tumour cellularity is the relative proportion of a tumour and healthy cells in a sample. Pathological examination of sectioned specimens based on deoxyribonucleic acid (DNA) analysis is the traditional technique to assess tumour cellularity. Nonetheless, this technique is subjective due to the heterogeneity within lesions. There are cellularity variances among the samples viewed during the pathological review (Song et al., 2012). Accordingly, as an alternative approach to indicate the cellularity of a brain tumour, apparent diffusion coefficient (ADC), which was extracted from diffusion-weighted MR image (DWI) technique was used.

The correlation between brain tumour cellularity and ADC was evaluated. It was found that the histopathological information associated with tumour cellularity was significantly correlated with ADC and hence it confirms its ability to classify different glioma grades (Sugahara et al., 1999, Kono et al., 2001). Similarly, it was assessed successfully to determine the malignancy levels of a liver tumour (Taouli et al., 2003). However, this technique needs an expensive and advanced MRI techniques and has limited availability in MRI clinical centres.

2.10 Malignancy Diagnosis of Brain Tumour Descriptors

Several tumour descriptors are used in clinical diagnosis, and can be extracted from conventional MR images such as necrosis, edema, post-contrast enhancement and non-enhancement of a brain tumour, which are used significantly to assess in the malignancy level of a brain tumour (Moore and Kim, 2010). Visualisations of these brain tumour descriptors (Figure 2.3) show different recognised regions for a brain tumour, which are extracted using T1 with enhancement (T1c). The figure also illustrates how the identified tumour descriptors can appear through T2 images. It can be noted that the inner regions include a solid tumour portion where the solid portion comprises a necrotic area and an active

part. The active portion of a brain tumour could be enhanced with contrast partially or fully or could be non-enhanced. The solid portion of a tumour may also be surrounded by edema which is also known as peritumoral (Corso et al., 2008).

One of the major indicators to favour the diagnosis the grade of glioma is the presence of a necrotic region. Necrosis can be described in pathology as a dead cell in brain tumour tissue, whereas on an MRI it appears as areas of non-enhancing hypointensity on T1 (Chow et al., 2000), commonly encircled by a ring-like zone of contrast enhancement. The presence of necrosis has been recommended as a significant predictor for the diagnosis of a higher grade. Nonetheless, the presence of tumour necrosis has not been essential for the diagnosis of glioma grade IV (Barker et al., 1996).

The signal intensity of a necrotic region can also be identified from DWI. To elaborate in term of visual assessment, necrosis portion has significantly low signal intensity near to that of cerebrospinal fluid (CSF), slightly low signal intensity between that of CSF and that of healthy brain tissue, or slightly high, or significantly high compared with healthy brain tissue (Lai et al., 2002). Tumour necrosis was employed as a predictor for glioma grades depending on whether a necrotic area is absent or not.

Generally, a radiologist examines the tumour descriptors visually from MR images, particularly, evaluating the presence of post-contrast enhancement, non-enhancement necrosis and edema, estimating their differences from each other to make a clinical decision. When these descriptors are visually present, this raises the probability of whether an unknown tumour is of a high or low grade.
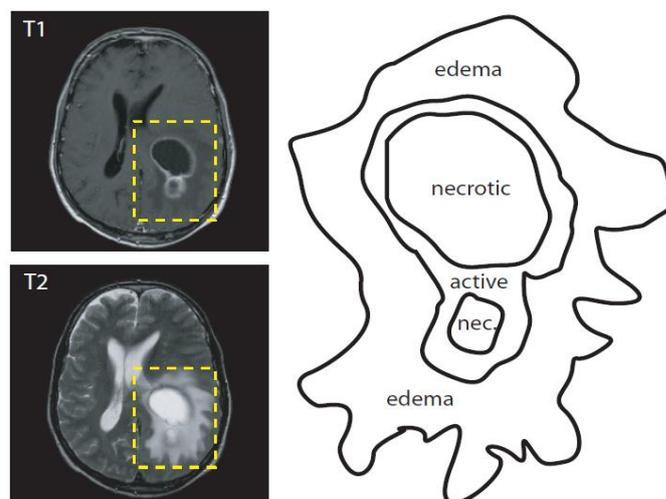
Figure 2.3 An example of the recognised regions of a brain tumour in the presence of the tumour descriptor. These MR images are for grade IV of glioma (Corso et al., 2008).

Many studies aimed at separating and recognising these tumour descriptors objectively. An example to demonstrate this task is the research carried by Mahmoud-Ghoneim et al. (2003), in which a comparative study between 3D against 2DGLCM was performed to differentiate necrosis, solid tumour, edema and surrounding white matter. It was found that the use of 3DGLCM as a 3D texture analysis provides better results for the segmentation of the brain tumour subregions. Another possible example would be the research conducted by Li et al. (2006) in which the presence of necrosis, edema and post contrast enhancement combined with many other clinical data are assessed to predict the malignancy grades of a brain tumour. However, these features are computed manually with the help of domain experts. Likewise, some of the tumour sub-regions, which were delineated manually, and then used to provide multiple features, were employed to predict different grades of glioma (Zacharaki et al., 2009). Similarly, the presence of a necrotic portion was assessed against histology features to improve the diagnostic performance of patient survival time. It was shown that patients who have necrosis could get worse survival times. It was found that the diagnostic accuracy can be further improved by incorporating necrosis compared to the use of the histology features alone (Lasocki et al., 2015). In the same way, (Geneidi et al., 2015) employed different statistical measures including mean, minimum and maximum of fractional anisotropy (FA) measured from DTI. These measures were extracted from the tumour mass, necrotic regions and edema. These predictors were examined to assess the correlation between these features and classification of glioma grades into low against high grades. Hence, it was found that FA measured from necrotic and tumour mass has a great correlation with glioma grading while there was no positive correlation with features extracted from the edema.

To summarise, the presence of the descriptors of a brain tumour as well as the features extracted from them have been assessed as predictors to contribute in the diagnosis of the malignancy grades of a brain tumour. However, these predictors have been identified visually with the help of a domain expert. The visual assessment has the limitations of subjectivity and suffers from inter and intera variabilities. Consequently, the inference findings based on the visual analysis to identify the tumour descriptors may lead to uncertainty in the clinical diagnosis of a brain tumour. Therefore, to have an accurate and automated diagnosis of malignancy growth of a brain tumour, investigation in the statistical analysis of the presence of these tumour descriptors is significantly required.

2.11 Brain Tumour Segmentation

In order to extract features only from the region of interest (ROI), i.e., a tumour, it is most significant to initially split the MR image into 'meaningful' (brain tumour) and 'meaningless' (other undesired structures); this process is known as segmentation. Segmentation is valuable in the medical applications by classifying medical image pixels to different structural areas such as blood vessels, bones and tissues. Also, it is employed to classify the pixels of pathological structures such as cancer and multi sclerosis tumours (Dvořák et al., 2013). The recommended criteria to achieve a good-quality segmentation of ROI are as follows. First, the segmented ROI should have a smooth boundary. Second, the adjacent areas to the ROI should have a noticeable difference. Third, the internal region of the segmented ROI should be homogenous and not include any holes (Al-Waeli, 2017).

There are three categories of segmentation: manual, semi-automated and fully automatic. The manual segmentation where a qualified expert delineates the extent of the lesion visually depending on user interaction and the domain- knowledge of the expert, and thus the resulting output is generally recognised as the gold standard or the ground truth. Fully automatic segmentation needs no user interaction and requires less processing time. However, it is likely to accomplish less satisfactorily on medical images because of the complexity and inhomogeneity of anatomical texture. Semi-automatic segmentation techniques are accomplished based on integrating the user's supervision with the computer-aided algorithm, and are used when the pathological region is more accessible to recognise visually but not automatically (Tantisatirapong, 2015). Extensive works dedicated to automated brain tumour segmentation which are widely expanded because of the rapid development in the medical imaging techniques as well as due to the fast progress in computer-aided image analysis (Menze et al., 2015, El-Dahshan et al., 2014, Bauer et al., 2013).

2.12 Features Selection and Relevance Analysis

Relevance analysis aims to identify subsets of the most vital features and removes irrelevant and redundant features that affect the classification performance. The irrelevant feature can be eliminated without any effect on the classification accuracy. Therefore, discarding one of them will not affect the classification accuracy (Gómez et al., 2012). Feature relevance analysis is used for the following reasons. It increases the accuracy of a classification model through selects an accurate feature subset, shrinks the complexity and reduces the storage

requirements especially for a system that deals with hundreds or thousands of features making the features easier to interpret. It also supports machine-learning algorithms to train faster (Tantisatirapong, 2015, Pantelis, 2010). Feature relevance analysis is one of the efficient and common methods of the feature selection applications used to select the most appropriate feature subset. It was stated that a good feature subset is one that includes features highly correlated with the truth class label and uncorrelated with each other (Hall, 1999). Feature selection applications can be categorised into two main approaches, namely, filter and wrapper methods. The filter approach that can also be denoted as the relevance analysis method measures the amount of the feature relevance with its truth class label so that the fundamental properties of the data for each feature are independent of each other. A feature relevance score is usually obtained and then features are ranked. After that, features with low scores are eliminated. Afterwards, the selected subset of features becomes the input to the classification algorithm (Saeys et al., 2007). A possible examples would be the researches carried by (Baboo and Sasikala, 2010, Hsieh et al., 2017b, Citak-Er et al., 2018, Gómez et al., 2012, Eliat et al., 2012) in which the features are selected prior to the classification process based on the filter approach. The purpose was to select only the features that meet the significant p-value of less than the critical value while excluding other features that did not reject the null hypothesis. The other approach includes wrapper techniques. This differs from the filter methods as they consider the interaction between features and the classification outcome and incorporate a learning algorithm to search for the best subset of features (Li et al., 2006, Zacharaki et al., 2009).

Filter techniques that are performed independently to a machine-learning algorithm are fast, easy to implement and need low computational cost. However, this approach ignores the interaction between feature subsets and the classification outcome. Hence, this may lead to the selection of features that do not support the classification outcome. Furthermore, ignoring the correlation between features potentially causes higher redundancy in the feature space (Luts et al., 2007). While the disadvantage of the wrapper method is that the selection is highly dependent on the combination of features being examined. Additionally, the framework for this approach has higher computational cost and requires an extensive heuristic process to find the optimal feature subset (Guyon and Elisseeff, 2003, Saeys et al., 2007).

One-way analysis of variance (ANOVA) is used successfully to investigate the relevance analysis (Hasan et al., 2016b, Li et al., 2015), while Pearson correlation is widely used to

measure the linear correlation and dependence between two variables (Ly et al., 2018, Labani et al., 2018). The details of these methods are explained in the next subsections.

2.12.1 One-Way Analysis of Variance

The one-way analysis of variance is an efficient and common statistical method used for evaluating the difference level between two or more independent groups of samples in a feature vector. This is performed by testing whether the means of multiple groups are significantly different. In this technique, the null hypothesis is that there is no difference in the means of groups. This technique relies on the assumption that all instances are normally distributed independently with equal variances of features in different classes. This method is used as a guide to select the most significant features based on p-value and F-ratio. The p-value is the probability of the test is at least equal to or less than the critical value of the test, i.e., 0.05 or less. When applying "the ANOVA on a two-class scenario, it is equivalent to the two-sample t-test assuming equal variances" (Dubitzky et al., 2007). After applying ANOVA, the p-value is used to indicate the features that have the high difference between different means of groups of samples, where features that have a p-value less than the critical value are selected. Hence, features with higher discrimination power are selected and others are discarded.

The value of the F-ratio offers an indicator of class separation where the significant value refers to the higher separation. The F-ratio (Eq. 2.3) is computed from applying ANOVA and is measured by calculating the ratio of the between-class variance (Eq. 2.1) to the within-class variance (Eq.2.2) (Johnson and Synovec, 2002).

$$\sigma_{BC}^2 = \frac{\sum_i^n (X_i - \overline{X})^2 n_i}{(1 - k)} \qquad 2.1$$

where σ_{BC}^2 is the between-class variance, X_i is mean of the ith class, $\overline{X}$ is the overall mean, n is the number of features for each class and k is the number of classes.

$$\sigma_{WC}^2 = \frac{(\sum_i \sum_j (X_{ij} - \overline{X})^2) - (\sum_i (X_i - \overline{X})^2 n_i)}{(N - k)} \qquad 2.2$$

where σ_{WC}^2 is the within-class variance, X_{ij} is the ith features of the jth class and N is the total number of features for all classes.

$$F - ratio = \frac{\sigma_{BC}^2}{\sigma_{WC}^2} \qquad 2.3$$

After measuring the F-ratio and referring to the ANOVA and critical values table, if any feature has a p-value less than the critical value then the feature is selected. Usually when the p-value is less than 0.05, the feature is considered as a significant feature. ANOVA has been used widely in many applications as a rapid statistical method and a good indicator to measure the significance level of extracted feature before classification. Nevertheless, similar to any other filter method used for features selection and reduction task, the relevance analysis examined for a feature is demonstrated independently to any other feature. Therefore, the interaction between features is not taken into account as well as the impact of a subset of features on the classification outcome is ignored. Therefore, in this study, this problem is addressed and overcome by using a wrapper strategy through using different machine learning algorithms incorporated with ANOVA and the search process is guided by Pearson correlation and several machine learning algorithms.

2.12.2 Pearson Correlation

Pearson Correlation is a well-known efficient statistical method used to measure the linear association between two vectors. It examines the strength of the linear relationship between the two continuous variables. The correlation coefficient is determined upon a range that varies from -1 through 0 to +1. No correlation is represented by 0 while the perfect correlation between the two vectors is determined by either -1 or +1. When the correlation is positive, it means that both vectors are increasing. While, when one increases as the other decreases represents a negative correlation. Pearson correlation coefficient is calculated using Eq. 2.4 (Swinscow and Campbell, 2002).

$$Pearson\ Correlation\ (r) = \frac{\sum_i^N (x_i - \overline{x})(y_i - \overline{y})}{(N-1)SD(x)SD(y)} \qquad 2.4$$

where x and y are two feature vectors; $\overline{x}$ and $\overline{y}$ are the mean of the feature vectors respectively; N is the total number of samples and SD is the standard deviation.

The null hypothesis is that there is no association between the two variables undertaken the test. After the Pearson correlation is measured, Eq. 2.5 (Swinscow and Campbell, 2002) is applied, then both results of the Pearson correlation and the degree of freedom (D_f) are used to get the significance level p-value. Note that if the p-value is less than the critical value, then the null hypothesis is rejected indicating strong evidence that there is a linear correlation between the two variables.

$$P_t = r \times \sqrt{\frac{D_f}{1 - r^2}} \qquad\qquad 2.5$$

where P_t is the t-test, r is the Pearson correlation coefficient, D_f is the degree of freedom which is equal to N-2, where N is the total number of samples.

2.13 Feature Classification

Classification is the categorisation of objects into classes such as grouping samples into abnormal or normal. Image feature classification is a crucial step for automation and integration of the diagnostic system. Classification methods are categorised into two approaches: unsupervised and supervised. Also, the classification methods can be developed based on two trends: single classifier system and multiple classifier systems. Further details are described in the following subsections.

2.13.1 Single Classifier System

A single classifier system is accomplished based on using only one classification model such as Decision Tree (DT), Support Vector Machine (SVM), Artificial Neural Network (ANN), K-Nearest Neighbour (KNN), Linear Discriminate Analysis (LDA) (Deepa and Devi, 2011). In medical image analysis, many studies have commonly used single classifier system to perform different tasks such as the segmentation, detection and classification, utilising both supervised and unsupervised approaches.

Possible examples that used the supervised approach would be the research carried by (Devos et al., 2005, Li et al., 2006, Luts et al., 2007, Zacharaki et al., 2009), in which a single classifier system based on SVM was used successfully for tumour grade identification. However, DT and SVM classifiers have shown superior results as compared to other classifiers including logistic regression, K-nearest neighbour, and the linear discriminate analysis in binary and multi-class classification of brain tumour subtypes using pathological medical images (Das et al., 2018). Nevertheless, the KNN classifier has been reported to be an excellent classifier in the segmentation and binary classification for medical images of cervical cancer (William et al., 2018). Similarly, the KNN classifier outperforms other classifiers such as SVM and DT in the detection of sclerosis issues from healthy controls in MR brain images (Zhang et al., 2016).

On the other hand, possible examples that used the single classifier system in unsupervised classification approach would be the research conducted by (Ye et al., 2002, Javed et al.,

2013) in which the single classifier system was demonstrated based on fuzzy rules. Likewise, single classifier system in unsupervised approach was used based on clustering methods for the classification of malignancy grades of brain tumour in MR images (Inano et al., 2014, Subashini et al., 2015).

To summarise, unsupervised algorithms extract patterns from the input data using statistical techniques. Examples of this approach are fuzzy rules and clustering techniques. Unsupervised algorithms, in general, are easy to implement and relatively fast because they do not require any training process. However, they have limited accuracy by ignoring prior knowledge related to the given training samples. Furthermore, the patterns driven using unsupervised approaches may produce spurious classes. The second category is the supervised classification approach where the classification procedures include two phases of processing: training and testing. In the initial training phase, a description of each training samples based on image features is determined and used to design a classification model. In the testing phase, the model based on the same feature space is used to classify the unseen sample. It is worthwhile noting that this approach is the most widely used in medical images (Erickson et al., 2017) due to its superior results in classification accuracy, based on takes advantage of training on given data, and optimised model that can be used effectively to predict the label of unseen samples. Consequently, in this thesis, a supervised classification approach is considered.

There are so many classifiers available and are used in different applications. A detailed description of the most common single classifier models (Kulkarni et al., 1998, Kotsiantis et al., 2007, Lu and Weng, 2007) and their advantages and disadvantages are introduced in the following subsections.

2.13.1.1 Decision Tree

A decision tree (DT) is defined as tree-like structures in which the nodes represent the input features and each branch indicates a value that the node can produce an output decision. The process starts at the root node and instances are classified and arranged to correspond to the input features values (Kotsiantis et al., 2007). This classification algorithm offers many advantages such as it is a non-parametric approach and is not dependent on the input data distribution. Also, it can deal with the non-linear relationship between class labels and features. Finally, it is easy to interpret as it has a relatively simple classification structure (Friedl and Brodley, 1997). However, DT also has the limitation as it can underfit or overfit the model, especially when using a small sample size of the dataset (Song and Lu, 2015).

2.13.1.2 Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a parametric classification (Fisher, 1936), in which the linear relationship between feature vectors is used to discriminate between classes. To guarantee the maximal separation between classes, it maximises the ratio of between-class variance to the within-class variance. This method has the advantage of its simplicity, and it is fast as it performs well when there are less overlapping between the sample distributions. However, it is affected by the data distribution thereby performing poorly when there is a significant overlapping among class's data (Lu et al., 2005).

2.13.1.3 Support Vector Machine

Support vector machine (SVM) identifies a hyperplane that separates input data into two classes. Input data is mapped into a features space with higher dimensionality. Maximising the gap between the two classes, thereby, the wide possible distance between the hyperplane and the two categories are then used to predict the class label by mapping them into the same space based on which side of the hyperplane enclose to the new instances. The key strength of SVM is its robustness to fuzzy values and noise in the dataset as well as its ability to handle higher dimensional input spaces and smaller data sets, which could lead to an improvement in the generalisation performance of the classification. To elaborate further, SVM can deal efficiently with learning tasks where the number of training instances is smaller with respect to the number of features. This is because SVM usually selects a small number of support vectors during the learning process (Dubitzky et al., 2007). The typical limitation of this classification model is associated with the selection of the kernel function (Prajapati and Patle, 2010). The ultimate aim of SVM is to find the optimal hyperplane that maximises the generalisation ability of the trained model. However, if the training data is not linearly separable, the obtained classifier may not have high generalisation ability even though the hyperplanes are determined optimally. Thus, to enrich linear separability, the original input space is mapped into a high-dimensional dot-product space called the *feature space*. Hence, to avoid dealing directly with that high-dimensional space and to retain nearly the simplicity of separating hyperplane of SVM, different additional supported functions known as kernel are used. The Kernel function used with SVM can be linear or non-linear; examples for the nonlinear kernel are Quadratic and Gaussian (Das et al., 2018, Hamel, 2009).

2.13.1.4 K-Nearest Neighbour

The K- Nearest Neighbour (KNN) algorithm is simply based on the searching process to the proximity of k-samples between a test sample and other instances with similar behaviour (Cover and Hart, 1967). This searching process is guided by a distance function such as Euclidean, Camberra and Manhattan function. The distance metric plays a significant role to measure the distance between a test sample and the training samples. This metric should minimise the distance between samples belong to the same class label and maximise the distance between samples from the different class label. The final decision for the class label of a new test instance is made based on identifying the most frequent class label of k-nearest instances. This algorithm has achieved an efficient performance in solving different classification problem (Chen and Shah, 2018, Wu et al., 2018). However, it is highly sensitive to noise as well as to the choice of similarity metric used to compare instances (Kotsiantis et al., 2007).

2.13.1.5 Artificial Neural Networks

The artificial neural networks (ANNs) are efficient methods able to discover the intricate and non-linear relationship between the input features and the desired output. This method is inspired by the working way of biological nervous systems in the human brain and has been used widely and successfully in several applications such as diagnostic purposes, security systems, forecasting, pattern recognition and still wider. Many advantages can be achieved using ANNs including adaptive-learning, fault tolerance and parallelism strength (Deepa and Devi, 2011). However, a large number of training samples are generally recommended to achieve the most significant enhancement in classification accuracy (Kotsiantis et al., 2007, Sahiner et al., 2008). Different types of neural networks (NNs) are designed to solve different problems for various applications. Feedforward neural networks (FFNN) is the common type used successfully for pattern recognition, classification and object recognition (Birry, 2013).

FFNN is well-known, and popular model of ANNs used to address different problems in the diagnosis of various medical applications. It has been widely used due to its ability to sense the nonlinear and complicated relationship between inputs and outputs (Hwang and Hu, 2001, Jiang et al., 2010, Othman and Basri, 2011). The design of the NNs includes several layers. The first layer that receives information from the input features to be processed. The output layer where the outcomes of the processing are produced; and one or more layers in

between known as hidden layers. There is no neuron in the first layer while a different number of neurons can be assigned to the hidden layers whereas the number of neurons in output layers depends on the number of classes available in the input data. In the structure of FFNN, signals are allowed to travel one way only from the input layer to the output layer where all neurons are fully connected (Graupe, 2013). The network is trained on a set of data to produce input-output mapping. As a result, the weights of the connections of NNs are determined and then the network is used to demonstrate the classifications of a new set of data (Kotsiantis et al., 2007).

The signal is transferred within the network structure through an activation function assigned in each neuron in the NNs. The most common activation function used widely in the application of pattern recognition is a hyperbolic tangent function (Eq. 2.6) (Graupe, 2013, Lekutai, 1997). In opposed to other activation functions, the hyperbolic tangent function is more deferential (Özkan and Erbek, 2003, Negnevitsky, 2005) based on mapping the input signal into the non-linear, smooth and large scale from +1 to -1, where the +1 and -1 output values represent plus and minus infinity respectively (Figure 2.4). This enables the function to take into account both sides of the output single computed from the activation function without neglecting to the outcome produced from the negative part.

$$f(x) = tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \qquad 2.6$$
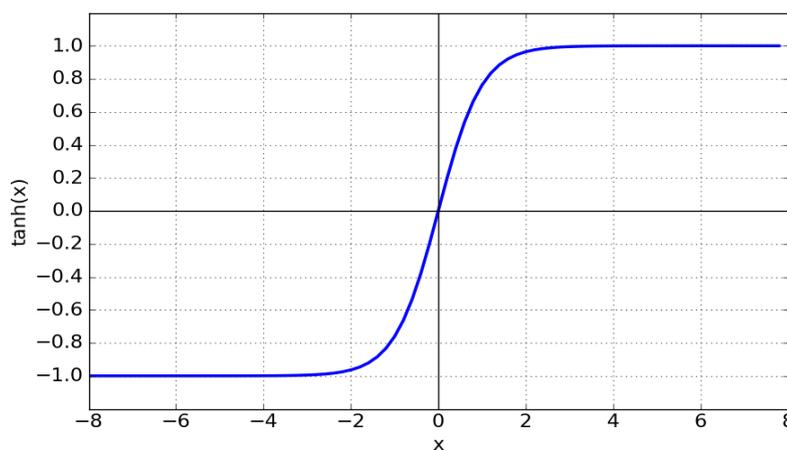
Figure 2.4 The hyperbolic tangent function used as an activation function for neural networks

The main aspects that determine the outcome behaviour of NNs are input and activation functions of the neurons, the weight of each input connection. The most popular and commonly used learning algorithm to predict the weights of the connections in the neural network is the backpropagation algorithm (Lekutai, 1997). This is based on iterative adjusting the weights of each neuron to lower the local error, minimising the errors between the produced output of the networks and target. It performs several weight modifications through iterative updating process before it reaches a proper weight configuration.

Practically, carefully determining the appropriate size of the hidden layer as well as the number of neurons per the layers is a significant challenge. This is because an overestimate neurons can lead to overfitting and eventually make the search for the global optimum more problematic, while an underestimate of the neurons number results in poor approximation and generalisation capabilities (Camargo and Yoneyama, 2001). It has been proven that using two hidden layers and a sufficient number of neurons in the hidden layer can approximate an arbitrarily complex mapping within a finite support (Hwang and Hu, 2001). Generally, the common practices to select the number of hidden layers and number of neurons in each hidden layer are investigated heuristically through performing several runs of NNs using a different number of neurons in the hidden layers.

2.13.2 An Overview of Convolutional Neural Network

Convolution neural network (CNN) is the most common deep learning approach dedicated for the classification and recognition for large-scale image datasets, and it is used widely in many applications, for examples music information retrieval (Han et al., 2017), remote-sensing image classification (Maggiori et al., 2017), and speech recognition and language processing (Qian et al., 2016, Abdel-Hamid et al., 2012, Swietojanski et al., 2014). CNN is designed based on three main layers namely the convolution layer, the pooling layer (subsampling layer) and fully connected layer (Karpathy, 2016).

This approach has the advantage of not needing to incorporate the feature extraction or selection process before being applied suggesting a strong point. However, training a CNN from scratch is time-consuming and challenging, as it needs a huge labelled dataset for training before the model is ready for classification, which is not always available particularly for medical data (Mohsen et al., 2018). Furthermore, it is a computationally expensive architecture and requires tuning a large number of parameters that need optimisation, lead to a high risk of overtraining particularly for a small sample size of data and subsequent low performance on data that are not utilised in the training process

(Chakraborty et al., 2019). It also requires advanced hardware for example, processing large number of filters for the large size of images (Litjens et al., 2017, Ravì et al., 2017). Furthermore, it is highly affected by class imbalance problem which is common in medical image datasets and it is necessary to combine a solution to this problem such as upsampling from the minority class (Buda et al., 2018), which can also lead to overfitting and low performance in the testing phase (Zhang et al., 2017). Accordingly, the implementation and application based on this approach are not feasible for such research problem undertaken in this thesis and effective results may or may not granted particularly for a small number of samples and limited hardware resources. Therefore, this approach is not investigated in this research work.

2.13.3 Multiple Classifier Systems

The best enhancement in the classification accuracy is not always guaranteed by using the single classifier. In the literature, many existing experimental works report the success of the multiple classifier systems in the classification task for various application domains (Oza and Tumer, 2008). Multiple classifier systems are based on a combination of more than one classification model. In developing the design of MCS, there are common questions that should be addressed carefully. Firstly, how to select the best set of classifiers, and secondly how to combine these classifiers. To answer these questions, it is essential to consider the classification accuracy as the most important criterion for selecting a base classifier. Other classifier abilities can also be considered such as the capability to handle noise and outliers. Additionally, the classifier should be sensitive to variations in the input data, training run or initialization (Hastie et al., 2001, Kuncheva, 2014). The other crucial aspect is to ensure that the ensemble members are not identical; that is, if the outcomes of the members are the same, then there will be no difference in the results compared to a single classifier approach. Therefore, it is necessary to ensure diversity of the classifier outcomes to achieve an improvement in the performance. More details are relayed in the following subsections.

2.13.3.1 Ensure Diversity

It is significant to ensure the diversity of the output decisions produced from different classifiers as this has an important impact on enhancing classification performance. It was reported that an ideal combination of classifier ensembles should have high diversity and low classification error (Kuncheva, 2014). In more details, the diversity can be generated by the management of either individual classifier inputs, outputs, or models (Giacinto et al.,

2000). The most popular approaches to ensure the diversity of classifier ensembles are feature subsampling and data resampling. Dividing the input data into different subsamples and using them to train classifiers individually leads to significant variance in the classifier outcomes. Bagging (Breiman, 1996), and boosting (Freund, 1995, Polikar, 2006) are the most common methods in data resampling. The Bagging technique determines subsampling with replacement to gain independent training datasets for each classifier, and a majority vote is used to obtain the final decision. Boosting on the other hand adjusts the input data distribution perceived by each classifier from the results of classifiers trained previously, and a weighted voting rule is applied to generate the final decision. However, boosting is more prone to overfitting the training data, which could reduce the output accuracy of data classification (Abdallah et al., 2018). Regarding feature subsampling, the Random Subspace Ensemble method, for example, is more adequate to a large number of features. It uses different subsets of features randomly sampled to train MCS members.

Ensuring the diversity of ensemble outcomes based on the subsampling of input data and features has a notable influence on improving classification accuracy. Nevertheless, it sacrifices part of either input instances or features mainly for a small number of input samples. Hence, this may have a negative impact on the learning phase, and it would truncate the chance to recognising unseen cases that could lead to reducing the classification quality. Therefore, to avoid this drawback, in this research work, and specifically for data with a small number of samples, the diversity ensemble outcomes is generated based on the manipulation of classification models' design that is performed based on utilising different setting and parameters of the ensemble members.

2.13.3.2 Combiner Design

The possible approaches to designing the fusion stage of MCS are as follows:

1. Non-trainable: An example of this category is the majority vote that is usually applied in many MCS. The correct class for a test sample is decided based on counting the vote for each class predicted by the base classifiers and selects the majority class. 50% of the vote +1 is generally used. Most of the ensemble method used this approach due to its simplicity and efficiency in the implementation.

2. Trainable: The basic and typical example of this approach is weighted voting.

3. Meta classifier: This category includes a two-stage learning phase. The first is constructed by the ensemble of multiple classifiers whereas the second stage treats output decisions of the ensemble members as inputs into a new machine learning algorithm.

The fusion development based on both majority voting or weighted voting is an efficient method being less complicated and easy to implement. With the merits of these approaches, it is possible to have a direct design of an ensemble of a considerable number of weak classifiers, without necessarily involving the most efficient classification model in the ensemble design. However, improving system performance is not guaranteed compared to a single classification model unless a heuristic evaluation is carried out.

Meta classifier approach is further advanced techniques of the ensemble methods and has shown promising results in different medical applications (Tsirogiannis et al., 2004). However, they require more complex design and the dimensionality of the output space rises rapidly with the number of classifiers and classes. Also, it is hard to identify or interpret the characteristics of the produced feature space and this is because the meta-combiner should be trained with a dataset different from the one used for the individual classifiers (Ponti Jr, 2011).

In conclusion, notably, each classification method based on single classifier system has a limitation which can lead to misclassification errors. However, patterns that are misclassified by different classifiers are not necessarily the same. Therefore, it is anticipated that the use of multiple classifiers can improve the decision about the patterns under classification. Utilising appropriate methods and techniques to fusion multiple classifiers minimises the overall effect of these errors and can overcome the drawback of weak classifiers and thus enhance the classification performance. Therefore, in this thesis, both single and multiple classifier systems will be examined and evaluated for the classification of malignant grades of glioma.

2.14 Multi-Class Classification

Developing a machine learning algorithm to accomplish multiclass classification poses a challenge compared to binary classification. Most classification models are designed for two-class problems and cannot be used in multiclass problems. Moreover, in some cases, they show insufficient efficiency and lower performance when applied to multiclass classification problems. Hence, it is more difficult to handle multiclass datasets than two-class problems (Zhou and Liu, 2006, Iram et al., 2014). Various methods were proposed in the literature for carrying out multiclass classification. In more detail, three major categories of methods were highlighted. The first group includes classifiers that can be extended to handle multiclass classification problems such as DT, NNs, SVM and k-NN. The second

category includes converting the multiclass classification problem into several binary classification problems. For instance, methods such as the one against all, one against one, and error correcting output coding (ECOC) can be used for multiclass problem conversion. The third group is illustrated by a hierarchical classification (HC) approach (Mehra and Gupta, 2013). Further details on the hierarchical method are explained as follows. HC was proposed to solve multiclass classification problems based on taking advantage of binary classification construction inside the tree of HC. This method was proposed by Kumar et al. (2002) and was called Binary Hierarchical Classifier. This approach uses M−1 binary classifiers to classify M-class problem. The binary classifiers are arranged in a tree-like structure with M leaf nodes, each corresponding to a given class (Aly, 2005). HC was reported as an efficient approach and used to classify different datasets and was shown to gain promising results compared to ECOC (Rajan and Ghosh, 2004). Similarly, (Chen et al., 2004) have used the HC and each node of the Hierarchical tree is based on SVM in which the obtained results show improved performance compared to bagged classifiers using remote sensing data. Accordingly, to deal with multi-class classification, this research work considers the hierarchical structure in the development of multi-class classification system for glioma grades.

2.15 Performance Evaluations

The final phase for a classification system is the evaluation of classification performance. Several techniques are employed for this purpose. The confusion matrix in general is the common evaluation source which provides several evaluation metrics including true positive (TP), false negative (FN), false positive (FP), true negative (TN), sensitivity or recall or true positive rate (TPR), specificity or true negative rate (TNR), classification accuracy or correct classification rate (CCR), positive predictive rate (PPR) or precision, negative predictive rate (NPR). These metrics are widely used to evaluate the classification performance in medical image based applications (Mahmoud-Ghoneim et al., 2003, Deepa and Devi, 2011, Das et al., 2018).

Determining the confusion matrix is a fundamental step in reporting the performance evaluation of a classification system (Figure *2.5*), whereby it is used to measure most common performance evaluation tools, a further detailed definition of these evaluation tools are as follows. TP means that the sample is originally positive, and the prediction system makes it true and classifies it as positive. FN means the sample is originally positive and the

prediction system makes it is false and classifies it as negative. FP means the sample is originally negative and the prediction system makes it false and classifies it as positive. TN means the sample is originally negative and the prediction system makes it true and classifies it as negative.

		Predicted Class	
		Class 0 (N)	Class 1 (P)
Actual Class	Class 0 (N)	TN	FP
	Class 1 (P)	FN	TP

Figure 2.5 Confusion matrix

To further elaborate, sensitivity measures the proportion of true (actual) positives which are correctly identified, for example, the percentage of sick people, who are correctly identified by the diagnostic system as having the disease (Deepa and Devi, 2011). It is measured by dividing the number of samples that are classified correctly using a prediction system over the number of samples involved in the experiment and taken from the same class. It is also known as the proportion of positives that are correctly classified; an actual positive rate with the probability of detection or recall defined by Eq. 2.7.

$$Sensitivity = \frac{TP}{TP + FN} \qquad 2.7$$

Whereas, specificity determines the proportion of negatives that are correctly recognised, for example, the percentage of healthy people, who are correctly identified by the diagnostic system as healthy samples. It also has the same value of sensitivity that is measured from the other class. For example, for two classes: class A and B, if the sensitivity is computed for class A, it can be considered as the specificity for class B and vice versa, the sensitivity of class B is considered as the specificity for class A. Specificity is known as a true negative rate which is defined by the Eq. 2.8.

$$Specificity = \frac{TN}{TN + FP} \qquad 2.8$$

The correct classification rate (CCR or ACC) measures the total number of samples correctly classified using a prediction system and is calculated from all classes involved in the experiment. It is computed by the summation of the total number of true positives and true

negatives divided by the total population used in the classification, which is defined by Eq. 2.9.

$$ACC = CCR = \frac{TP + TN}{TP + TN + FP + FN}$$
<div align="right">2.9</div>

Precision or positive predictive rates for class 0 and negative predictive rate for class 1, are defined as in Eqs. 2.10 and 2.11, respectively.

$$Postive\ Predictive\ rate\ (PPV) for(class\ 0) = \frac{TP}{TP + FP}$$
<div align="right">2.10</div>

$$Nagative\ Predictive\ rate\ (NPV) for(class\ 1) = \frac{TN}{TN + FN}$$
<div align="right">2.11</div>

F-metric is a popular evaluation metric, which is more appropriate to sense the difference in the system accuracy if there are unbalanced distributions of samples in the dataset. F-metric represents the harmonic average that trade-off between sensitivity and precision and it is defined by Eq. 2.12 (Bashir et al., 2016, Gu et al., 2009).

$$F - measure = 2 * \left[\frac{Sensitivity * Precision}{Sensitivity + Precision} \right]$$
<div align="right">2.12</div>

2.16 Recent Findings and Comparison

Different approaches are used for the classification of glioma grades and thus to facilitate the comparison of the reviewed literature mentioned early, the summarisation of the main elements of the studies that were developed to classify malignant grades and types of brain tumour is summarised in Table 2.2 and is discussed as follows:

Hsieh et al. (2017b) proposed a grading system to discriminate between grade IV and the lowest glioma grades (II, III). The MR images are segmented manually by an expert. They have used the combination of 2DGLCM and histogram features, which are extracted from T1c-weighted MR images. A filter approach is used to select the significant features. In the classification task, they have suggested that a logistic regression classifier outperforms neural networks. They evaluated the classification model using one dataset with 107 patients through the LOO cross-validation technique.

Table 2.2 Summary of the recent finding and comparison of different approaches dedicated for the classification of glioma grades

Ref.	Approach	Cross-validation	MRI Modality	Feature Extraction	Feature Selection	Classification Model	Classification types	No. of Patients	Accuracy %	Sensitivity %	Specificity %	Precision %
(Hsieh et al., 2017b)	Single classifier system	LOO	T1c	2DGLCM+ histogram	Filter	Logistic regression	Binary	34 glioblastomas and 73 lower-grade gliomas. Sensitivity here is for low grades	88	82	90	80
(Hsieh et al., 2017a)	Single classifier system	LOO	T1c	Texture-LBP	Filter + wrapper approach	Logistic regression	Binary	34 glioblastomas and 73 lower-grade gliomas. Sensitivity here is for low grades	93	97	92	85
(Subashini et al., 2016)	Single classifier system	Splitting the data	T2	Texture-GLCM shape and intensity	wrapper approach	Naïve Bayes	Binary	200 samples (100-low-grade, 100 high-grade)	91	-	-	-
(Zacharaki et al., 2009)	Single classifier system	LOO	T1, T1c, T2, FLAIR and perfusion	texture-Gabor, shape and intensity	Filter + wrapper approach	SVM	Binary	102 brain tumours, II (22), III (18), glioblastomas (34).	88	85	96	-

Table 2.2 continued

Ref.	Approach	Cross-validation	MRI Modality	Feature Extraction	Feature Selection	Classification Model	Classification types	No. of Patients	Accuracy %	Sensitivity %	Specificity %	Precision %
(Zacharaki et al., 2009)	Single classifier system	LOO	T1, T1c, T2, FLAIR and perfusion	texture-Gabor, shape and intensity	Filter + wrapper approach	SVM	Multi-class	102 brain tumours, II (22), III (18), glioblastomas (34).	62.50	-	-	-
(Khawaldeh et al., 2017)	Deep learning	Partition the data	FLAIR	-	-	CNN	Multi-class	109 subjects	91.16	92.25	-	91.79
(Ryu et al., 2014)	-	-	DWI, ADC	Texture-GLCM	-	t-test	Binary	40 patients (II(8), III(10), IV(22))	80	87.50	78.10	-
(Geneidi et al., 2015)	-	-	DTI, PWI, T2	Statistical MRI-features	-	t-test	Binary	15 (low-grade), 9 (high-grade)	100	100	100	100
(Inano et al., 2014)	Single classifier system	-	DTI	Statistical MRI-features	-	Kmean clustering	Binary	14 (low-grade), 19 (high-grade)	80.40	84.80	75.50	-

The classification performances in terms of classification accuracy, sensitivity, specificity, and precision are 88%, 82%, 90%, and 80% respectively. Similarly, Hsieh et al. (2017a) used the same methodology to differentiate between grade IV and the lowest glioma grades (II, III) while the differences are the use of local binary pattern (LBP) for features extraction and the combination of filter and wrapper methods for feature selection in attempting to enhance the classification accuracy of glioma grades. The results showed that the classification performances in terms of classification accuracy, sensitivity, specificity, and precision are 93%, 97%, 92% and 85% respectively.

Subashini et al. (2016) conducted a classification system for glioma grading into low-grade and high-grade tumours using 200 samples that includes 100 low-grade images and 100 high-grade images. They validated the classification model through training the model on 164 samples and then the trained model was used to test 36 samples. The combination of image texture features based on GLCM, shape and intensity features were used as an input features to the system. A comparison of three classifiers was performed and the Naïve Bayes classifier achieved the highest accuracy compared to the others at 91%.

Zacharaki et al. (2009) developed a classification system to determine different brain tumour types and grades. Several types of features including texture- Gabor, shape and intensity are extracted from manually segmented brain tumour images using the combination of conventional and advanced MRI namely T1, T1c, T2, FLAIR and perfusion. The combination of filter and wrapper approaches are utilised to select the best subset of features. For the classification task, the classification accuracy of SVM was found to achieve better results than both KNN and LDA classifiers. The classification performance to classify high-grades (III, IV) versus low-grade (II) in terms the classification accuracy, sensitivity, and specificity are 88%, 85% and 96% respectively. They also performed multi-class classification for glioma grades, the classification accuracy achieved for glioma grading into three grades (II, III, and IV) is 62.5%. The multi-class classification was conducted using the same framework mentioned above with using the one-versus-all strategy of binary classification and the majority vote scheme.

Khawaldeh et al. (2017) demonstrated deep learning approach based on CNN to classify brain images into health, low-grade and high-grade glioma using FLAIR-MR images. They used 109 subjects. The result showed a classification accuracy of 91.16%.

Other studies used a statistical approach based on student's t-test with threshold process to classify glioma into low and high grades. MRI Texture features based on GLCM were extracted from DWI and ADC modalities using 40 patients where the obtained results showed classification accuracy of 80% (Ryu et al., 2014). Similarly, MRI statistical features were derived from DTI and PWI modalities based on 24 patients where the results indicated a classification accuracy of 100%. However, the major limitations of this approach are as follows (i) they used an advanced MRI modalities that have limited availability in any MRI clinical centre, (ii) the conclusion is built on small sample size, (iii) various thresholds are required if different dataset is used, which reduces the generalisation of this approach to perform well using different brain tumour images.

Inano et al. (2014) applied an unsupervised classification approach to discriminate glioma grades into low versus high grades. They have extracted MRI statistical features from DTI using 14 samples of low-grades and 19 of high-grade glioma. K-mean clustering algorithm was used to enable unsupervised clustering of input features. The results have shown a classification accuracy of 80.40%.

Most of the recent studies are performed based on a single classification system due to its efficiency in achieving an objective and automated classification as well as the low sample size and complexity required in developing a procedure of the classification system design compared to the other approach such as the deep learning. However the single classification approach has high variation in the classification accuracy due to various sensitivities to input data distribution and it can behave differently if tested with a different dataset. Furthermore, improved classification accuracy may and may not be granted.

2.17 Conclusion

This chapter presents the definition of brain tumours with an emphasis on glioma. The common visual descriptors for different malignant grades of glioma were presented. This chapter also introduced the general basic representation of MRI modalities. The chapter also provides a comprehensive literature survey for the classification of malignant brain tumours in MRI images. This review was conducted in terms of feature extraction, selection and classification schemes. The malignancy assessment of brain tumours is generally a complex task. Therefore, many techniques were evaluated and discussed. This was conducted to select the most appropriate methods and techniques in term of determining an accurate and automated classification of the malignant brain tumours from MR images. Particularly, the

focus was given to the methods and approaches that are used to quantitatively classify glioma grades. The chapter also explored several predictors of malignant brain tumour, with a concentration on tumour heterogeneity, Necrosis, edema, enhanced and non-enhanced tumour. It besides discussed different classification approaches to solve the diagnostic problem of glioma grades. The survey also covered the effectiveness of using multiple classifier systems to improve the accuracy of the classification system further.

Based on the outcome of the literature review, the main limitations of the existing studies can be summarised, which have a significant impact on identifying the direction of this research work, as follows:

1- The traditional method to assess the malignant degree of brain tumours is mainly based on visual diagnosis and clinical analysis of multiple tumour descriptors such as tumour heterogeneity, the presence of necrosis and contrast enhancement. However, the malignancy assessment of a brain tumour based on the visual diagnosis is a complex task. This is due to the mixed visual characteristics of these descriptors among different grades of glioma, potentially leading to inaccurate diagnoses and misclassification. Moreover, the clinical confirmation in some cases requires biopsy or aggressive clinical surgery both of which is invasive and include many clinical complications. Indeed, less attention is given to the impact and usefulness of the quantitative measures of tumour descriptors including tumour necrosis, edema, non-enhancement and enhancement on the diagnosis of malignant brain tumours. Therefore, it is necessary to assess the importance of these descriptors on the classification accuracy of glioma grades. The objective analysis of these descriptors of a brain tumour is anticipated to enhance the quality of glioma grading in term of classification accuracy.

2- Some studies have significantly relied on using the combination of conventional MRI and advanced imaging modalities to gain further improvement in the classification accuracy of glioma grades. However, advanced MRI techniques, as opposed to conventional techniques, require more expensive equipment, more experience and relatively more time to extract tumour attributes. They also have limited availability in MRI clinics. Consequently, developing a classification system for glioma grades based only on conventional MRI modalities is of great interest for those who have only access to the conventional MRI techniques.

3- Many studies recommended two-dimensional textural analysis using GLCM due to their efficiency in representing the textural information of an image, which leads to promising results in different applications. On the other hand, other studies suggest that 3D analysis based on GLCM can lead to better classification results. Hence, there is, so far, no explicit clue of which analysis has the most impact in the classification of glioma grades.

4- MRI image patterns have a high correlation with each other and hence extracted features inducing high correlation and raising redundancy that could degrade the classification accuracy. However, to overcome this problem and eliminate the redundant features, the fast and efficient approach suggests the use of the relevance analysis between features and their corresponding targets, which is performed independently to classification outcome, ignoring the interaction among the features.

5- Many of the existing works for glioma grading is based on using a single classification approach, and very few investigated the advantage of MCS. However, to further improve the accuracy in the classification of glioma grades, developing an effective MCS has a significant impact on improving the classification accuracy for glioma grades

6- It is necessary with the application of backpropagation Neural Network to find the optimal convergence point that maximises the classification accuracy of NNs. Indeed, at present, no such method gives a general or standard solution to overcome this problem. Hence, to optimise the performance of NN, studies used few trials of NNs then track the accuracy results to report the highest one. However, both solutions suffer from a lack of generalisation. Furthermore, many existing studies have ignored the impact of varies initial weights or the merit of using different validation set on the overall performance of neural networks, and it is possible to produce an enormous range of different results for using the same NNs design by manipulating these two factors.

7- The existing works that have developed a hierarchical scheme to solve multi-class classification problem give less interest to the development of each node of the hierarchical strategy and its impact on the classification performance. However, some of these studies developed only a single classification approach or used different classifiers in different nodes of the hierarchical scheme. The development of these nodes on the other hand, with MCS, has received less attention.

To overcome the above-mentioned limitations, many methods and techniques are proposed to develop an objective and accurate classification of glioma grades in MRI images. Brief descriptions of the proposed methods throughout this study are listed below:

1. Comprehensive texture analysis is developed using 2D and 3DGLCM derived from MR images. The extracted textural features of both 2D and 3D have been examined to enhance the classification accuracy of glioma grades.

2. New features are proposed and investigated based on the objective analysis of different brain tumour descriptors, including Necrosis, Edema, non-enhancement and enhancement tumours. These features incorporated with different machine learning algorithms are used to develop the classification system for glioma grades.

3. A hybrid feature selection method based on the combination of filter and wrapper approaches are utilised and incorporated with different machine learning methods to guide the search process. This method is proposed to overcome degrading the classification accuracy due to the effect of the redundant features.

4. A meta-trainable ensemble approach is proposed and developed based on the development of two stages of learning in multiple classifier systems. Using Backpropagation NNs in the fusion stage incorporated with the proposed deep iteration neural networks (DINN) has a significant impact on improving the classification accuracy. The advantage of the proposed DINN is to optimise the performance of NNs, in a systematic way, achieving the optimal accuracy of NNs. The proposed meta-trainable ensemble approach improved the classification accuracy based on the integration of multiple classifiers and compensating the possible drawback that can occur due to weak classifiers.

5. A new hierarchical ensemble approach is proposed and developed to solve the multi-class classification problem (multi glioma grades) based on integrating the meta-trainable ensemble approach in each node of the hierarchical scheme; this proposed approach is named Hierarchical Meta-Trainable Multiple Classifier System (HMTMCS).

CHAPTER 3 : Texture Extraction, Selection and Classification

Overview

This chapter presents the implementation and results analysis of the extraction, selection and classification phases for the texture features extracted from MR images. It also presents the proposed hybrid features selection algorithm (HFSA), which was developed to enhance the classification of glioma grades. The proposed method has the merit of integrating the filter and the wrapper methods. This is based on using ANOVA as a filter method and ranking the feature space, incorporating the Pearson correlation and several machine learning algorithms to guide the selection process, which is updated by the outcome of the final classification accuracy of different classifiers. The main purpose of this chapter is to develop an automated classification system for glioma grades based on the objective analysis of the tumour heterogeneity. The other purpose of this chapter is to evaluate the ability of the proposed method to select the most efficient feature set and eliminate redundant ones. Thus, leading to further development of the classification system for glioma grades.

This chapter starts with the details relating to the preparation of the region(s) of interest (ROI) of brain tumour images using four datasets of MR images. This work also covers a comprehensive analysis of texture features extracted from ROI of brain tumour images using 2D and 3DGLCM. This chapter covers a demonstration of several experiments conducted to evaluate and examine the behaviour of the developed system based on the proposed method. Classification performance was analysed and evaluated by comparing the proposed method against ANOVA in terms of classification accuracy, sensitivity, precision, specificity and F-measure. Furthermore, the discrimination ability of the proposed method was evaluated by examining the final performance of the developed system using many common classifiers, including single classifier and ensemble approaches, which were trained and tested individually. The single classifier consists of different classification models namely DT, SVM, KNN and LDA. The ensemble approaches include Ensemble Subspace Discriminate Analysis (ESDA) and Ensemble Bagged of Decision Tree (EBTree).

3.1   Introduction

It is crucial to differentiate low-grade gliomas and high-grade gliomas preoperatively, as this impacts the prognosis and treatment of the patient who has brain cancer (Theeler and Groves, 2011, Siker et al., 2006, Lasocki et al., 2015). This motivates the development of a non-invasive, objective and automated system to determine the malignant grade of a brain tumour. To achieve this aim, a classification system based on machine learning is developed and new methods and techniques are proposed. This system will offer a reproducible and efficient method to automate and enhance the classification of a malignant brain tumour.

MRI is widely used for evaluating brain pathologic lesions because it is a common imaging technique and a safer medical imaging method (El-Dahshan et al., 2014, Larroza et al., 2016). Analysing the MRI morphological descriptors of brain tumours can support clinicians in making more objective and accurate decisions (Hsieh et al., 2017c). Both conventional and advanced MRI techniques are used for the identification of the malignancy level of brain tumours (Kono et al., 2001, Porto et al., 2014). However, advanced MRI techniques are limited in terms of their availability in MRI clinical centres and come with high costs for advanced equipment. Therefore, in this research work, the classification system is developed based only on conventional MRI methods to differentiate between different glioma grades.

Several conventional MRI modalities can be used to extract image features and utilise the classification of malignant brain tumours. For example, T1 modality with contrast enhancement (T1c-weighted) has been used to distinguish grade IV against the lower glioma grades (Hsieh et al., 2017b). However, this MRI modality is an invasive approach due to the involvement of the contrast agent. Also, the enhancement can be seen only in areas where the blood barrier inside the brain lesion has become permeable. Hence, it is highly dependent on the contrast leakage (Geneidi et al., 2015). In T2 modality, most of the brain tumours appear as hyper-intense compared to the surrounding parts. Thus the brain lesion is visually easier to identify and commonly used to conduct an initial assessment, identifying brain tumour types and differentiating non-tumour from tumours tissues (Tonarelli, 2013). T2-weighted is a non-invasive technique and is the most common MRI modality utilised for the segmentation and classification of brain tumour types and grades (Hasan and Meziane, 2016, Kharrat et al., 2010, Ananda Resmi and Thomas, 2010, Mohsen et al., 2018, Al-Waeli, 2017). The proposed classification system was therefore designed based only on T2-weighted MRI images.

Computer-aided diagnosis using the image features of brain tumours has been put forth as a significant approach in improving radiological diagnosis performance (Herlidou et al., 1999, Kassner and Thornhill, 2010, El-Dahshan et al., 2014). The malignancy of brain tumours can be predicted by the assessment of tumour heterogeneity (Ryu et al., 2014). Automated classification of different heterogeneity levels of brain tumours offers more objective and accurate decision-making than a human reader. Texture analysis of the surface patterns of a lesion is an important approach to measure tumour heterogeneity. Texture analysis is utilised widely and plays a key role in the identification and recognition of morphological characteristics of brain tumours (Nielsen et al., 2008, Holli et al., 2010, Roy et al., 2013, Mohan and Subashini, 2018). Among the different texture feature methods, the grey level co-occurrence matrix (GLCM) can significantly access the spatial distribution of image intensities and the local texture features leading to an efficient representation of image textural features and promising classification results. Furthermore, it has been commonly used in various applications in the classification of medical images (Yazdi et al., 2007, Hasan and Meziane, 2016, Kovalev and Kruggel, 2007, Bonilha et al., 2003, Wibmer et al., 2015, Subashini et al., 2016, Liu et al., 2018).

The motivations and contributions of this chapter are summarised as follows:

While two-dimensional textural analysis using GLCM is recommended by many research works and shown remarkable results in the evaluation of the malignancy level of brain tumours (Larroza et al., 2016, Nakagawa et al., 2018), other studies suggest that three dimensional-analyses based on GLCM can lead to better classification results (Chen et al., 2007, Chen et al., 2009, Sanghani et al., 2018). However, there is, so far, no explicit conclusion as to which one of these texture analyses has the best impact on the classification of glioma grades. Consequently, it is necessary to investigate a comprehensive analysis of the MRI-based 3D textural features, which can lead to achieve an optimised diagnosis of the accurate level of the malignancy growth of glioma grades. This creates the incitement towards the three-dimensional textural feature analysis based on GLCM, which could be an effective approach for the classification of glioma grades. This leads to the first contribution of this chapter, which is investigating a comprehensive 3D textural analysis based on GLCM incorporating different machine-learning algorithms for the classification of glioma grades in MR images (Al-Zurfi et al., 2019). The 3DGLCM matrix is mapped over all slices for each patient along the Z-dimension as well as the classic X- and Y-dimensions. A comparison of the classification performance based on 3D and 2D texture analysis was

conducted in terms of different evaluation matrices such as classification accuracy, sensitivity, specificity, precision, and F-measure. The texture analysis was also developed using the proposed method and was examined using several different classification models. Selecting the most efficient features is one of the main challenges to develop an efficient classification system, which is necessary to optimise the classification performance. For this concern, a filter approach can be used to select the crucial features. The filter method has been utilised in the classification of malignant brain tumours (Hsieh et al., 2017b), due to its simplicity and efficiency. However, the filter method could lead to limited classification accuracy because the selection process by this method has not taken into account the outcome of the classification stage. Other possible approaches to demonstrate the selection process are based on the wrapper method, which can achieve better accuracy (Subashini et al., 2016, Zacharaki et al., 2011). However, it is computationally an expensive approach. MRI image patterns have a high correlation with each other and therefore the extracted features induce high correlation. Hence, the features that have a high correlation with others and have less relevance can raise redundancy in the feature space (Hall, 1999, Al-Waeli, 2017) and can degrade the classification accuracy for glioma grades. Consequently, it is a significant challenge to select the optimal set of features without considering the issue of the interaction among features as well as take into account the classification outcome. This forms the ground of the second contribution of this chapter: proposing a hybrid feature selection algorithm that is able to capture the most crucial features from a wide range of features generated in this work. The proposed method has taken the merit of integrating the filter and wrapper approaches. The filter method was applied using the ANOVA technique. The wrapper approach was performed by incorporating different machine-learning algorithms where the search process is guided by the Pearson correlation and the outcome generated by using different subset of features and different classifiers.

3.2 Input Materials

Four MR image datasets that are publicly available were used to evaluate the proposed system for the classification of glioma grades. These datasets have a confirmation of histopathological diagnosis. The first three datasets are known as BRATSS2013 and BRATS2015 and BRATS2018; these provided with standard segmented MR images (Menze et al., 2015). The BRATS2013 dataset contains thirty patients, with low and high-grade histopathological diagnosed gliomas. The group of low-grade gliomas (I and II) includes ten patients. The second group of high-grade gliomas (III and IV) contain twenty patients. The

low-grade tumours are diagnosed as astrocytomas or oligoastrocytomas. The high-grade tumours are diagnosed as anaplastic astrocytomas or glioblastoma multiform tumours. This dataset was collected at four different centres: Debrecen University, Bern University, Heidelberg University and Massachusetts General Hospital, over the course of several years, using different MRI scanners with different field strengths (1.5T and 3T respectively). The BRATS2015 dataset includes 274 patients, covering 54 patients with low-grade gliomas (I, and II) and 220 patients with high-grade gliomas (III, and IV). The multimodal MRI data are available in these two datasets. For each patient, the FLAIR, T2, and T1 images were co-registered into the T1c data, which has the finest spatial resolution, and then all the images were resampled and interpolated into $1\times1\times1$ mm^3, with image dimensions of $240\times240\times155$ for all MR slices collected. The image file format and bit depth for BRATS2013 and BRATS2015 datasets are MHA format with 16 bits (Dong et al., 2017). The BRATS2018 dataset includes 285 patients, with 75 patients of low-grade gliomas (I, II) and 210 patients of high-grade glioma (I, II). The image file format and bit depth are NIFTI format with 16 bits (Bakas et al., 2017).

The fourth dataset includes three tumour grades of glioma (Clark et al., 2013). We have given the name 'Cancer dataset' to these MR images to distinguish it from the other datasets used in this work. This dataset is also publicly available and confirmed by the histopathological diagnosis. This collection contains ten patients of grade IV (Glioblastomas), ten patients of grade II, and ten patients of grade III. Each patient has a varying number of slices ranging from 20 to 120, with varying post imaging parameters such as different gap spaces and slice thicknesses, ranging from 2 to 7.5 mm. The image file format for this dataset is DICOM with 16 bits depth.

3.3   MRI Pre-Processing

The overall flow chart of the general stages of the classification system of glioma grades is shown in Figure 3.1. It starts with feeding T2-weighted MRI images into the classification system. Then the images are pre-processed to prepare them for the feature extraction stage, followed by the selection of the significant features and finally the process is ended by the classification and performance evaluation stages. In the classification stage, all samples in the dataset are passed through two phases of training and testing, where the classification performance is then evaluated based on the testing phase. The pre-processing is aiming to make the remaining stages more applicable. This includes the preparation of ROI of brain

tumours, cropping the image to keep only the ROI; the process is ended with the intensity normalisation. The pre-processing steps for MR images are further detailed in the following subsections.
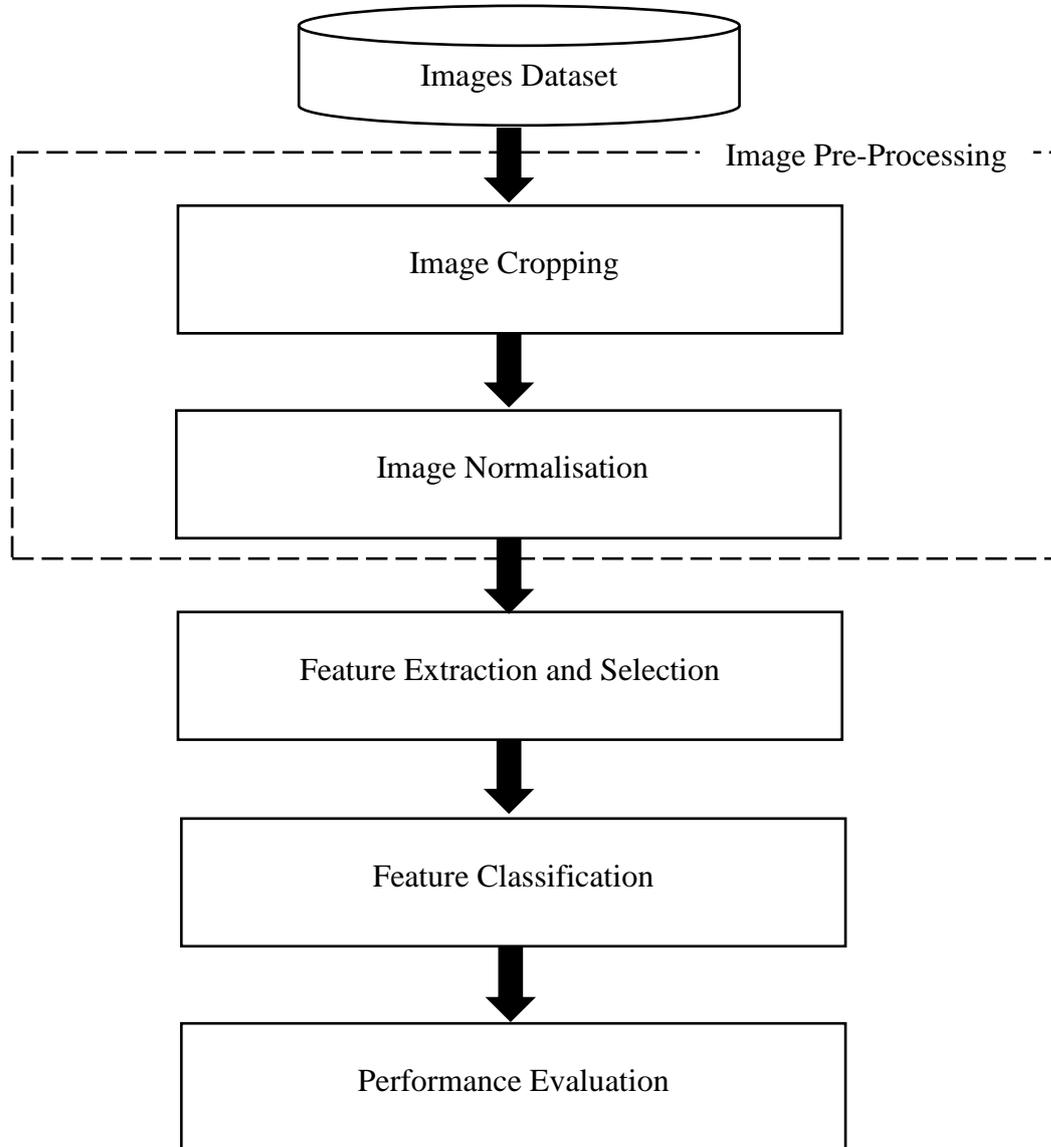
Figure 3.1 Overview of the general stages of the proposed classification system for glioma grades.

3.3.1   Preparation of Region of Interest of MR Images of Brain Tumour

The first step in the pre-processing of a classification system is to prepare the region of interest (ROI) of MR images of brain tumours carefully to avoid any distortion in the feature space; the distortion can occur if features from other regions out of the region of interest are involved in the feature extraction. Thus, extraction of features from only the ROI will lead to maintaining the quality of feature classification.

For the BRATS2013 and BRATS2015 datasets, standard segmented MR images are provided which are supported by a label identification layer. This layer has an index for the regions of the segmented tumour. This identification layer was used to identify the ROI of all MR images of brain tumours. Consequently, each patient in the dataset was represented by different numbers of MR slices, ranging from 20 to 60 MR images. These MR slices contain only the ROI of the brain tumour images where other MR slices without the presence of brain tumours are discarded.

For the Cancer dataset, which is different from the BRATS dataset, the provided images are not supported with segmented ROI. Therefore, to prepare the ROI of brain tumour images, it was necessary to apply a segmentation process. Since the segmentation task is out of the scope of this thesis, therefore to yield the ROI of brain tumour from MR images, the segmentation algorithm developed by Al-Waeli (2017) was used. This algorithm segments the ROI automatically from the MR images. It applied an automatic localisation of brain tumours using genetic algorithms based on bounding 3D-boxes (Hasan et al., 2016c). This algorithm relies on randomly creating hundreds of 3D-boxes with different locations and sizes in both the right- and left-brain hemispheres. The boxes in the right hemisphere are then compared to the corresponding 3D-boxes in the left-brain hemisphere using the objective function. This process is iterative and is based on the result of the objective function; these 3D-boxes are moved and updated toward the region that maximises the outcome of the objective function. The objective function is computed based on measuring the absolute value of subtracting the means of the intensities inside the produced 3D-box in the right-brain site from the corresponding 3D-box in the left-brain site. The objective function is thus computed between the two states (i.e. initial and next state). The value of the objective function is low when standing on soft tissue and high when the 3D-box stands on the lesion area because the tumour is always more hyper-intense than the surrounding soft tissue of the brain in T2-weighted images. Based on the recommended iteration range of the genetic algorithm (85 to 18 iterations) which is associated with the corresponding size

population ranging from 20 to 100 slices (Al-Waeli, 2017), and since the number of images in the obtained dataset ranged from 20 to 120, the maximum and minimum number of iterations were set to 85 and 18 iterations respectively. This was done to control the termination process of the genetic algorithm to search for the best optimal solution in locating the brain tumour in the MR images.

After the brain tumour is automatically localised, the next task is the segmentation of ROI of the brain tumour; this is performed automatically using three-dimensional active contours without edge (Hasan and Meziane, 2016). This algorithm is known as the Chan-Vese model; it can detect the object boundary not necessarily defined by the gradient, and it is independent of whether the boundaries are discontinuous or smooth. The parameters of this algorithm that have been evaluated and recommended to optimise the performance of this contour evolution are as follows: the length of penalty μ was set to 10^6, which enabled the algorithm to detect and segment the object accurately. The parameters $\lambda 1$, λ_2 control the competition force between the internal and external regions of the contour. Generally they hold the same values and usually $\lambda_1 = \lambda_2 = 1$, leading to a fair competition between these two regions (Nixon, 2008, Hasan et al., 2016a).

3.3.2 Image Cropping

This task involves removing the unnecessary parts out of the ROI of MR images of brain tumours. The purpose of this step is to avoid any redundant processing that can be consumed for other image parts outside the ROI. This can lead to reducing the computation time, which is considered as advantageous when developing an efficient classification system. This is performed based on eliminating zero background through an automatic cropping of each MR slice. The process of image cropping is conducted based on searching the image through four margins: top, bottom, right, and left, to produce a small window that has the ROI of the MR images of a brain tumour (Figure 3.2). At the same time, this process should not cause any reduction in the ROI of the tumour image. Therefore, to avoid potential loss in the tumour region, the dimensions of the produced window were assigned to be less than the largest presence of the tumour in all slices by one row and one column. In this reduction procedure, the pixel location and intensity were maintained, which is important for the feature extraction based on the GLCM. To elaborate, the construction of the GLCM is dependent on the pixel pair relationship that requires the pixel locations for the generation of these relations. The outputs of this procedure are slices with smaller dimensions compared

to the input images without loss of the tumour information. The original intensity of the MR slices and the pixel locations were not changed or transformed through this procedure. An example of the pre-processing steps were applied to the MR image, with the dimensions 216 by 176 pixels to obtain the ROI of the brain tumour, started by preparing the ROI of the MR slice (Figure 3.2). This was performed as based on the masking process, after which a slice with the same dimensions was produced and which had only the ROI of the brain tumour (Figure 3.2B). Then, the final stage involved cropping an MR image based on the movement of four margins in the produced image, this movement is designed by comparing each two neighbouring pixels in the x-axis to control the movement of left and right margins, and in the y-axis to control the movement of the top and bottom margins. The movement of the margins is stopped when both neighbour pixels are equal to zero, and finally an image with lower dimensions at 51 by 42 pixels is produced (Figure 3.2C).
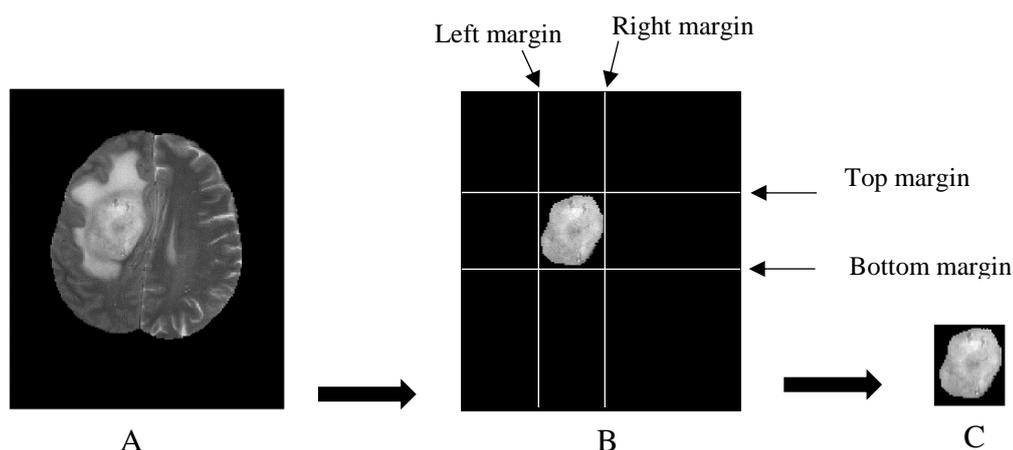
Figure 3.2 Preparation of region of interest of MR image of brain tumour and the cropping process of MR slice. A) Original MR slice with dimensions 216 by 176 pixels, B) Segmented tumour in MR slice with the same dimensions of input image C) Cropped MRI slice with dimension 51 by 42 pixels.

3.3.3   Intensity Normalisation

In the medical field, there is usually a wide variation of intensities in MR images. The reasons behind such variation are as follows: the image acquisition from different MRI scanners varies, the scanners can come from different manufacturers, there are different scanner models, and different models use different magnetic fields. Additionally, different acquisition settings of MRI units lead to variations in the intensities of MR images. Consequently, it is important for medical image analysis to demonstrate the normalisation

of the intensity of an image, to have a consistent range of intensity for all MR images involved in the feature extraction stage. MR image normalisation is useful for reducing the computation time for the analysis of images with a large range of intensities. The MRI image dataset obtained in this study comes from different MRI centres, scanners, with different magnetic fields; therefore, it was necessary to perform intensity normalisation. It is valuable in texture analysis to standardise the intensity range to eliminate dependence on an individual MRI setting which can disturb image contrast (Kjaer et al., 1995, Tantisatirapong, 2015). Image normalisation consists of adjusting the scale of the intensity of all images to produce a standardised range for all the MRI images.

Furthermore, Intensity normalisation is used widely in texture classifications (Hsieh et al., 2017b, Hsieh et al., 2017c, Hsieh et al., 2017a). Therefore, in this thesis, the image intensity for each T2-weighted image was normalised. The normalisation process is defined by Eq. 3.1 and 3.2 (Nyúl and Udupa, 1999, Loizou et al., 2009).

Let G be an input image while the normalized output image is K.

$$R_n = \frac{G_{in} - Gmin}{Gmax - Gmin}$$
3.1

$$K_n = R_n \times (Kmax - Kmin)$$
3.2

Where G_{in} is the input intensity of the input image G being considered, $Gmin$ and $Gmax$ are respectively the minimum and maximum intensities of G. Assuming the minimum intensity of the grey level is zero, leads to R_n representing the transformation ratio having values in the interval [0, 1], and n has a range from 1 to the total number of pixels of an image. For example, if an image has the dimensions of 256 by 256, n will be in the range [1- 56536]. In the normalisation process, the input intensities of an image are mapped from the range (*Gmax, Gmin*) into a new range (*Kmax*, *Kmin*), where *Kmax* and *Kmin* represent the maximum and minimum values of the normalised image K respectively, and K_n is the normalised image produced.

3.4 Texture Extraction

It was shown in the literature review that tumour heterogeneity is one of the most significant descriptors in assessing the malignancy degree of a tumour. This descriptor is widely used, and it has shown promising results in determining the grade of malignant brain tumours accurately. The heterogeneity of a brain tumour can be measured by analysing image texture

features to predict progression in the malignancy of tumour. It was also shown that GLCM can achieve remarkable results in measuring the texture features, leading to considerable classification results. The details of both 2D and 3D analyses of GLCM are explained in the following subsections:

3.4.1   Co-occurrence Matrix Based on 2D and 3D Analysis

GLCM investigates the relationship between a pixel pair in the region of interest of MR images of brain tumours. It measures the relative spatial information of a texture image for several directions and distances between pixel pairs. The main difference between 2D and 3D analysis for demonstrating GLCM lies in the number of directions for which θ is being considered (Chen et al., 2007). For 2DGLCM, for a certain distance d, four independent directions are considered corresponding to θ = 0∘, 45∘, 90∘, 135∘ (Figure 3.3); while for 3DGLCM, nine angles are constructed in addition to the four angles of the 2D data, producing a total of 13 angles for GLCM. The nine directions of the 3DGLCM (Figure 3.4) are as follows: θ = $(0^0, 45^0)$, $(0^0, 0^0)$, $(0^0, -45^0)$, $(45^0, 0^0)$, $(-45^0, 0^0)$, $(45^0, 45^0)$, $(-45^0, -45^0)$, $(45^0, -45^0)$ and $(-45^0, 45^0)$.

The construction of a three-dimensional GLCM is similar to that of a two-dimensional GLCM; both are designed by searching the probability of a pixel pair for a given distance and angle. The significant difference relies on the direction of the searching process. In the 2DGLCM the search process considers only the two dimensions of the x and y-axes through an image matrix. In the 3DGLCM, the search process also considers the third dimension. The third dimension of the 3DGLCM is built based on searching the probability of a pixel pair along the z-dimension for all MR slices that have a brain tumour. In 3DGLCM, the searching process investigates the relationship between a pixel in a reference slice and its neighbour in the next slice; this search includes nine angles (Figure 3.4). For example, if the neighbour in the next slice has the same coordinates, it is considered as angle 0^0. Texture analysis based on 3DGLCM was conducted in this thesis to add more information to the texture analysis of MR images of brain tumours.
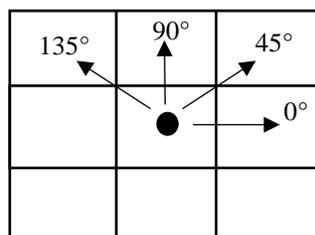
Figure 3.3 Two-dimension co-occurrence matrix generated with directions = 0°, 45°, 90°, and 135°, and distance = 1 for each reference pixel and its neighbouring pixels.
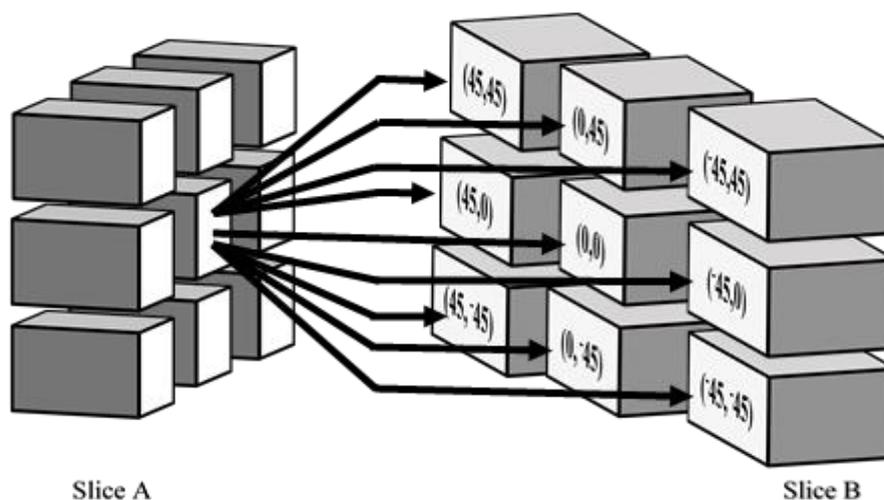
Slice A   Slice B

Figure 3.4 The relations between a pixel in a reference slice A and its neighbours in the next slice B for the Z-axes of 3DGLCM.

To select an adequate grey-level for texture extraction by GLCM, Gómez et al. (2012) have examined different quantisation grey-level, these are 8, 16, 32, 64, 128, and 256, and they have found that the quantisation levels do not improve or worsen the discrimination power of texture features. Although the GLCM is an efficient method for spatial texture extraction from grey images, it needs further computational time as more levels are included in the GLCM. Consequently, the common settings for the construction of a GLCM that were adopted in this research work are as follows; range of grey value that is used for grey level intensity is 0 to 255, and the distance is equal to one (Hsieh et al., 2017b, Hsieh et al., 2017c, Hsieh et al., 2017a, Kharrat et al., 2010). Thirteen angles with the application of GLCM were selected. These angles represent all possible directions of GLCM, which can be developed between a pixel and its neighbour pixels located in other slices. Furthermore, the texture features based on GLCM were extracted for all slices that have shown a brain tumour. The slices that have not presented any portion of brain tumour were discarded from processing stages.

For a given image, the co-occurrence P and its associated matrixes are defined as follows (Haralick et al., 1973): the probability co-occurrence is measured for the pair d and θ, where d and θ are the selected distance and direction respectively and defined by Eq. 3.3.

$$P=P \ (i, j \mid (d, \theta)) \qquad\qquad 3.3$$

where $P \ (i, j)$ is the $(i, j) \ th$ entry element of a normalised spatial probability matrix, which is obtained by measuring the ratio of each element in the spatial probability matrix divided by the total summation of all elements of the probability matrix, and i, j are the pixel coordinates in the P (i, j).

Different statistical predictors incorporated with the GLCM are used to represent the texture variability. These textural predictors have been used to measure image texture variance (Haralick et al., 1973). The ability of these predictors to classify different texture patterns is highly dependent on the complexity and nature of the grey tone transitions of an image. MR images of glioma are identified as presenting high heterogeneity between low and high grades. Therefore, statistical predictors were used to measure the variation in the texture, leading to discrimination between low- and high-grade gliomas. In general, for an inhomogeneous image, the co-occurrence matrix will have a large number of entries of small values, while for a homogeneous image, the matrix will have a small number of entries of a large value (Chen et al., 2007).

In this thesis, eighteen statistical predictors that represent the most common textural predictors were driven from the GLCM (Haralick et al., 1973, Gómez et al., 2012, Hasan and Meziane, 2016, Birry, 2013, Tantisatirapong, 2015). These predictors are as follows: autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of squares, sum average, sum variance, sum entropy, information measure of correlation 1, information measure of correlation 2, inverse difference normalised, and inverse difference moment normalised. These predictors are calculated for each of the co-occurrence matrixes for all angles Ɵ for each patient in the dataset. For 2DGLCM, eighteen predictors for each of the four co-occurrence matrices, which yield each patient represented by seventy-two features. For the implementation of the 3DGLCM, each patient is represented in total by eighteen predictors multiplied by thirteen directions, which produces two hundred thirty-four textural features. The definition and mathematical construction of the statistical textural predictors are described in APPENDIX B.

3.5 Proposed Hybrid Feature Selection Algorithm

Eliminating the redundant features can lead to an improvement in the classification performance. It also leads to reducing computational complexity by transforming high-dimensional data into a meaningful representation of a reduced one. A redundant feature is an attribute that is highly correlated with one or more of the other features so that the irrelevant predictor can be discarded without affecting the classification accuracy (Hall, 1999, Saeys et al., 2007). The proposed method aims to identify subsets of the most significant features affecting the accuracy of classification performance. To select the most efficient features, ANOVA technique can be used, which is an effective statistical method for detecting the significance level for each predictor in the feature space (Jafari and Azuaje, 2006). ANOVA is used to predict the significance of a predictor using P-values. The predictor that has a small P-value, less than the critical value that is being considered, will be significant and would thus be selected. For instance, if the P-value is less than 0.05 or 5%, the feature will be selected. This method is an efficient and fast approach. However, it measures the significance of each feature individually without considering the interactions between the predictors and ignoring the outcome of the classification. Therefore, to overcome this drawback, a hybrid feature selection algorithm (HFSA) was proposed that takes the advantages of ANOVA technique and Pearson correlation integrated with different classification models in an iterated search process, whereby the optimal classification accuracy and best set of features are achieved (Figure 3.5).

The automated system starts with the extraction of 3DGLCM from T2-weighted MR images of brain tumour using the full set of features. In the classification stage, different common classification algorithms were trained and tested individually based on the features extracted. After that, the classification performance was evaluated and analysed. The proposed hybrid features selection algorithm (HFSA) includes two main stages; the first one is the initialisation, and the second is the search algorithm in an iterated process, in which the best classification accuracy and best set of features are selected.

Further details of the proposed algorithm are as follows; at the initialisation, ANOVA was herein applied to all features, the features that have more than the P-value (0.5) were eliminated. In this stage also, the feature space is ranked from the lowest to the highest correlated against the reference features. The reference feature represents the one that has highest classification accuracy compared to all other predictors. The Pearson Correlation method was used to assess the correlation between features (Chan, 2003, Swinscow and

Campbell, 2002), and can take values within the range of the three values, -1, 0 and 1, where 0 indicates there is no correlation and 1 or -1 point to a higher correlation.
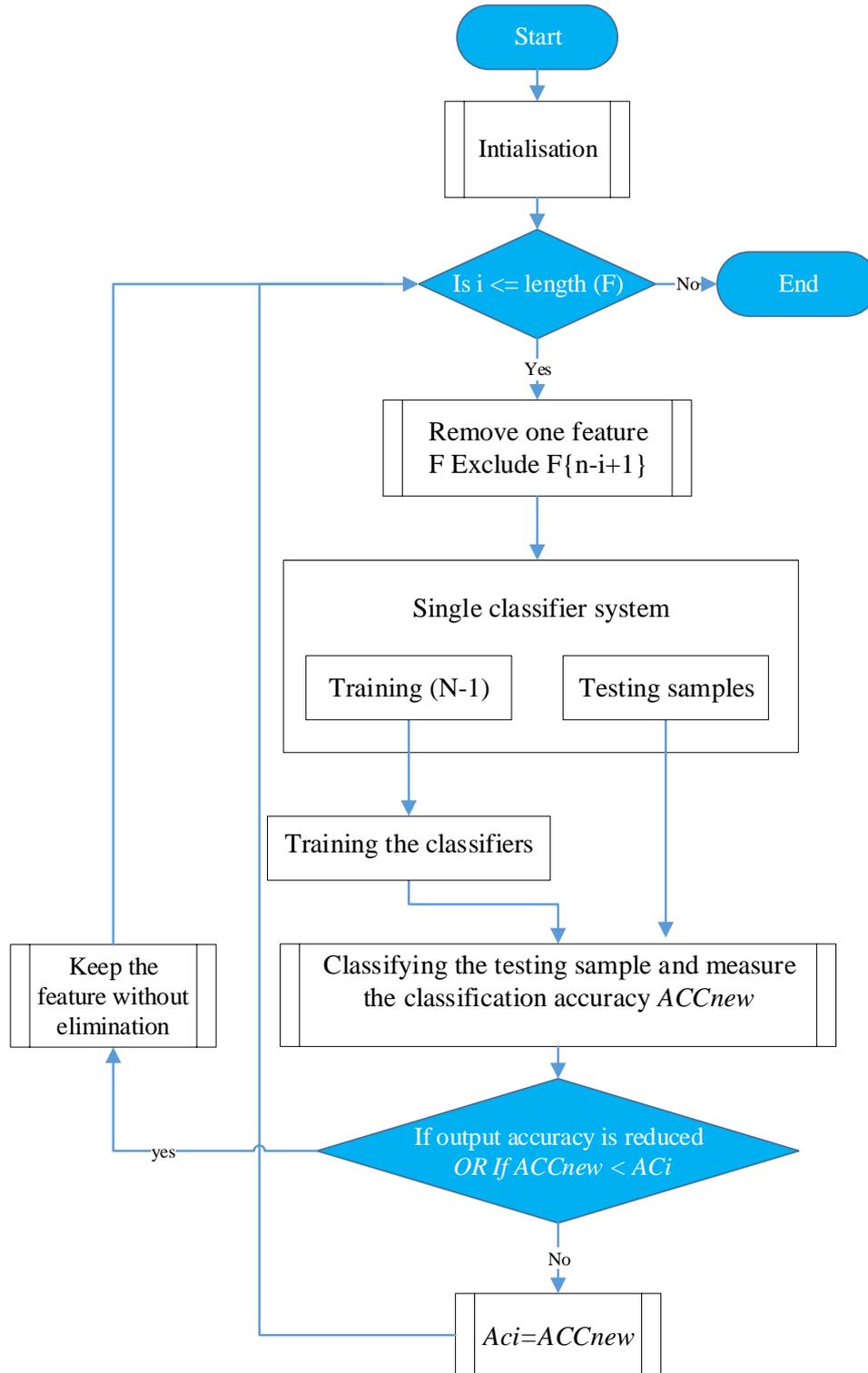
Figure 3.5 The flow chart of the proposed hybrid features selection algorithm based on machine learning algorithm developed for the automated classification of glioma grades.

73

The Pearson correlation was proposed in combination with the ANOVA method to remove the redundant features and select the best set of features based on measuring the outcome of the interactions among features. The output of this stage is the ranked features F in ascending order based on the measurement of Pearson correlation between these features and the reference predictor, thereafter the feature set F is fed to the iteration process.

Let i=1. Let a classifier set $\Psi = (\Psi_1,..., \Psi_N)$, N is the total number of classifiers, feature set $F = (F_1, F_2, \ldots , F_n)$; n is the total number of predictors. Let Aci represent the initial classification accuracy, which set experimentally based on the result obtained from the first run to all classification models and F.

The loop process is started using the feature F to find the maximum possible accuracy. The proposed method uses an iterative process of decrementing the feature space F by one feature to generate new feature set M and then performing training and testing for all classifiers Ψ in a comparable procedure to choose the best classification accuracy (Eq. 3.4 and Eq. 3.5).

$$ACC = \frac{1}{S} \sum_{k=1}^{S} P(correct| (w1, w2))$$ 3.4

where P is the probability of correct prediction; w_1, w_2 refers to class one and two respectively, S is the total number of samples of w_1, and w_2.

Maximum classification accuracy=arg MAX (ACC (Ψ, M))   3.5

The criteria that is used to select the best set of features is as follows; the process of eliminating features starts from the highest correlation to the lowest one, then the produced feature set is used to train and test different classifiers. Consequently, if the result of the classification accuracy based the generated subset of features is less than the previous state (*Aci*). In other words, if the classification accuracy is reduced compared to the previous states then that feature is inserted back and kept in the features set for the next examination. Otherwise if the resultant output accuracy is the same as the maximum previous state or increased then the elimination process is continued, and the initial accuracy is replaced with the new accuracy. This process is iterated and repeated until all features are examined and the classification accuracy is investigated. The final output represents the best selected feature subset that has the optimal classification accuracy.

3.6   Features Classification Using Single and Ensemble Classifiers

Features classification is a key step in the automation of the diagnostic system. It refers to the categorisation of objects into classes. It includes two phases of processing: training and testing. In the initial training phase, a description of each classification category based on image features is determined and utilised to train classification model. In the testing phase, the trained model is used to classify unknown samples.

A comprehensive analysis of MRI textural features was accomplished to evaluate the impact of both 2D and 3D textures in classifying glioma grades. The features extracted from both 2D and 3D data of GLCM were used to train different common machine learning algorithms, namely Linear Discriminate Analysis (LDA), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Decision Tree (DT), EBTree, and ESDA. Thereafter, the testing phase, based on leave-one-out (LOO) was determined. Several decision choices and techniques were applied within these classification methods to generate additional classifiers (Table 3.1). Three main justifications for these methods and design choices are as follows; first they are the most commonly used, which lead to promising results, and therefore they are selected as a default configuration in the Matlab learner application (Abdallah et al., 2018, Ashour et al., 2018a, Das et al., 2018, Al-Waeli, 2017). Furthermore, the selected classification models are the most suitable to handle both small and large datasets, which can lead to successful results in the decision-making process (Kuncheva, 2014). Moreover, these techniques and choices are demonstrated in this work to ensure the diversity in the output decisions of the classification models whereby robust evaluation and objective comparison can be established. Also, approaching the diversity of the output decisions based on these predefined design choices are further utilised to enhance the quality of the classification performance through the development of the classification system for glioma grades, further details are explained in section 5.1. The textural features produced were prepared for the classification stage by performing features normalisation, mapping these features into new forms, which were more suitable for the classification process. This step is vital in avoiding features with large ranges outweighing those with smaller ones. Linear normalisation was performed to standardise the extracted features while maintaining the same relationships between the original features (Georgiadis et al., 2009).

Table 3.1 The parameters settings for the classifiers undertaken. These include four main single classifiers namely DT, LDA, SVM, KNN, and two Ensemble methods. In total, thirteen classifiers are generated using these different parameters and techniques.

Classifier Name	Parameters
DT	Maximum number of split is 4. Split criterion is Gini's Diversity Index. Maximum surrogate per node is 10.
LDA	Regularisation is diagonal covariance.
SVML	SVM with Linear kernel function.
SVMQ	SVM with Quadratic kernel function.
SVMCUB	SVM with Cubic kernel function.
SVMG	SVM with Gaussian kernel function
KNNF	KNN where the number of k-neighbours is 1, distance metric is Euclidean, and the distance weight is identical.
KNNM	KNN where the number of k-neighbours is 10, distance metric is Euclidean, and the distance weight is identical.
KNNCOS	KNN where the number of k-neighbours is 10, distance metric is cosine, and the distance weight is identical.
KNNCUB	KNN where the number of k-neighbours is 10, distance metric is cubic, and the distance weight is identical.
KNNW	KNN where the number of k-neighbours is 10, distance metric is Euclidean, and the distance weight is squared inverse.
EBTree	Ensemble classification model where Bagging strategy is used, thirty learners of DT as a base classifier are used.
ESDA	Ensemble classification model where feature subspace strategy is used, thirty learners of discriminate analysis as a base classifier are used.

All the classifiers were validated using the leave-one-out cross-validation method, which is essential for avoiding overfitting problem. The ultimate purpose of this stage was to select the best classifier that achieved better accuracy and thus develop a classification model that can classify unseen new data correctly.

3.6.1 Ensemble Bagged Tree

Ensemble Bagged Decision Tree (EBTree) is based on bootstrap aggregation, manipulating training data (Abdallah et al., 2018, Kuncheva, 2014). The training subset is bootstrapped (resampled randomly with a replacement) to generate a different training subset. The manipulation involves splitting the input instances (in the training phase), then feeding them as inputs to the classifiers. In this experiment, thirty learners of DT as a base classifier are used and a majority vote (50% +1 rule) as a combination strategy was conducted (Breiman, 1996).

3.6.2 Ensemble Subspace Discriminate Analysis

Ensemble Subspace Discriminate Analysis (ESDA) uses different feature subsets to train the members of the ensemble and has been reported as being an efficient method in many application domains characterised by high-dimensional features (Tin Kam, 1998, Ashour et al., 2018a). Examples include data classification and cancer diagnosis (Bertoni et al., 2004, Bertoni et al., 2005, Armano et al., 2011). This method involves randomly discriminating the dimension of features space into several different subsets of features, these subsets of the features are used to train and test the members of the ensemble, and finally the decision is made by majority vote. A subspace dimension of the feature subset can be adjusted to select the best set of features that improve the classification accuracy. Thirty learners of discriminate analysis as a base classifier were used and a majority vote was used to combine the base classifiers.

3.7 Experimental Results and Discussion

The hybrid features selection method was proposed to find and identify an optimal subset of features that can improve the performance of the classification system. A comprehensive evaluation of the results in terms of classification accuracy, sensitivity, specificity, precision, and F-measure for both the 2D and 3DGLCM was conducted. The purpose of this work is to examine the quality of the proposed hybrid feature selection method integrated with different classification methods to achieve an accurate classification of glioma grades in MR

images. This work will start with a comparison of the results obtained from training and testing different classification method using 2D against 3DGLCM. Thereafter, a comprehensive analysis of the 3DGLCM will be conducted corresponding to different orientations of the 3DGLCM to show the behaviour of the features associated with each angle in the classification of glioma grades. These results will highlight the need to use the proposed features selection method to select the essential features that can achieve a significant improvement in classification performance.

3.7.1   Results Comparison between 2D and 3D Analysis of GLCM

In this experiment, BRATS2013 dataset set was used to compare both behaviours of the 2D and 3DGLCM. The obtained results show that the maximum classification accuracy was achieved by SVML classifier at 93.3% for both 2D and 3D analysis of GLCM (Table 3.2 and Table 3.3). The strategy that was used in extracting 2DGLCM was based on using standard segmented datasets of volumetric data and by considering all the slices that include the ROI of the brain tumour enables the classification method based on 2DGLCM to show high results. On the other hand, it was noticed that no improvement was achieved in classification accuracy using 3DGLCM compared to 2DGLCM, where the same results of classification accuracy were maintained at 93.3% by SVML classifier (Table 3.3). It is worthwhile to note that using the full set of features associated with 3DGLCM could lead to increase the redundancy between features and this could affect negatively the classification accuracy. Therefore, to gain a clearer understanding, a comprehensive analysis of 3DGLCM in terms of thirteen angles was performed. This was based on discriminating the feature space of 3DGLCM based on the angles into the 2D analysis including the four angles (0^0, 45^0, 90^0, and 135^0), and third dimension of the 3D analysis that cover the nine angles (0^0, 45^0), (0^0, 0^0), (0^0, -45^0), (45^0, 0^0), (-45^0, 0^0), (45^0, 45^0), (-45^0, -45^0), (45^0, -45^0) and (-45^0, 45^0). Then, they were individually used to train, and test several classification methods. Classification performance was analysed in term of the results of the confusion matrices for different classifiers trained and tested individually using 2D and 3D of GLCM respectively (Table 3.2 and Table 3.3).

Table 3.2 Comparative results of different classifiers incorporated with the full set of features associated with 2DGLCM using BRTAS2013 dataset. Class1 and Class0 refer to high and low grades respectively.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
LDA	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
SVML	Class0	9	1	90.00	90.00	90.00	93.33
	Class1	1	19	95.00	95.00	95.00	
SVMQ	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
SVMCUB	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
SVMG	Class0	8	2	80.00	88.90	84.21	90.00
	Class1	1	19	95.00	90.50	92.68	
KNNF	Class0	7	3	70.00	70.00	70.00	80.00
	Class1	3	17	85.00	85.00	85.00	
KNNM	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
KNNCOS	Class0	9	1	90.00	75.00	81.81	86.67
	Class1	3	17	85.00	94.40	89.47	
KNNCUB	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
KNNW	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
EBTree	Class0	7	3	70.00	63.60	66.66	76.67
	Class1	4	16	80.00	84.20	82.05	
ESDA	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	

Table 3.3 Comparative results of different classifiers incorporated with the full set of
features associated with 3DGLCM using BRTAS2013 dataset.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
DT	Class0	6	4	60.00	66.70	63.15	76.67
	Class1	3	17	85.00	81.00	82.92	
LDA	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
SVML	Class0	9	1	90.00	90.00	90.00	**93.33**
	Class1	1	19	95.00	95.00	95.00	
SVMQ	Class0	6	4	60.00	66.70	63.15	76.67
	Class1	3	17	85.00	81.00	82.92	
SVMCUB	Class0	7	3	70.00	70.00	70.00	80.00
	Class1	3	17	85.00	85.00	85.00	
SVMG	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
KNNF	Class0	7	3	70.00	70.00	70.00	80.00
	Class1	3	17	85.00	85.00	85.00	
KNNM	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
KNNCOS	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
KNNCUB	Class0	7	3	70.00	77.80	73.68	83.33
	Class1	2	18	90.00	85.70	87.80	
KNNW	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
EBTree	Class0	8	2	80.00	66.70	72.72	80.00
	Class1	4	16	80.00	88.90	84.21	
ESDA	Class0	4	6	40.00	33.30	36.36	53.33
	Class1	8	12	60.00	66.70	63.15	

For the 2D analysis of GLCM in term of the four angles, it was found that the highest classification accuracy was achieved at 93.3% by the classifier SVML and DT by the angles 45^0, and 135^0 respectively (Figure 3.6), followed by 90% that was achieved by both KNNF, and KNNCUB classifier at the orientations 0^0, and 90^0 respectively. For the behaviour of 3D analysis of GLCM in term of the classification accuracy (Figure 3.7) for the all the nine angles of GLCM, the obtained results show that the maximum classification accuracy was achieved by DT at 96.7% by the angle $(0^0, 45^0)$ outperforming all other classifiers, followed by 93.3% by the same classifier at the direction $(-45^0, 45^0)$. The SVM model also achieved its highest accuracy at 90% at both angles $(0^0, 45^0)$, and $(-45^0, 0^0)$. Similarly, KNN achieved its highest accuracy at 86.7% for all the angles $(0^0, 45^0)$, $(-45^0, -45^0)$, $(-45^0, 45^0)$, $(45^0, -45^0)$, and $(45^0, 45^0)$ (Figure 3.7). Classification performance in terms of sensitivity of the high-grade gliomas corresponding to the GLCM angles (Figure 3.8, and Figure 3.9) and specificity (or sensitivity of low grades) (Figure 3.10, and Figure 3.11) were also measured, the results indicated that the angles of both 2D and 3DGLCM showed various behaviours, where the best sensitivity of high grades was achieved at 100% by both the DT and ESDA classifiers at the orientation $(0^0, 45^0)$ (Figure 3.9). While for specificity, the maximum results were achieved at 100% by KNNM, KNNCOS, and KNNCUB classifiers at the angle 90^0 (Figure 3.10). Ultimately, the results of the analyses of both 2D and 3D show the effects of applying different orientations on classification performance. Also, the results show the relevant angle that can achieve the best diagnosis quality of glioma grading. Significantly, it was noted from results that different orientations of 2D and 3DGLCM show different behaviours for the classification of glioma grades. This is due to the significant variance of the spatial arrangements of tumour patterns in MRI images by using different texture angles. It is seen that there is a significant difference in the behaviours while using the full set of features of 3DGLCM and the outcome of individual angles, where the highest accuracy of the former was only 93.3% while the obtained results of latter were 96.7% by the angle $(0^0, 45^0)$. Hence, this is a notable indication that redundancy generated in features when the features are combined in one set of features, and this causes a relative reduction in classification accuracy. Therefore, to tackle this problem and remove these problematic features, in an attempt to improving the classification accuracy, the proposed hybrid feature selection method was demonstrated and evaluated. To obtain a wide range of evaluations for the proposed method, four public MRI datasets were used to investigate the ability of the proposed method in selecting the best set of features that leads to improved classification accuracy for glioma grades.
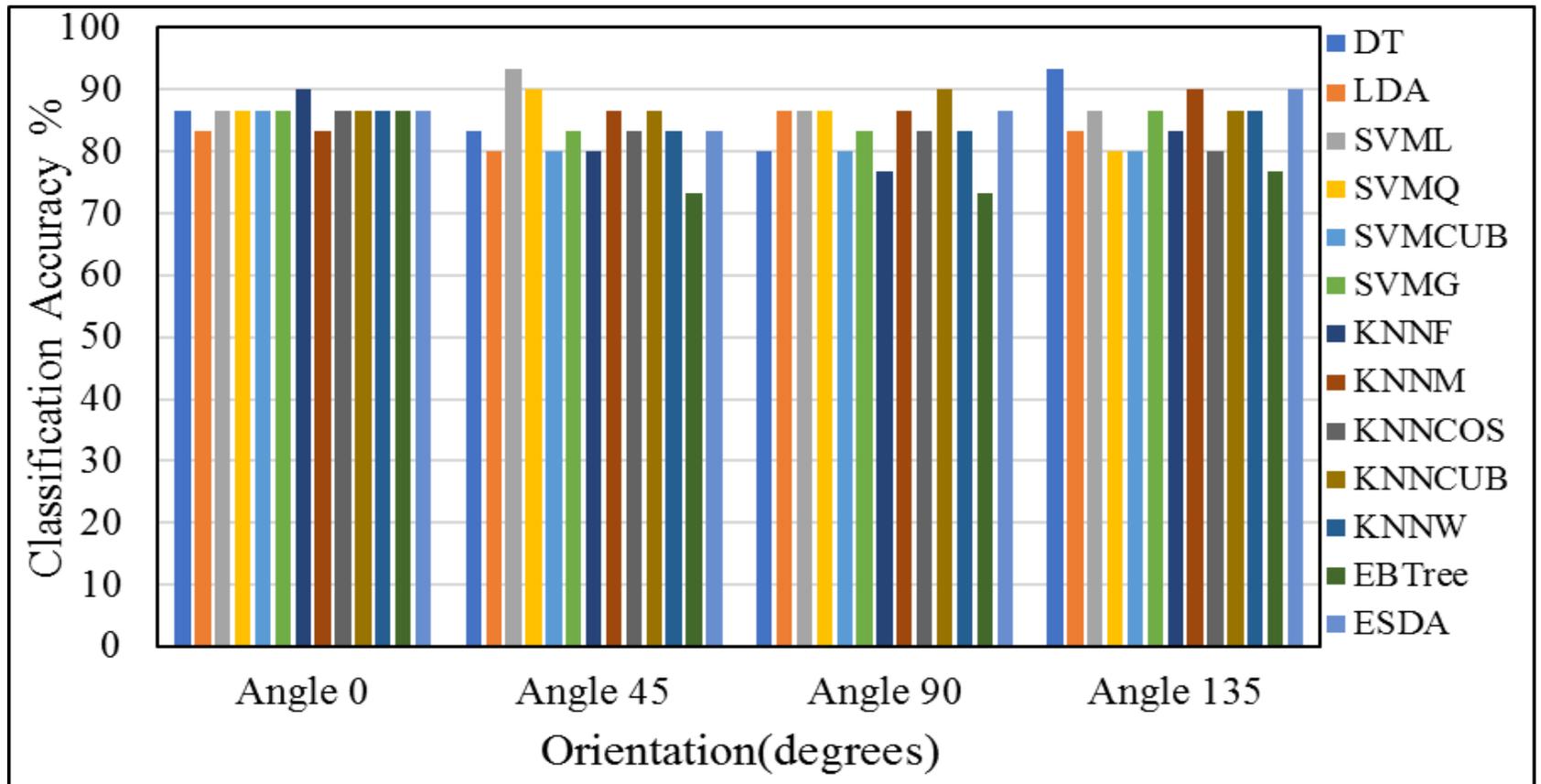
Figure 3.6 The behaviour of the four orientations of 2DGLCM application investigated with different classifiers in term of classification accuracy for BRATS2013.
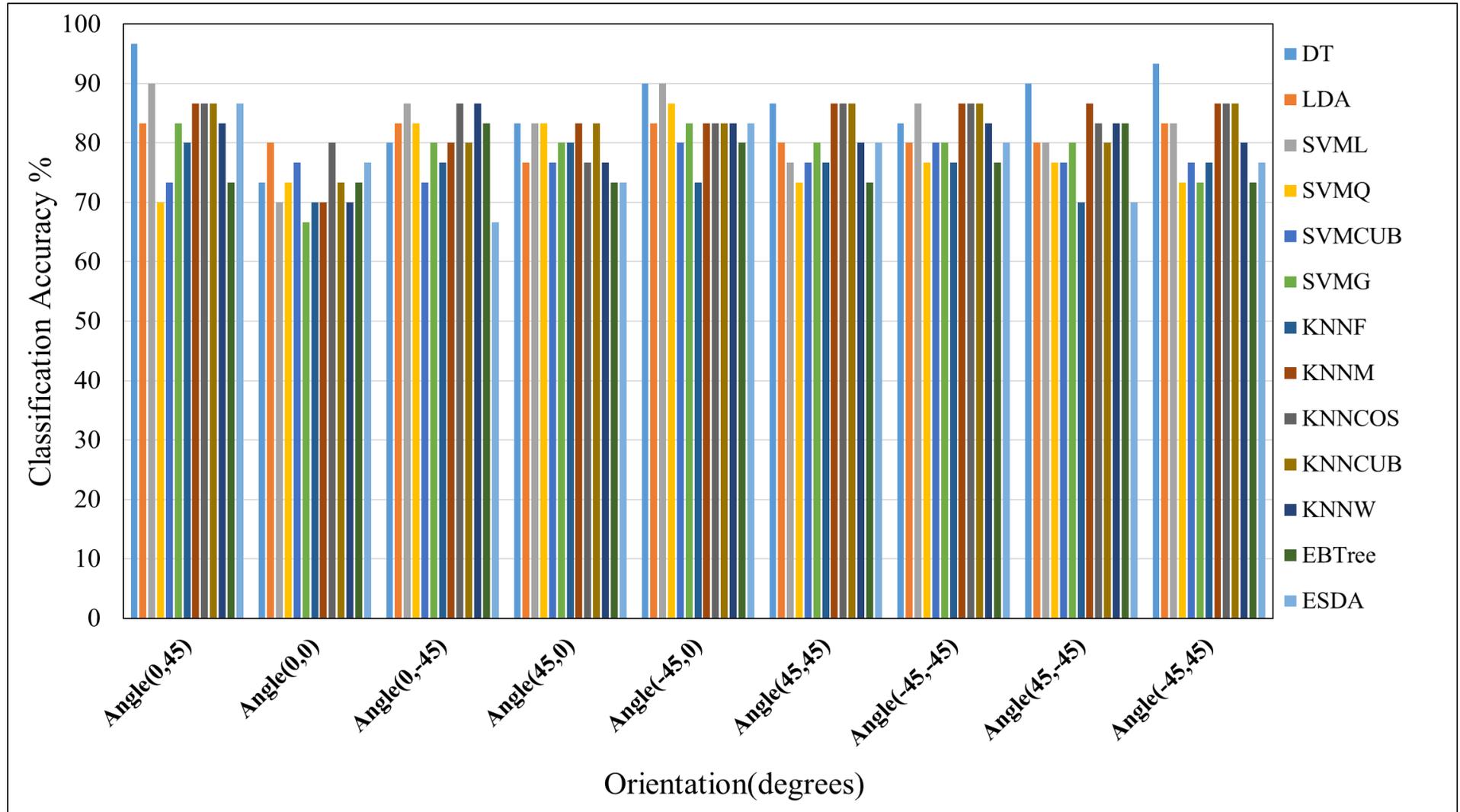
Figure 3.7 The behaviour of the nine angles of the third dimension of 3DGLCM application investigated with different classifiers in term of classification accuracy for BRATS2013.
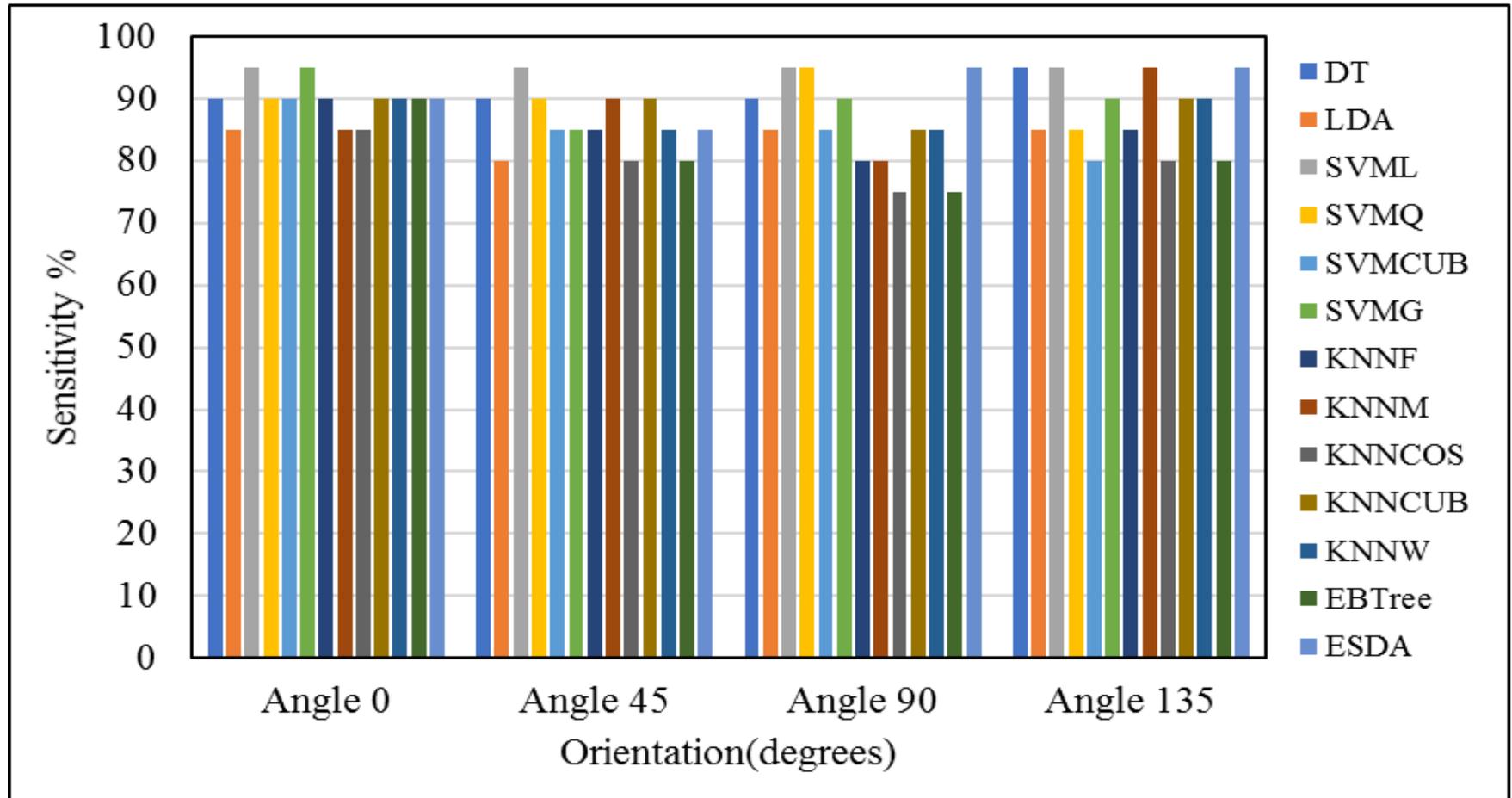
Figure 3.8 The behaviour of the four orientations of 2DGLCM application investigated with different classifiers in term of sensitivity of high-grades glioma for BRATS2013 dataset
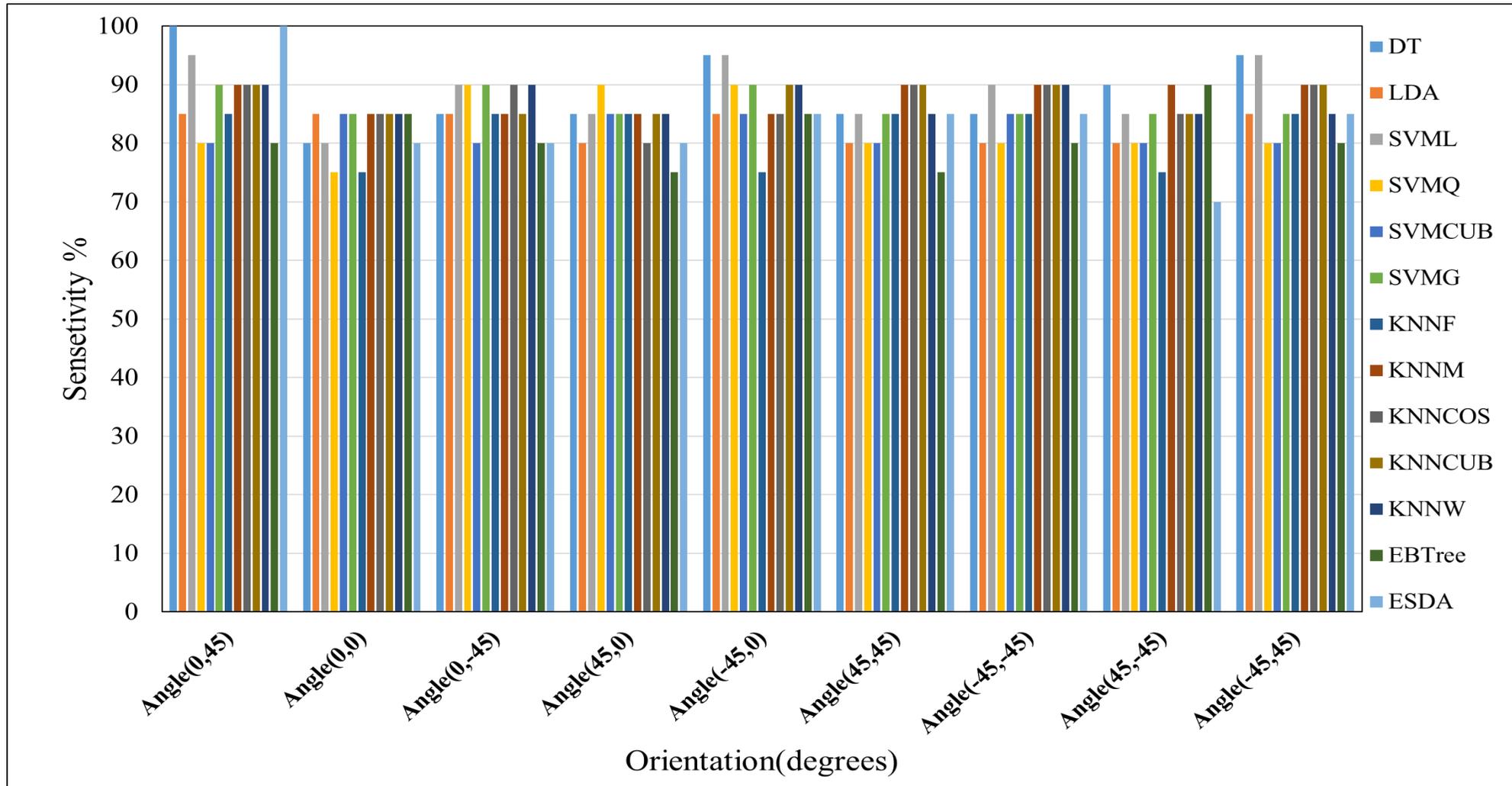
Figure 3.9 The behaviour of the nine angles of the third dimension of 3DGLCM application investigated with different classifiers in term of sensitivity of high-grades glioma for BRATS2013 dataset.
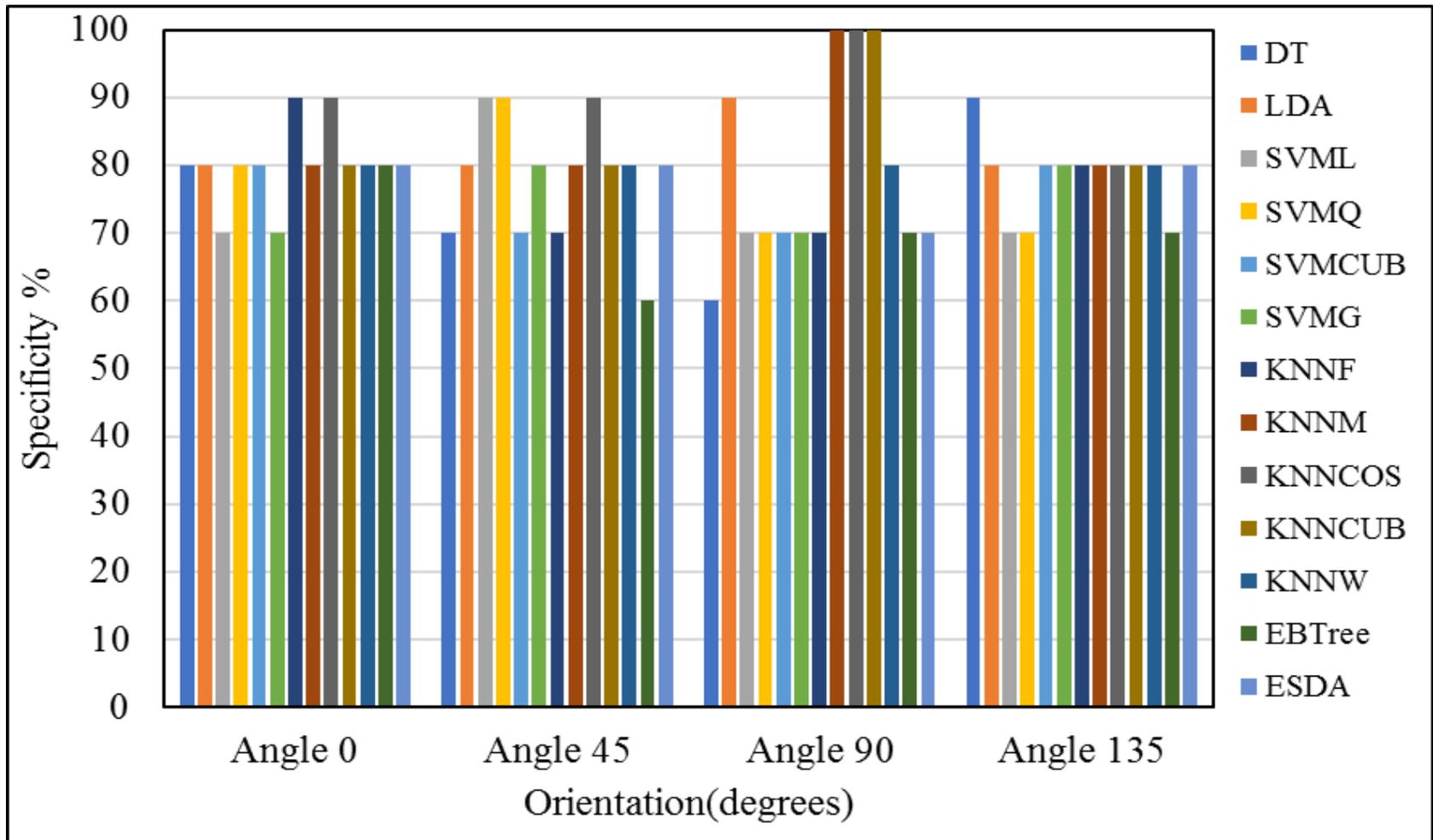
Figure 3.10 The behaviour of the four orientations of 2DGLCM application investigated with different classifiers in term of specificity (sensitivity of low-grades glioma) for BRATS2013 dataset
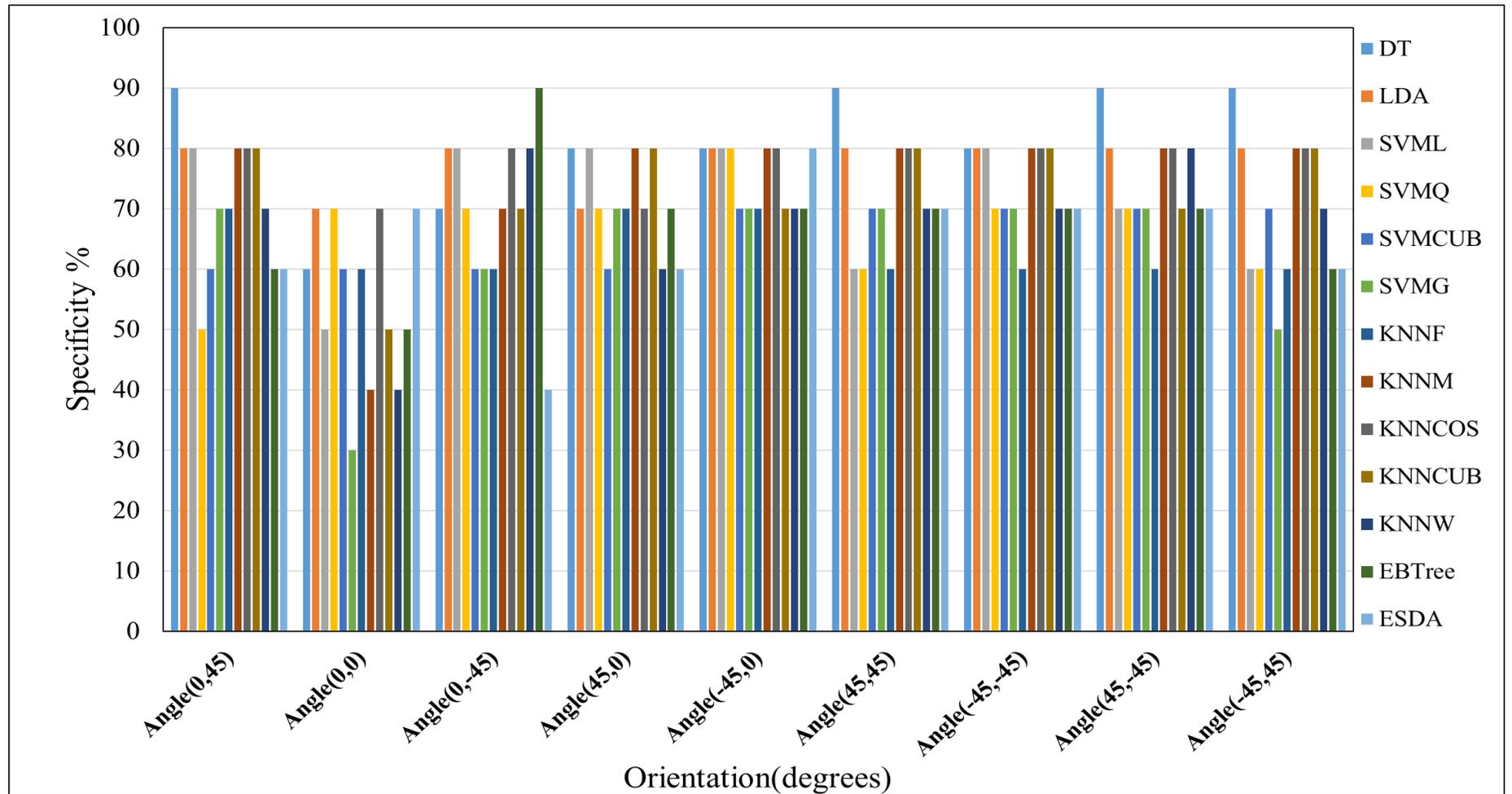
Figure 3.11 The behaviour of the nine angles of the third dimension of 3DGLCM application investigated with different classifiers in term of specificity (sensitivity of low-grades glioma) for BRATS2013 dataset.

3.7.2   Evaluation of the Proposed Algorithm Using the BRATS2013 Dataset

The proposed method for feature selection incorporated different classification methods is evaluated and compared against ANOVA technique. This is based on the steps as follows; the results obtained from feeding all sets of features associated with the 3DGLCM to all classifiers were analysed, then the results gained after applying ANOVA were presented, and finally the results acquired by the proposed features selection method are introduced and compared. These steps are repeated for all four datasets. It is crucial to select the most significant features that lead to improved classification performance. Therefore, the ANOVA technique is used and the significant predictors that showed a P-value less than 0.05 were selected. The ANOVA and Pearson correlation techniques were implemented using IBM SPSS Statistics software Version 20 (Burns and Burns, 2008). When the ANOVA technique was applied to the BRATS2013 dataset to select the best set of features, the experimental results revealed a slight improvement in the classification accuracy, where classification accuracy was improved to 80% and 90% by the classifiers DT and KNNCOS respectively (Table 3.4). Although the classification model SVML remained at the same performance level to show the highest classification accuracy at 93.3%, the dimensions of the features space were reduced from 234 to 166 features, reducing complexity while maintaining the same classification accuracy. The reason for there being no noticeable difference in the results between the use of the full set of features and the use of ANOVA was that the redundant features that were unrecognised by ANOVA were still active and had not been completely eliminated. Consequently, to tackle this limitation, the proposed HFSA was used to eliminate those remaining redundant features and hence improve classification performance. In order to evaluate the ability of the proposed HFSA to select the most significant features and investigate the impact of this method when integrated with different classification methods on the classification results, the proposed HFSA was applied to all features associated with the 3DGLCM for BRATS2013 dataset, which has 234 features. Due to the predictor (*autocorrelation*, 0^0) produces the maximum classification accuracy compared to all other features when it was tested individually; therefore, it was selected to be the reference feature. After applying the proposed algorithm for feature selection, the selected sets of features were evaluated using several classifiers. These classifiers included single classifiers, namely DT, LDA, SVM, and KNN, and ensemble classifiers including EBTree and ESDA. The results indicated that the selected set of features chosen by the

proposed HFSA showed optimal accuracy for classification of glioma grades at 100% using KNNF or ESDA classifies (Table 3.5).

Table 3.4 Comparative results of different classifiers incorporated with the selected set of features associated with 3DGLCM after applying ANOVA using BRTAS2013 dataset.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	7	3	70.00	70.00	70.00	80.00
	Class1	3	17	85.00	85.00	85.00	
LDA	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
SVML	Class0	9	1	90.00	90.00	90.00	**93.33**
	Class1	1	19	95.00	95.00	95.00	
SVMQ	Class0	7	3	70.00	77.80	73.68	83.33
	Class1	2	18	90.00	85.70	87.80	
SVMCUB	Class0	7	3	70.00	70.00	70.00	80.00
	Class1	3	17	85.00	85.00	85.00	
SVMG	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
KNNF	Class0	6	4	60.00	66.70	63.15	76.67
	Class1	3	17	85.00	81.00	82.92	
KNNM	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
KNNCOS	Class0	9	1	90.00	81.80	85.71	90.00
	Class1	2	18	90.00	94.70	92.30	
KNNCUB	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
KNNW	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
EBTree	Class0	7	3	70.00	63.60	66.66	76.67
	Class1	4	16	80.00	84.20	82.05	
ESDA	Class0	4	6	40.00	30.80	34.78	50.00
	Class1	9	11	55.00	64.70	59.45	

Table 3.5 Comparative results of different classifiers incorporated with the selected set of features associated with 3DGLCM after applying the proposed hybrid features selection method using BRTAS2013 dataset.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
DT	Class0	9	1	90.00	100.00	94.73	96.67
	Class1	0	20	100.00	95.20	97.56	
LDA	Class0	10	0	100.00	71.40	83.33	86.67
	Class1	4	16	80.00	100.00	88.88	
SVML	Class0	8	2	80.00	100.00	88.88	93.33
	Class1	0	20	100.00	90.90	95.23	
SVMQ	Class0	9	1	90.00	100.00	94.73	96.67
	Class1	0	20	100.00	95.20	97.56	
SVMCUB	Class0	9	1	90.00	90.00	90.00	93.33
	Class1	1	19	95.00	95.00	95.00	
SVMG	Class0	8	2	80.00	100.00	88.88	93.33
	Class1	0	20	100.00	90.90	95.23	
KNNF	Class0	10	0	100.00	100.00	100.00	**100.00**
	Class1	0	20	100.00	100.00	100.00	
KNNM	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
KNNCOS	Class0	9	1	90.00	100.00	94.73	96.67
	Class1	0	20	100.00	95.20	97.56	
KNNCUB	Class0	10	0	100.00	83.30	90.90	93.33
	Class1	2	18	90.00	100.00	94.73	
KNNW	Class0	8	2	80.00	100.00	88.88	93.33
	Class1	0	20	100.00	90.90	95.23	
EBTree	Class0	6	4	60.00	100.00	75.00	86.67
	Class1	0	20	100.00	83.30	90.90	
ESDA	Class0	10	0	100.00	100.00	100.00	**100.00**
	Class1	0	20	100.00	100.00	100.00	

Overall comparative results illustrate that the classification accuracy of most classifiers has gained a superior classification accuracy when compared to both the full set of features of 3DGLCM and the selected set of features by ANOVA (Figure 3.12). In this experiment, the classification results were improved from 76.67% to 96.67% by DT, from 83.33% to 86.67% by LDA, from 76.67% to 96.67% by SVMQ, from 80% to 100% by KNNF, from 86.67 to 96.67% by KNNCOS, and from 53.33% to 100% by ESDA classifier. It was noted that the proposed features selection method enabled most classifiers to achieve better results compared to the ANOVA technique with a better reduction in the dimensions of the features space, where produced features by ANOVA was 166 features while the selected set of features when using the proposed method was reduced to 14 features (Table 3.6). However, a few classifiers did not show a noticeable improvement, such as SVML classifier, and this was due to the fact that this classifier relies on linear separation between the two classes and ignores any non-linear relationships between different patterns while all other non-linear kernels that were used such as Quadratic, Cubic and Gaussian enabled SVM to achieve a significant improvement in the accuracy based on the selected set of features by the proposed method. KNNM classifier also has not achieved an improvement and remained on the same level of accuracy too, and this was due to the nature of the data distribution of input data besides the criteria that are used with this classifier can detect new samples based on the 10 nearest neighbours and this leads to confusing the KNN classifier in the prediction process to find the correct class; while smaller number of nearest neighbours, enables KNNF classifier uses only one nearest neighbour to predict the correct class, and achieve the optimal classification accuracy at 100%. It was noted that there are various behaviours in classification accuracy when using different classification models due to the use of different sets of features that have diverse data distributions in each input set of features. However, the results indicated that the selected set of features by the proposed method have achieved a dominant improvement in classification accuracy compared to both selected features by the ANOVA method and the original features in most of single classifiers and ensemble classification models.
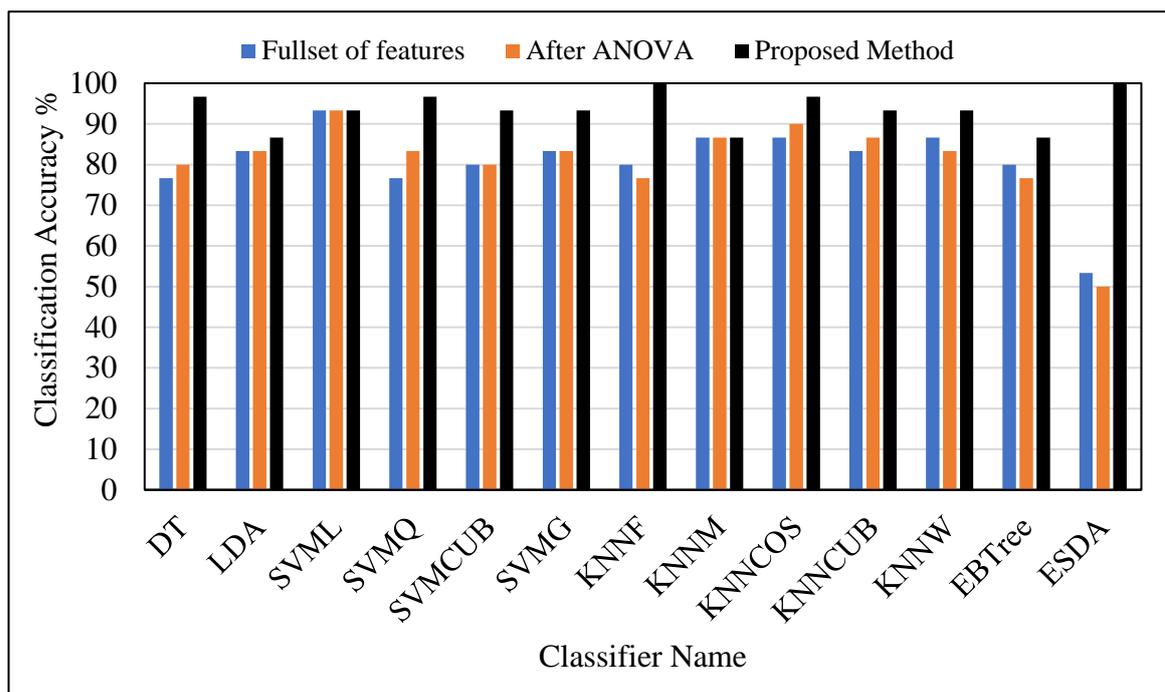
Figure 3.12 Overall comparative results for the application of three cases, the first case is the use of the full set of features, the second case is the use of ANOVA, and the third case is the use of the selected features chosen by the proposed method. All cases are integrated with the same classifiers. The dataset used in this experiment is BRATS2013.

Table 3.6 Optimal selected set of features by the proposed algorithm with their corresponding angles using BRATS2013 dataset

Features	Angles	Features	Angles
Autocorrelation, Correlation, Cluster Prominence, Information Measure of Correlation 1, Information Measure of Correlation 2	(0^0)	Excluded	$(45^0, 0^0)$
Correlation, Cluster Prominence, Information Measure of Correlation 2	(45^0)	Excluded	$(-45^0, 0^0)$
Correlation, Cluster Prominence	90^0	Excluded	$(45^0, 45^0)$
Correlation, Cluster Prominence, Homogeneity	135^0	Excluded	$(-45^0, -45^0)$
Contrast	$(0^0, 45^0)$	Excluded	$(45^0, -45^0)$
Excluded	$(0^0, 0^0)$	Excluded	$(-45^0, 45^0)$
Excluded	$(0^0, -45^0)$		

3.7.3   Evaluation of the Proposed Algorithm Using the Cancer Dataset

The framework that is used in the previous experiment to evaluate the proposed method is repeated using the Cancer dataset. This dataset includes the MR images of 30 patients with three glioma grades. To obtain a binary classification using this dataset, pre-labelling for the samples was performed in the preparation of the training phase, where the samples belonged to grades II were labelled as a first class, representing the low-grade gliomas and were assigned by index 0, the higher-grade gliomas (III and IV) were pre-labelled as a second class and assigned by index 1. Hence, the ten patients belong to grade II were labelled as the first class, and twenty patients belong to high-grade samples were labelled as the second class. The purpose of this experiment was to add further evaluation for the proposed feature selection method incorporating different machine-learning algorithms – a single classifier system and ensemble classification models to classify low-grade gliomas against high-grade gliomas.

When the full set of features associated with the 3DGLCM was used to train and test the same classifiers utilised in this work, the results indicated that the highest classification accuracy, sensitivity and specificity compared to all other single and ensemble classifiers was achieved by DT at 86.67%, 90%, and 80% respectively (Table 3.7). Afterwards, when the selected set of features by ANOVA was used, the results illustrated 90% classification accuracy achieved by DT, which outperformed all other classifiers, followed by 70% achieved by both SVML and KNNCOS classifiers (Table 3.8). When the proposed method was applied to all features associated with the 3DGLCM for the Cancer dataset. The results showed that the maximum results in terms of classification accuracy, sensitivity and specificity were achieved by DT at 93.33%, 95%, and 90% respectively, followed by 90%, 95%, 80% achieved by EBTree (Table 3.9). The results indicate that the classification accuracy was improved after the use of the selected set of features by the proposed hybrid feature selection method (Table 3.9).

Table 3.7 Comparative results of different classifiers incorporated with the full set of features associated with 3DGLCM using the Cancer dataset.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
DT	Class0	8	2	80.00	80.00	80.00	**86.67**
	Class1	2	18	90.00	90.00	90.00	
LDA	Class0	6	4	60.00	50.00	54.54	66.67
	Class1	6	14	70.00	77.80	73.68	
SVML	Class0	3	7	30.00	37.50	33.33	60.00
	Class1	5	15	75.00	68.20	71.42	
SVMQ	Class0	2	8	20.00	25.00	22.22	53.33
	Class1	6	14	70.00	63.60	66.66	
SVMCUB	Class0	2	8	20.00	22.20	21.05	50.00
	Class1	7	13	65.00	61.90	63.41	
SVMG	Class0	0	10	0.00	0.00	0.00	50.00
	Class1	5	15	75.00	60.00	66.66	
KNNF	Class0	4	6	40.00	44.40	42.10	63.33
	Class1	5	15	75.00	71.40	73.17	
KNNM	Class0	3	7	30.00	37.50	33.33	60.00
	Class1	5	15	75.00	68.20	71.42	
KNNCOS	Class0	5	5	50.00	45.50	47.61	63.33
	Class1	6	14	70.00	73.70	71.79	
KNNCUB	Class0	4	6	40.00	44.40	42.10	63.33
	Class1	5	15	75.00	71.40	73.17	
KNNW	Class0	0	10	0.00	0.00	0.00	50.00
	Class1	5	15	75.00	60.00	66.66	
EBTree	Class0	3	7	30.00	37.50	33.33	60.00
	Class1	5	15	75.00	68.20	71.42	
ESDA	Class0	3	7	30.00	17.60	22.22	30.00
	Class1	14	6	30.00	46.20	36.36	

Table 3.8 Comparative results of different classifiers incorporated with the selected set of features associated with 3DGLCM after applying ANOVA using the Cancer dataset.

| Classifier | Actual class | Confusion matrices | | Sensitivity % | Precision % | F-measure % | Accuracy % |
| | | Predicted class | | | | | |
		Class0	Class1				
DT	Class0	8	2	80.00	88.90	84.21	**90.00**
	Class1	1	19	95.00	90.50	92.68	
LDA	Class0	7	3	70.00	53.80	60.86	70.00
	Class1	6	14	70.00	82.40	75.67	
SVML	Class0	5	5	50.00	55.60	52.63	70.00
	Class1	4	16	80.00	76.20	78.04	
SVMQ	Class0	3	7	30.00	30.00	30.00	53.33
	Class1	7	13	65.00	65.00	65.00	
SVMCUB	Class0	4	6	40.00	36.40	38.09	56.67
	Class1	7	13	65.00	68.40	66.66	
SVMG	Class0	1	9	10.00	14.30	11.76	50.00
	Class1	6	14	70.00	60.90	65.11	
KNNF	Class0	4	6	40.00	44.40	42.10	63.33
	Class1	5	15	75.00	71.40	73.17	
KNNM	Class0	6	4	60.00	50.00	54.54	66.67
	Class1	6	14	70.00	77.80	73.68	
KNNCOS	Class0	7	3	70.00	53.80	60.86	70.00
	Class1	6	14	70.00	82.40	75.67	
KNNCUB	Class0	5	5	50.00	45.50	47.61	63.33
	Class1	6	14	70.00	73.70	71.79	
KNNW	Class0	1	9	10.00	16.70	12.50	53.33
	Class1	5	15	75.00	62.50	68.18	
EBTree	Class0	4	6	40.00	40.00	40.00	60.00
	Class1	6	14	70.00	70.00	70.00	
ESDA	Class0	5	5	50.00	35.70	41.66	53.33
	Class1	9	11	55.00	68.80	61.11	

Table 3.9 Comparative results of different classifiers incorporated with the selected set of features associated with 3DGLCM after applying the proposed hybrid features selection method using the Cancer dataset.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	9	1	90.00	90.00	90.00	**93.33**
	Class1	1	19	95.00	95.00	95.00	
LDA	Class0	5	5	50.00	55.60	52.63	70.00
	Class1	4	16	80.00	76.20	78.04	
SVML	Class0	4	6	40.00	50.00	44.44	66.67
	Class1	4	16	80.00	72.70	76.19	
SVMQ	Class0	6	4	60.00	60.00	60.00	73.33
	Class1	4	16	80.00	80.00	80.00	
SVMCUB	Class0	5	5	50.00	71.40	58.82	76.67
	Class1	2	18	90.00	78.30	83.72	
SVMG	Class0	7	3	70.00	63.60	66.66	76.67
	Class1	4	16	80.00	84.20	82.05	
KNNF	Class0	6	4	60.00	54.50	57.14	70.00
	Class1	5	15	75.00	78.90	76.92	
KNNM	Class0	6	4	60.00	60.00	60.00	73.33
	Class1	4	16	80.00	80.00	80.00	
KNNCOS	Class0	7	3	70.00	63.60	66.66	76.67
	Class1	4	16	80.00	84.20	82.05	
KNNCUB	Class0	6	4	60.00	60.00	60.00	73.33
	Class1	4	16	80.00	80.00	80.00	
KNNW	Class0	6	4	60.00	54.50	57.14	70.00
	Class1	5	15	75.00	78.90	76.92	
EBTree	Class0	8	2	80.00	88.90	84.21	90.00
	Class1	1	19	95.00	90.50	92.68	
ESDA	Class0	3	7	30.00	42.90	35.29	63.33
	Class1	4	16	80.00	69.60	74.41	

Overall comparative results obtained before and after the use of ANOVA, and after the use of the proposed feature selection method, indicate that the selected set of features by the proposed method have achieved superior accuracies with most classifiers compared to all experiments. They were improved from 86.67% to 93.33% by DT, 50% to 76.67% by SVMCUB, from 50% to 76.67% by SVMG, from 63.33% to 70% by KNNF, from 60% to 90% by EBTree (Figure 3.13).

The proposed method has also achieved a significant reduction in the number of dimensions of the features space, where the number of features was reduced from 243 to 3 features, and irrelevant features were discarded; this is better than the number of features selected by ANOVA, where the number of selected features by ANOVA was 109 features. The selected set of features by the proposed method were ($Autocorrelation$, 0^0), ($Homogeneity$, 90^0), and ($Homogeneity$, 0^0). The best classification accuracy was achieved by DT classifier at 93.33% outperformed all other results produced by other classifiers (Figure 3.13).
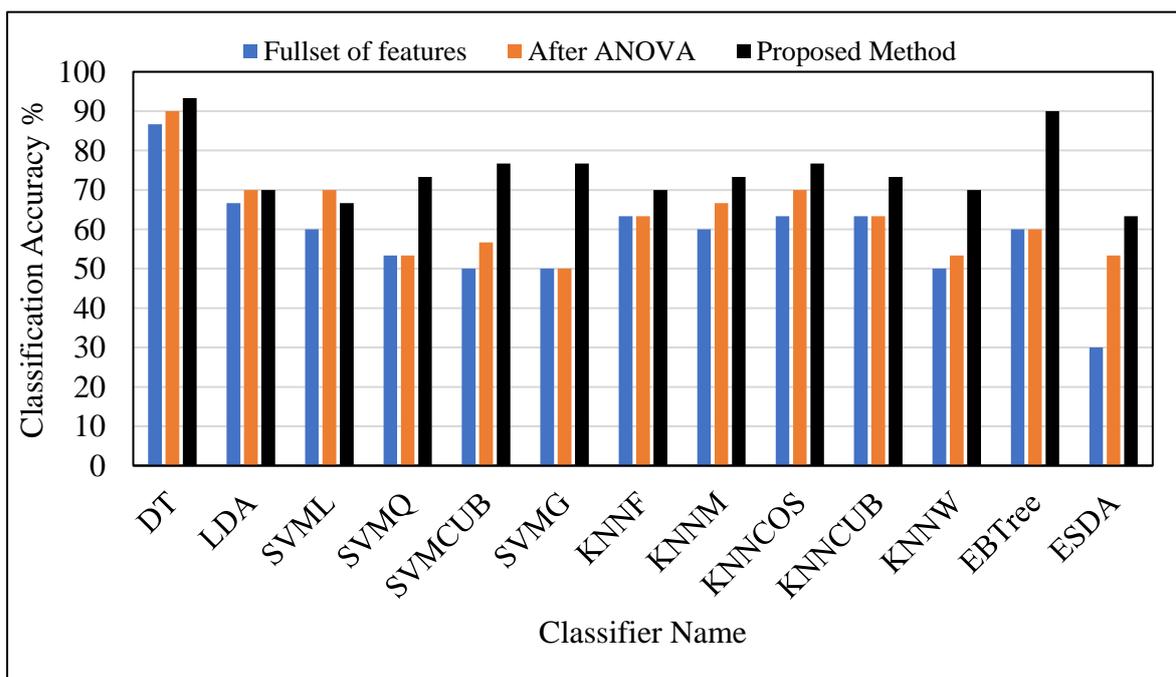
Figure 3.13 Overall comparative results between the application of the proposed method against ANOVA and the full set of features. The selected features chosen by these applications are integrated with different classifiers. The dataset used in this experiment is the Cancer dataset.

3.7.4   Evaluation of the Proposed Algorithm Using the BRATS2015 Dataset

The proposed algorithm for feature selection was applied to BRATS2015 to conduct further evaluation for the proposed method. This dataset has a standard segmented dataset of MR images with 274 patients, which includes 220 patients with high-grade gliomas and 54 patients with low-grade gliomas. T2-MR images were used to extract 3DGLCM, incorporated with 18 predictors with full use of the 13 angles of GLCM. The low-grade samples and the high-grade samples of this dataset were pre-labelled by the index 0 and index 1 respectively.

When the full set of features associated with the 3DGLCM was used, the results indicated that the highest classification accuracy was achieved by KNNF at 85.77%, followed by SVMQ at 85.40% (Table 3.10). After applying the ANOVA technique, the results illustrated that the best classification accuracy was also achieved by KNNF classifier, at 86.50% (Table 3.11). When the proposed features selection algorithm was conducted and integrated with different classification models, the results illustrated that the best classification accuracy compared to all other classifiers was achieved by KNNF classifier, at 87.96% (Table 3.12). Different classifiers have shown various results in terms of sensitivity, specificity and classification accuracy, such as SVMQ and SVMG, where their sensitivities were 93.64% and 95% respectively, while their specificities were 46.30% and 35.19% respectively, and their accuracies were 84.31% and 83.21% respectively (Table 3.12). However, the best classifier that showed the best results in terms of sensitivity at 93.18% and specificity at 66.67% was the KNNF classifier, leading to the best classification accuracy at 87.96% (Table 3.12). It was noted that there is a large difference in the classification performance in terms of sensitivity and specificity between low-grade and high-grade gliomas. This is because that there is a difference in the number of samples between the low grades and high grades; this can lead to different representations of each class in the training phase, which can reveal a higher sensitivity to the one that has a higher representation. The number of samples of higher-grade group is 220 compared to lower-grade class that has 54 samples. Therefore the high-grade class reflects higher sensitivity compared to the low-grade class (Table 3.10, and Table 3.11, and Table 3.12). When the proposed method was applied, the results indicate that the classification accuracy was improved after the use of the selected set of features by the proposed hybrid feature selection method (Table 3.12).

Table 3.10 Comparative results of different classifiers incorporated with the full set of features associated with 3DGLCM using BRTAS2015 dataset.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	6	48	11.11	16.70	13.33	71.53
	Class1	30	190	86.36	79.80	82.96	
LDA	Class0	43	11	79.63	45.30	57.71	77.01
	Class1	52	168	76.36	93.90	84.21	
SVML	Class0	27	27	50.00	65.90	56.84	85.04
	Class1	14	206	93.64	88.40	90.94	
SVMQ	Class0	28	26	51.85	66.70	58.33	85.40
	Class1	14	206	93.64	88.80	91.15	
SVMCUB	Class0	27	27	50.00	60.00	54.54	83.58
	Class1	18	202	91.82	88.20	89.97	
SVMG	Class0	19	35	35.19	57.60	43.67	82.12
	Class1	14	206	93.64	85.50	89.37	
KNNF	Class0	37	17	68.52	62.70	65.48	**85.77**
	Class1	22	198	90.00	92.10	91.03	
KNNM	Class0	27	27	50.00	50.90	50.46	80.66
	Class1	26	194	88.18	87.80	87.98	
KNNCOS	Class0	29	25	53.70	47.50	50.43	79.20
	Class1	32	188	85.45	88.30	86.83	
KNNCUB	Class0	26	28	48.15	52.00	50.00	81.02
	Class1	24	196	89.09	87.50	88.28	
KNNW	Class0	23	31	42.59	60.50	50.00	83.21
	Class1	15	205	93.18	86.90	89.91	
EBTree	Class0	22	32	40.74	57.90	47.82	82.48
	Class1	16	204	92.73	86.40	89.47	
ESDA	Class0	32	22	59.26	53.30	56.14	81.75
	Class1	28	192	87.27	89.70	88.47	

Table 3.11 Comparative results of different classifiers incorporated with the selected set of features associated with 3DGLCM after applying ANOVA using BRTAS2015 dataset.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	6	48	11.11	16.70	13.33	71.53
	Class1	30	190	86.36	79.80	82.96	
LDA	Class0	43	11	79.63	45.30	57.71	77.01
	Class1	52	168	76.36	93.90	84.21	
SVML	Class0	26	28	48.15	63.40	54.73	84.31
	Class1	15	205	93.18	88.00	90.50	
SVMQ	Class0	28	26	51.85	65.10	57.73	85.04
	Class1	15	205	93.18	88.70	90.90	
SVMCUB	Class0	32	22	59.26	59.30	59.25	83.94
	Class1	22	198	90.00	90.00	90.00	
SVMG	Class0	20	34	37.04	62.50	46.51	83.21
	Class1	12	208	94.55	86.00	90.04	
KNNF	Class0	35	19	64.81	66.00	65.42	**86.50**
	Class1	18	202	91.82	91.40	91.61	
KNNM	Class0	29	25	53.70	56.90	55.23	82.85
	Class1	22	198	90.00	88.80	89.39	
KNNCOS	Class0	27	27	50.00	64.30	56.25	84.67
	Class1	15	205	93.18	88.40	90.70	
KNNCUB	Class0	30	24	55.56	55.60	55.55	82.48
	Class1	24	196	89.09	89.10	89.09	
KNNW	Class0	27	27	50.00	64.30	56.25	84.67
	Class1	15	205	93.18	88.40	90.70	
EBTree	Class0	19	35	35.19	54.30	42.69	81.39
	Class1	16	204	92.73	85.40	88.88	
ESDA	Class0	29	25	53.70	49.20	51.32	79.93
	Class1	30	190	86.36	88.40	87.35	

Table 3.12 Comparative results of different classifiers incorporated with the selected set of features associated with 3DGLCM after applying the proposed hybrid features selection method using BRTAS2015 dataset.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	18	36	33.33	51.40	40.44	80.66
	Class1	17	203	92.27	84.90	88.45	
LDA	Class0	41	13	75.93	44.10	55.78	76.28
	Class1	52	168	76.36	92.80	83.79	
SVML	Class0	30	24	55.56	68.20	61.22	86.13
	Class1	14	206	93.64	89.60	91.55	
SVMQ	Class0	25	29	46.30	64.10	53.76	84.31
	Class1	14	206	93.64	87.70	90.54	
SVMCUB	Class0	33	21	61.11	61.10	61.11	84.67
	Class1	21	199	90.45	90.50	90.45	
SVMG	Class0	19	35	35.19	63.30	45.23	83.21
	Class1	11	209	95.00	85.70	90.08	
KNNF	Class0	36	18	66.67	70.60	68.57	**87.96**
	Class1	15	205	93.18	91.90	92.55	
KNNM	Class0	30	24	55.56	52.60	54.05	81.39
	Class1	27	193	87.73	88.90	88.32	
KNNCOS	Class0	30	24	55.56	50.80	53.09	80.66
	Class1	29	191	86.82	88.80	87.81	
KNNCUB	Class0	29	25	53.70	53.70	53.70	81.75
	Class1	25	195	88.64	88.60	88.63	
KNNW	Class0	29	25	53.70	64.40	58.58	85.04
	Class1	16	204	92.73	89.10	90.86	
EBTree	Class0	21	33	38.89	58.30	46.66	82.48
	Class1	15	205	93.18	86.10	89.51	
ESDA	Class0	32	22	59.26	47.10	52.45	78.83
	Class1	36	184	83.64	89.30	86.38	

Overall comparative results acquired before and after the use of ANOVA, and after the use of the proposed feature selection method show that the selected set of features by the proposed method achieved enhanced accuracies for most classifiers, where the classification accuracies were improved from 71.53% to 80.66% by DT, from 85.04% to 86.13% by SVML, from 83.58% to 84.67 % by SVMCUB, from 82.12% to 83.21% by SVMG, from 85.77% to 87.96% by KNNF, from 80.66% to 81.39% by KNNM, from 79.2% to 80.66% by KNNCOS and from 83.21 % to 85.04% by KNNW (Figure 3.14). It is evident that the performance of many classifiers in term of classification accuracy has improved and the best classification accuracy was obtained by KNNF classifier at 87.96%, which outperformed all other results produced by other classifiers (Figure 3.14). Both the ANOVA technique and the proposed HFSA achieved a notable reduction in the features space as well as they achieved an improvement in the classification accuracy that was notable with many classifiers. However, the number of selected features by the proposed HFSA was reduced from 243 to 129 features (Table 3.13), and irrelevant features were discarded, which is smaller than the number of features selected by ANOVA, which was 199 features. The results indicated that the proposed HFSA achieved better accuracy in the classification of glioma grades using many classifiers. Furthermore, the proposed algorithm integrated with machine learning algorithms reduces the dimension of the feature space as well as maintain a better classification accuracy.
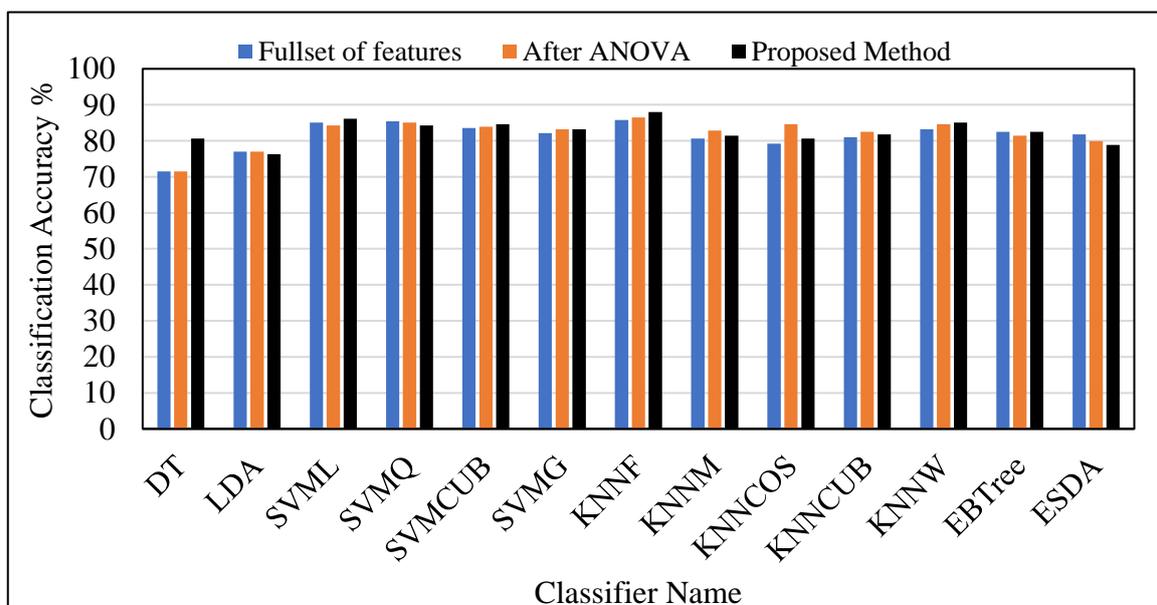
Figure 3.14 Overall comparative results for the application of ANOVA and the proposed feature selection method integrated with many different classifiers; the dataset used in this experiment is BRATS2015.

Table 3.13 Optimal selected set of features by the proposed HFSA with their corresponding angles using BRATS2015 dataset. The names of the features are referred by its abbreviations that were defined in the expression of their equations (APPENDIX B, B.1.1-B.1.14)

Features	Angles	Features	Angles
Autoc, contr, corrm, cprom, dissi homom, maxpr, sosq, savgh, svarh senth, inf1h, inf2h, indnc, idmnc	(0^0)	*Corm, cshad, dissi homom sosq, savgh, svarh, indnc*	$(45^0, 0^0)$
Autoc, cprom, dissi, maxpr, sosq savgh, svarh, indnc, idmnc	(45^0)	*Autoc, contr, cshad, dissi homom, sosq, savgh, svarh, senth*	$(-45^0, 0^0)$
Autoc, corrm, cprom, dissi homom, maxpr, sosq, savgh svarh, senth, inf1h, indnc	90^0	Autoc, cprom, cshad, homom, sosq, savgh, svarh senth	$(45^0, 45^0)$
Autoc, contr, corrm, cprom, homom maxpr, sosq, savgh, svarh, senth inf2h, idmnc	135^0	Autoc, corrm, cprom, cshad homom, sosq, savgh, svarh senth, indnc	$(-45^0, -45^0)$
Autoc, contr, cprom, cshad, dissi homom, sosq, savgh, svarh, senth idmnc	$(0^0, 45^0)$	Autoc, corrm, cshad, entro sosq, savgh, svarh, senth	$(45^0, -45^0)$
Autoc, contr, cprom, homom, maxpr, sosq, savgh, svarh	$(0^0, 0^0)$	Autoc, cshad, dissi, entro sosq, savgh, svarh, senth indnc	$(-45^0, 45^0)$
Autoc, corrm, cshad, dissi, entro, sosq, savgh, svarh, senth, indnc	$(0^0, -45^0)$		

3.7.5 Evaluation of the Proposed Algorithm Using the BRATS2018 Dataset

The developed classification system incorporated the proposed algorithm for feature selection were further evaluated using BRATS2018. This dataset has a standard segmented dataset of MR images with 285 patients, which includes 210 patients with high-grade gliomas and 75 patients with low-grade gliomas. T2-MR images were used to extract 3DGLCM, incorporated with the 18 predictors and thus patients were represented by 234 feature vectors.

The same proposed framework developed in the previous experiments, was implemented for this dataset and three cases incorporated machine learning algorithms were investigated; the first case is the use of the full set of features, the second case is the implementation of the ANOVA technique, and the third cases is the implementation of the proposed HFSA. The results for of these cases in terms of the confusion matrix, classification accuracy, sensitivity, specificity, precision, and F-measure are shown in APPENDIX C, Table C.1-Table C.3 . The best classification accuracy for the first case is achieved by SVML classifier at 86.32%. While, in the second and third cases the highest classification accuracy obtained at 87.02% and 88.07% respectively by the same classifier.

Overall comparative results indicated that the classification performance of most of classification models are improved using both ANOVA technique and the proposed HFSA (Figure 3.15). For example, the accuracy was improved from 86.31% to 87.02% through the use of selected features by ANOVA incorporated with SVML classifier. While the classification accuracy was enhanced to 88.07% after applying the proposed HFSA (Figure 3.15).

Both the ANOVA technique and the proposed HFSA accomplished a notable reduction in the features space as well as they achieved an improvement in the classification accuracy. However, the number of selected features by the proposed HFSA was reduced from 243 to 145 features (APPENDIX C Table C.4), and irrelevant features were discarded, which is smaller than the number of features selected by ANOVA, which was 224 features. The results indicated that the proposed HFSA achieved better accuracy in the classification of glioma grades when integrated with machine learning algorithms through eliminating redundant features, leading to reduce the dimensions of the feature space as well as maintaining a better classification accuracy.
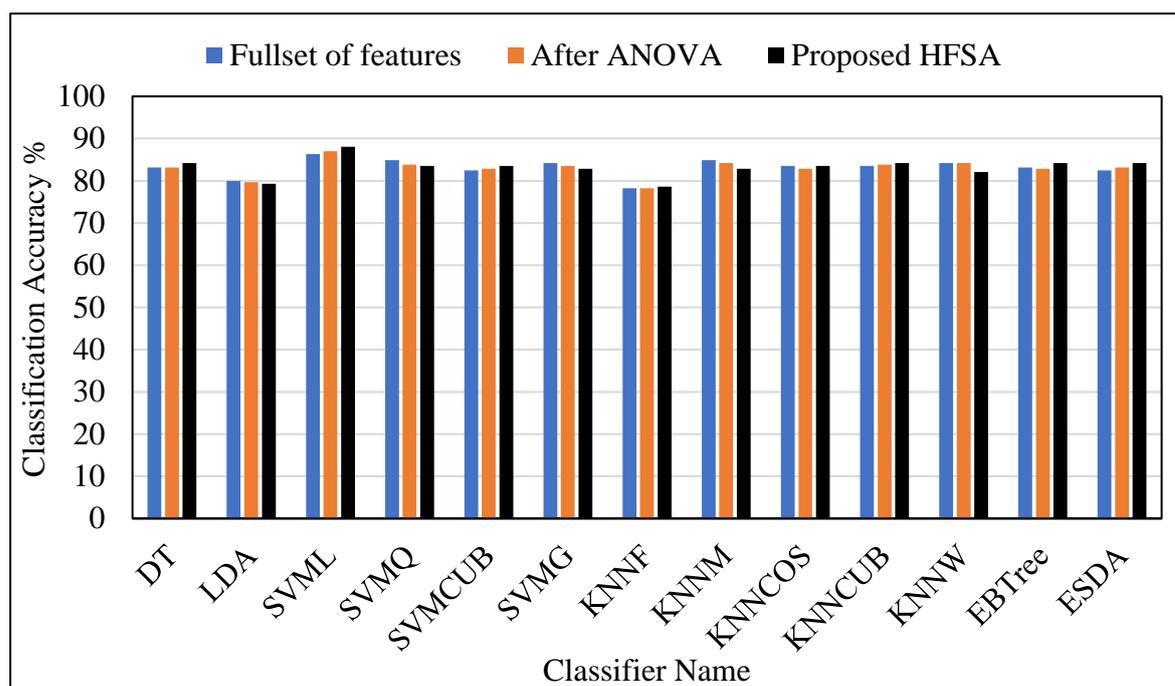
Figure 3.15 Overall comparative results for the application of the proposed feature selection algorithm integrated with many different classifiers against the application of ANOVA and the use of full set of features; the dataset used in this experiment is BRATS2018.

3.7.6 Implementation Time

The implementation time refers to the time required to conduct the experiments that were performed in this work. Particularly, it represents the time that was needed for the training and testing a machine-learning algorithm using the features extracted from the MR images. This time is based on using a personal computer with the specification of Core i7; the RAM was 16 G and parallel processing was enabled. Different implementation times were noted to train and test different classifiers based on LOO cross-validation techniques for both 2D and 3DGLCM (Table 3.14). It was noted that the time spent on the training and testing of the different classifiers was slightly higher for 3DGLCM than for 2DGLCM (Table 3.14). The results also showed that the time required for the implementation of the ensemble method was relatively higher than the time required to run the single classifier. This time was measured using the BRATS2013 dataset and Cancer dataset, and due to both datasets including the same number of samples the times needed to run the classification system were the same using both the BRATS2013 and the Cancer datasets (Table 3.14). While the time required to implement the training and testing phases for the classification system using BRATS2015 and BRATS2018 was relatively higher compared to using the other datasets (Table 3.15 and Table 3.16). The BRATS2015 and BRATS2018 dataset included a larger number of data, with 274 patients and 285 patients respectively, while each of the other

datasets contained 30 samples. This indicates that the time required during the development of the classification system is correlated with the number of samples, where a greater number of samples mean more training and testing times required to complete the classification process.

Table 3.14 Implementation time for the training and testing of a different classifier based on 2D and 3DGLCM.This time includes training and testing phase of different classifiers using LOO cross-validation technique for both BRATS2013 and Cancer dataset.

Classifier	Execution time for 2DGLCM (seconds)	Execution time for 3DGLCM (seconds)
DT	0.4447	0.5611
LDA	0.4333	0.4402
SVML	0.4707	0.5159
SVMQ	0.4062	0.4860
SVMCUB	0.4461	0.5026
SVMG	0.4401	0.4562
KNNF	0.4169	0.5230
KNNM	0.3941	0.4162
KNNCOS	0.3772	0.4323
KNNCUB	0.4938	0.5847
KNNW	0.3800	0.4388
EBTree	4.3680	5.2718
ESDA	5.4086	6.9563

Table 3.15 Implementation time for the training and testing of a different classifier based on 2D and 3DGLCM. This time includes training and testing phase of different classifiers using LOO cross-validation technique for BRATS2015 dataset.

Classifier	Execution time for 2DGLCM (seconds)	Execution for 3DGLCM (seconds)
DT	2.8179	3.5812
LDA	1.3961	2.1442
SVML	1.5821	1.9833
SVMQ	1.3400	1.7423
SVMCUB	1.5068	1.8582
SVMG	0.9710	1.3199
KNNF	0.7415	0.9295
KNNM	0.6420	0.8374
KNNCOS	0.6206	0.8286
KNNCUB	0.7254	0.9286
KNNW	0.6512	0.8155
EBTree	16.4809	21.3421
ESDA	17.8323	25.4095

Table 3.16 Implementation time for the training and testing of a different classifier based on 2D and 3DGLCM. This time includes training and testing phase of different classifiers using LOO cross-validation technique for BRATS2018 dataset.

Classifier	Execution time for 2DGLCM (seconds)	Execution for 3DGLCM (seconds)
DT	5.7626	6.3756
LDA	2.4700	3.2265
SVML	2.1192	2.4478
SVMQ	1.5687	1.6520
SVMCUB	1.7485	1.8439
SVMG	1.0389	1.4407
KNNF	1.3121	1.4514
KNNM	0.6546	0.9125
KNNCOS	0.6809	0.9249
KNNCUB	0.7569	0.9934
KNNW	0.6876	0.7924
EBTree	19.3906	22.5171
ESDA	20.1630	28.2650

3.7.7 Results Overview and Discussion

Several experiments and methods were investigated to evaluate the ability of the proposed classification system in achieving an accurate classification of WHO glioma grades. To achieve this, a wide range of evaluation metrics was used. Also, four public datasets were utilised to evaluate the general behaviour of the developed classification system. Several different classification methods were used to add further validity to the evaluation and to identify the best classification model, which could provide the highest classification accuracy to distinguish different WHO glioma grades.

When implementing the classification system using BRATS2013 dataset, the results indicated that the third-axis of 3DGLCM, integrated with the DT classifier, outperformed the others and achieved an accuracy of 96.7% at the orientation $(0^0,45^0)$ followed by 93.33% achieved by both SVML and DT classifier at the orientations 45^0, and 135^0 respectively.

It was noted that the most robust classifiers appropriate for use with the selected set of features by the proposed algorithm for feature selection were KNNF, KNNW, DT, SVMCUB, SVMG, and EBTree classifier; these classifiers showed improved classification accuracy when evaluated with all the datasets. The other classifiers showed various behaviours and generally illustrated improved classification accuracy for most datasets.

In the evaluation of the classification system using BRATS2015, the results indicated that the improvement in classification accuracy achieved was relatively lower compared to the results gained when using the other datasets. The results acquired for BRATS2013, the Cancer dataset, BRATS2015 and BRATS2018 were 100%, 93.33%, 87.9% and 88.07% respectively (3.7.2, 3.7.3, 3.7.4 and 3.7.5). To investigate the reason(s) behind this difference, the BRATS2013 dataset, the BRATS2015 and BRATS2018 dataset were analysed. The characteristics of the datasets in terms of the homogeneity levels of the brain tumour images were investigated. The rationale for a further investigation into the level of homogeneity was due to the belief that there may have been a difference in the homogeneity levels of the brain tumour images datasets. Furthermore, there is a significant correlation between the level of homogeneity of brain tumour images and generating an accurate representation of the texture features extracted from the MR images. It is stated that the images of this datasets have been further homogenised (Menze et al., 2015). However, there is not much detail regarding the specific level of homogeneity of these datasets. Therefore, the local homogeneity levels of MR brain tumour images for all samples that were associated with each glioma grade in the datasets were measured based on APPENDIX B, Eq. B.15. The results obtained by applying this formula for the MR images in the datasets were analysed using the boxplot technique (Ferreira et al., 2016). This technique was used to show the differences in data distribution and level of homogeneity of the datasets. Considering the differences in the median level of the homogeneity (Figure 3.16), the experimental results indicated that the average homogeneity levels measured from MR brain tumour images in the BRATS2015 at 0.248 and BRATS2018 at 0.241 were higher than the average level in the BRATS2013 dataset at 0.234 (Figure 3.16). Accordingly, high level of image homogeneity can reduce the chance of obtaining sufficient representation for the texture features. For further elaboration, increasing the level of homogeneity of tumour images can lead to ignoring small details of a tumour and not being detected accurately by the texture extraction process. Consequently, this increase in the level of homogeneity can reduce the amount of texture that is required to differentiate between different patterns of tumours, and hence the classification accuracy can be negatively affected and reduced. It is worthwhile to note that the low level of homogeneity of brain tumour images can be advantageous for better recognising small details in tumour patterns and to discriminate them easily from other tumour regions or patterns. A high degree of homogeneity can relatively lead to ignoring some significant structures of the tumour and this can lead to confusion in the classification process.
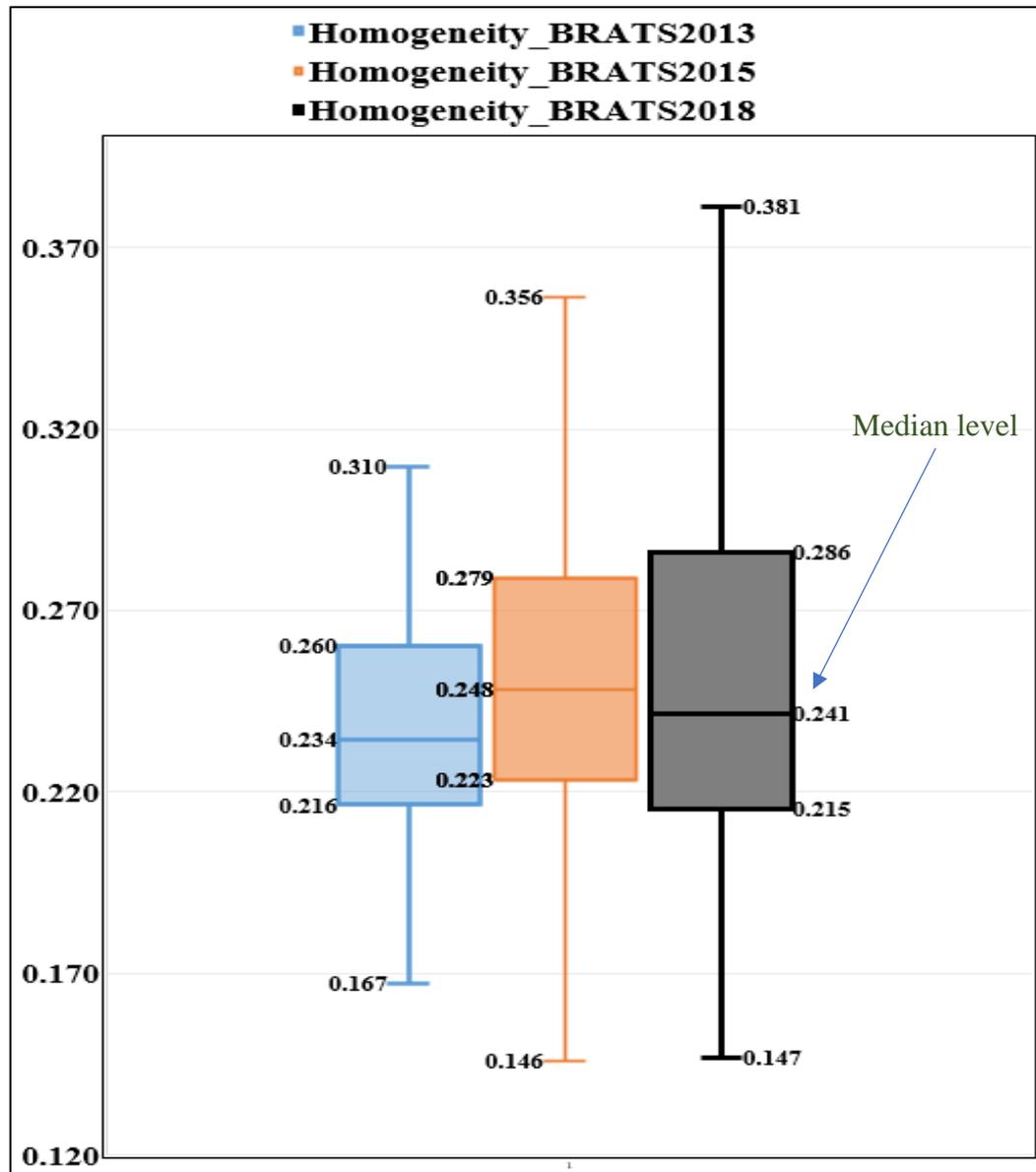
Figure 3.16 Results comparison for the level of homogeneity measured from MR brain tumour images of the BRATS2013, BRATS2015 and BRATS2018. The difference in median levels of homogeneity can be seen in the middle of the figure.

The most challenging issue in the classification of glioma grades is to discriminate between Grade II and Grade III, as these grades have mixed heterogenetic characteristics that make their classification problematic and, therefore, it is a significant challenge to achieve sufficient classification accuracy between them (Zacharaki et al., 2009). However, an efficient classification algorithm was proposed to discriminate between the low grades that include grade II, against high glioma grades that include grade III. It has been proven based on the results obtained that the proposed classification system has achieved a perfect

classification result (100% accuracy) in differentiating low grade-grade gliomas (I, II) from high-grade gliomas (III, IV) using BRATS2013 dataset.

Several methods were deployed to evaluate and validate the proposed HFSA, including examining the proposed method using four public datasets of MR images that have confirmation of histopathological diagnoses. In addition, a range of different classification models was incorporated into the classification system and their performances were compared. The results indicated that maximum classification results were achieved in discriminating between low- and high-grade gliomas using the BRATS2013 dataset. The proposed method also achieved an improvement in classification performance when used with the other datasets. Furthermore, the proposed HFSA achieved a greater reduction in the features space dimensions compared to the ANOVA technique over all the datasets.

It was observed that the best classification result was achieved using the BRATS2013 dataset, where the results revealed that the proposed HFSA boosted the KNN and ESDA classifiers to achieve the optimal classification accuracy of 100%. The experimental results confirmed that the proposed algorithm is an effective approach to overcoming the limitations of redundancy in feature space and enhancing the performance of many machine-learning algorithms to achieve improved classification of glioma grades. It improves the classification performance by eliminating the feature space from the most redundant features that degrade the classification accuracy.

3.8 Conclusion

The classification of glioma grades is a challenging task due to the mixed characteristics of different heterogeneity levels in the tumour images, which makes for significant difficulty in achieving an accurate discrimination between different glioma grades. Despite the high level of subjectivity in visual diagnoses of malignant brain tumours, experts can make decisions and achieve a certain degree of accuracy in the diagnostic process. However, there is always the need for more objective and accurate decision-making to be facilitated. Therefore, to overcome these limitations and to improve the classification process of glioma grades, an automated classification system was developed to achieve an objective and accurate classification of glioma grades. The classification was developed based on the texture features that have been widely investigated and studied to enhance the diagnosis and classification of the malignancy degree of brain tumours in MR images. The study started with the pre-processing of T2-MR images including the preparation of MRI-ROI of brain

tumour, the automatic cropping of ROI of brain tumours from MR images, and intensity normalisation processes. These were conducted to prepare the MR image of ROI for the next steps – features extraction and classification. A comprehensive analysis was then conducted based on two and three dimensions of MRI textural features to discriminate between low- and high-grade gliomas. This was based on GLCM incorporated eighteen statistics, commonly and widely used in the scholarly literature. A new algorithm was proposed for selecting the most crucial features in attempting to enhance classification accuracy. The proposed algorithm investigated the interactions between features in coincidence with their relationships with the output accuracy guided by Pearson correlation method and the outcome from different classifiers. The proposed method takes the merits of using Pearson correlation and ANOVA, integrated with different classification models to achieve improvement in the performance of the classification of malignant brain tumour. Several machine-learning algorithms that are popular and commonly used namely DT, LDA, SVM, KNN, EBTree and ESDA were investigated. These classifiers were incorporated with the selected set of features that were chosen by the proposed HFSA. The purpose of this was to select the best classification model that could achieve optimal performance in the classification of glioma grades. Different common and significant metrics were used to evaluate the performance of the classification system for glioma grades such as classification accuracy, sensitivity, specificity, precision and F-measure.

It is concluded that the proposed algorithm integrated with the machine-learning method is an efficient approach and can achieve significant results in term of enhancing classification accuracy and achieving a better features dimension reduction. However, improvement varied across the four datasets whereby accuracy was achieved 100% when using the BRATS2013 dataset, 93.3% when using the Cancer dataset, 87.9% when using the BRATS2015 dataset, and 88.07% for the BRATS2018. Therefore, further work will be conducted to develop and improve this classification system to achieve constantly improving accuracy in the classification of glioma grades. The next chapter will develop an automated classification system for glioma grades based on other morphological descriptors of brain tumours and it will evaluate and examine the impact of the quantitative features of these tumour descriptors on the classification of glioma grades.

CHAPTER 4 : MRI Classification System for Glioma Grades Based on Necrosis, Edema, Non-Enhanced, and Enhanced Tumour

Overview

This chapter presents a novel method to extract significant features of brain tumours from MR images, which reflect the objective analysis of different brain tumour descriptors, namely necrosis, edema, non-enhanced and enhanced tumours. The purpose of this work is to evaluate the discrimination ability of the proposed features, integrated with machine-learning algorithms, to achieve an accurate and automated classification of malignant grades of glioma. These new features were used to train different classification models to differentiate high-grade gliomas (III and IV) from low grades (I and II). The proposed classification system was evaluated using three datasets, namely BRATS2013 with 30 patients, BRATS2015 with 274 patients and BRATS2018 with 285 patients. The classification results were then compared using different evaluation metrics, namely classification accuracy, sensitivity, specificity, precision and F-measure. The classification performance of the developed system was validated and generalised using the leave-one-out cross-validation technique.

4.1 Introduction

The analysis of medical images is highly complex due to the fact that they reflect various attributes and different structures of the human body (Toennies, 2017). Therefore, it is crucial to have a high level of experience to achieve an accurate diagnosis of medical images. The rapid advances in medical image technologies and the large amount of medical data create a big challenge in medical fields where experts consume a lot of time and effort to achieve an accurate assessment for large amounts of data. Malignant brain tumours have different morphological descriptors that can be extracted from medical images and it is a significant challenge to assess an accurate malignant grade of brain tumour due to the

complexity of the tumour descriptors appearing in medical images. Pre-operative diagnosis of glioma grades is essential in order to manage the proper treatment and suitable prognosis for patients who develop cancer. The visual diagnosis can even be difficult if there is a suspicion that there is malignant growth of the tumour where the patient has to be sent to have a clinical invasive process conducted, such as a biopsy. However, the visual diagnosis of a malignant brain lesion using only the naked eye is subjective and time-consuming. Developing an automated classification system for glioma grades based on a statistical analysis of tumour descriptors, means being able to achieve an accurate differentiation between different grades of glioma and thereby offering a significant aid to support clinicians in determining the accurate grade of glioma.

Brain tumour descriptors, including necrosis, edema, non-enhancement and enhancement of tumours, are important indicators used in the clinical diagnosis of malignant brain tumours (Moore and Kim, 2010). The presences of the tumour descriptors are visually assessed within the clinical diagnostic procedure to determine the malignant grades of glioma. However, it highly depends on the level of the qualification and experience of the expert. It is also time-consuming and suffers from inter and intra subjectivity (Saad et al., 2015). Furthermore, it is stated that despite the optimisation of MRI protocols and sequences, the glioma grading through the visual diagnosis on MR images is sometimes unreliable, with the sensitivity for classification of glioma grades ranging from 55.1% to 83.3% (Geneidi et al., 2015). Accordingly, in attempting to enhance the quality of classification accuracy, an objective and efficient approach based on the incorporation of the statistical analysis of the tumour descriptors with the machine learning algorithm is developed in this work. The rationale of this development is to overcome the variations and subjectivity in the assessment of the tumour descriptors by the traditional visual diagnosis. Furthermore, developing the new features and integrating them with the machine learning algorithm is because the developed predictors associated with the tumour descriptors are independent of the variation of image intensities and the resolution of tumour images. To elaborate, this variation can play a significant role in providing sufficient representation of texture features that derived from image intensities and thereby affecting the classification performance (Tantisatirapong, 2015). Consequently, to achieve accurate and objective classification of glioma grades, this work starts from the standpoint that the statistical analysis of the brain tumour descriptors integrated with machine learning algorithm can improve the quality of classification of glioma grades. The other standpoint that drives this work is the statistical analysis in relation

to contrast enhancement comparing to the presence of other tumour descriptors can play an important role in the differentiation of glioma grades. This creates the incentive to examine and measure different statistical ratios measured from these tumour descriptors and the influence thereof, on the classification of glioma grades. This work will investigate the difference between glioma grades in terms of the proposed statistical measure of the tumour descriptors and thus gain a better understanding to identify which one of the proposed measures is the most efficient and has the highest role if integrated with a machine learning algorithm to develop an automated classification system for glioma grades.

The automated classification system based on the proposed measures integrated with a machine learning algorithm is developed to classify different degrees of glioma with accurate and objective results. This also offers a reproducible methodology to determine the level of malignancy of brain tumour. This is the first study, which thoroughly investigates the influence of the statistical analysis of the proposed measures integrated with a machine learning algorithm on the automatic classification of glioma grade.

4.2   The Proposed Method

In this chapter, three standard segmented datasets were used, namely BRATS2013, BRATS2015 and BRATS2018 (Menze et al., 2015). These datasets have a labelled identification layer that was created previously. The identification layer is used to generate four masks and to individually bring in labelled regions, including necrosis, edema, non-enhanced and enhanced tumours.

Image's pixels that are included within each labelled region of the tumour descriptor are utilised to measure the presence of these descriptors. A search process is conducted to compute the total number of pixels in each region for all slices. This procedure is accomplished for all patients in the dataset. Then, an average of the results is performed according to each patient. As a result, four features are produced, namely *Nec_M, Edm_M, tnC_M*, and *tC_M*, which represent the presence of necrosis, edema, non-enhanced and enhanced tumours, respectively. These four features are defined by Eq. 4.1.

$$Descriptor\ presence\_M \ = \frac{1}{Z} \sum_{i=1}^{x} \sum_{j=1}^{y} \sum_{z=1}^{Z} \begin{cases} 1 & ,if\ SEG(x,y,Z) = Descritpor\ label \\ 0 & ,Otherwise \end{cases} \qquad 4.1$$

Where *SEG* is the label identification layer provided by the dataset, *Z* is the total number of MR slices that has tumour, *x* and *y* are the coordinates of the MR slices. *Descriptor presence_M* is the resultant average of the presence of tumour descriptors; for example, to measure the necrosis presence, this factor will be *necrosis_M* or *Nec_M* and the descriptor label will be the necrosis label provided by the dataset. As a result, four features are produced, namely *tC_M*, *tnC_M, Edm_M* and *Nec_M*, indicating contrast enhancement, non-enhancement, edema, and necrosis, respectively.

In addition, four new features are proposed based on measuring the ratio of the presence of each tumour descriptor with respect to the total summation of other appearances of tumour structures; these ratios are defined by Eq. (4.2) – (4.5).

$$tC\_R = \frac{tC\_M}{tC\_M + Nec\_M + Edm\_M + tnC\_M} \qquad 4.2$$

$$tnC\_R = \frac{tnC\_M}{tnC\_M + Nec\_M + Edm\_M + tnC\_M} \qquad 4.3$$

$$Edm\_R = \frac{Edm\_M}{tnC\_M + Nec\_M + Edm\_M + tnC\_M} \qquad 4.4$$

$$Nec\_R = \frac{Nec\_M}{tnC\_M + Nec\_M + Edm\_M + tnC\_M} \qquad 4.5$$

Where *tC_M*, *tnC_M, Edm_M* and *Nec_M* are the average presence of contrast enhancement, non-enhancement, edema and necrosis respectively, whereas *tC_R*, *tnC_R*, *Edm_R* and *Nec_R* are the resultant ratios of tumour enhancement, non-enhancement, edema and necrosis respectively.

4.3 Assessment of the Brain Tumour Descriptors

The statistical features extracted from the presence of the brain tumour descriptors including *Nec_M* (Figure 4.1)*, Edm_M* (Figure 4.2), *tnC_M* (Figure 4.3), *tC_M* (Figure 4.4) were analysed to assess the predictive power of these features on the diagnosis of glioma grades using the BRATS2013 dataset. The results of the presence of necrosis (*Nec_M*), edema (*Edm_M*), non-enhancement (t*nC_M*) and enhanced tumours (*tC_M*), were presented and investigated for both low and high grades of glioma. All the results are measured for low-grade glioma (from patient 1 to patient 10) and high-grade glioma (from patient 11 to patient 30). The x-axes in the figures represent the patient ID number given sequentially throughout the experiment where each number is associated with a patient. The numbers from 1 to 10 on the x-axis in the figure represent the low-grade samples and the numbers ranged from 11 to 30 are given to the high-grade samples.

The results show that apart from 1 and 2, most low-grade patients (1 to 10) did not develop contrast enhancement. Tumour necrosis was also absent in most patients in the low-grade group, while the high-grade group (11 to 30) developed all tumour descriptors (Figure 4.1 and Figure 4.4).
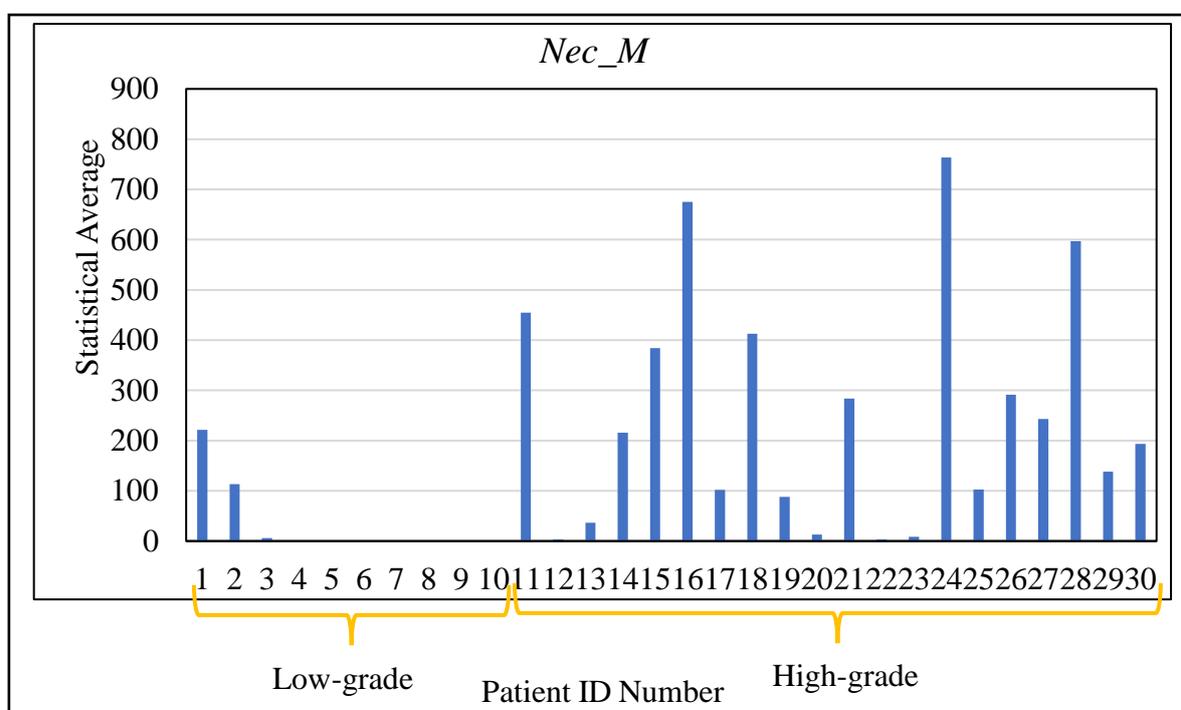
Figure 4.1 Feature results extracted from brain tumour descriptor (necrosis) in MR images.
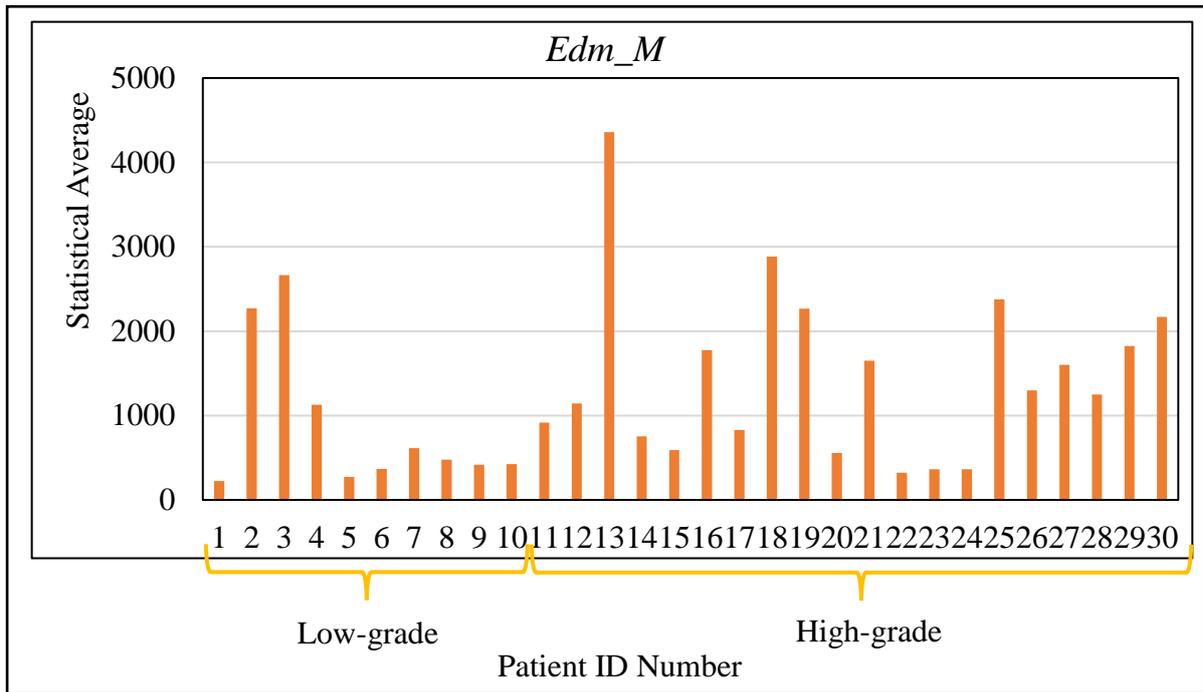
Figure 4.2 Feature results extracted from brain tumour descriptor (edema) in MR images.
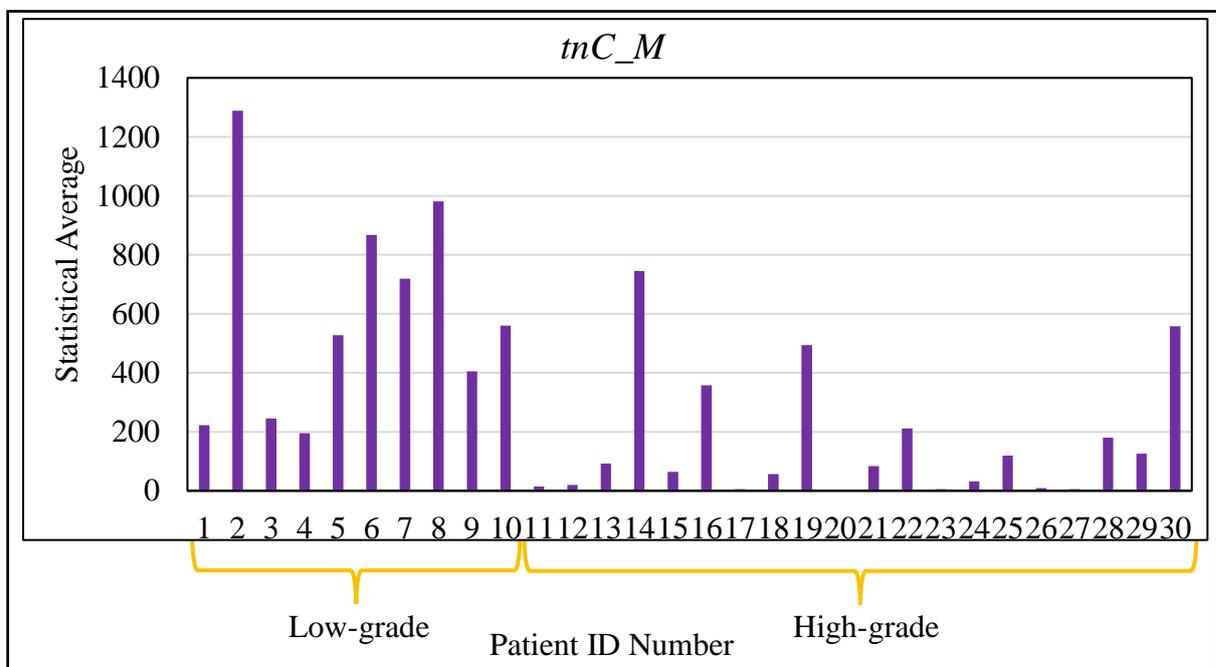
Figure 4.3 Feature results extracted from brain tumour descriptor (non-enhanced) in MR images.
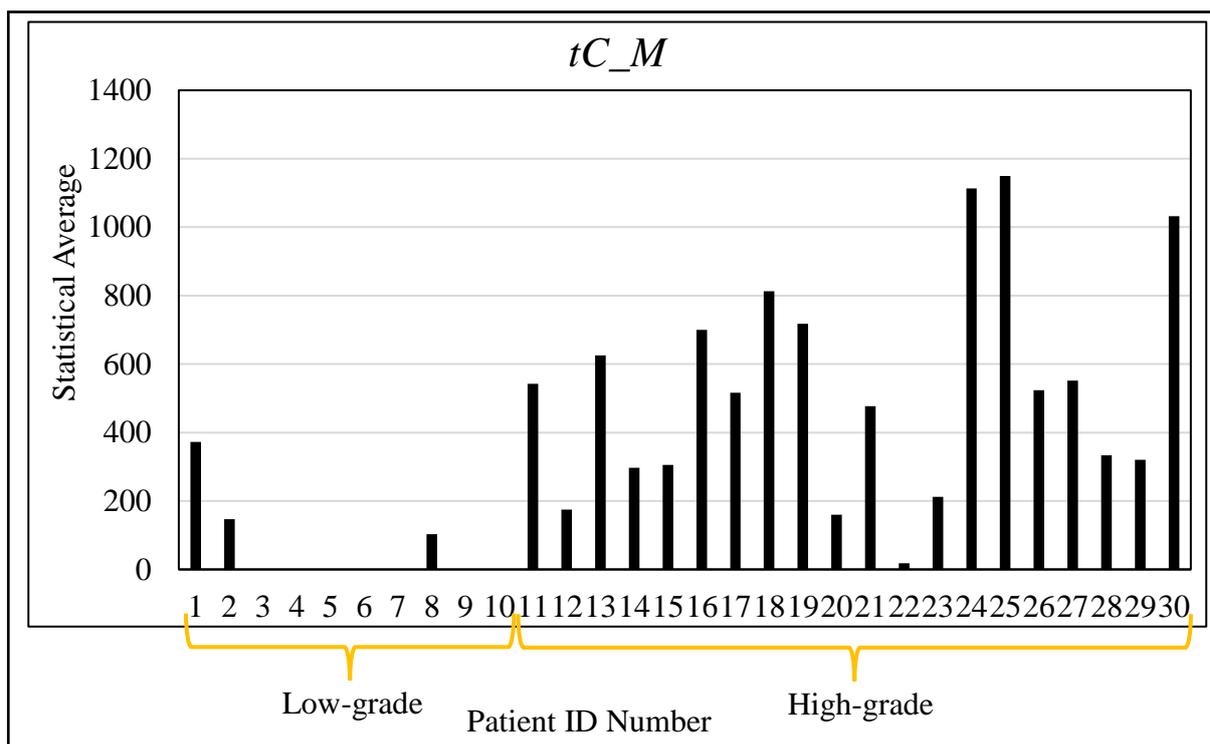
Figure 4.4 Feature results extracted from brain tumour descriptor (enhanced tumours) in MR images.

Accordingly, these features offer valuable indicators where unknown malignant brain tumours can be diagnosed as low-grade if there is neither contrast enhancement nor necrosis development. However, two patients belonging to the low-grade group behaved differently, these are patients 1 and 2 (Figure 4.1 and Figure 4.4). Consequently, it can be considered that these two descriptors are good features compared to the others in discriminating between low and high glioma grades, especially when there is no evidence for enhancement or necrosis in the MR image of a brain tumour. However, the decision can be disconcerted in some cases. For example, Patients 1 and 2 show tumour enhancement and necrosis at the same time while both subjects belong to the low-grade group (Figure 4.4). Therefore, the lesion can be mistakenly graded as high-grade as both low and high-grade tumours can develop an enhancement and necrosis. This is a considerable challenge for traditional visual diagnosis when assessing the tumour descriptors visually for glioma grading (Law et al., 2003, Thust et al., 2018). Consequently, this creates the motivation for further analysis on these features and assesses the predictive power of the integration of the proposed features with machine learning algorithm to determine an automated classification of glioma grades.

4.4 Results Analysis and Discussion

The feature classification stage was determined based on a supervised learning process where all patients were divided into both the training and testing phases based on the leave-one-out cross-validation technique. Eight generic features, namely *Nec_M, Edm_M, tnC_M, tC_M and tC_R, tnC, Edm_R, and Nec_R,* were extracted using the proposed method. The aim of this work is to investigate the ability of the automated classification system based on the proposed features in differentiating glioma grades into low and high grades, towards developing an automated, objective, and reproducible methodology to accurately determine glioma grades. Two experiments were conducted, and the results were analysed to answer the following questions:

- ➢ Which one of the tumour descriptors and proposed measures thereof, has the highest influence on developing an automated classification system for glioma grading?
- ➢ Which one of the classification models achieves better accuracy and is more appropriate for developing an automated classification system that can classify new data correctly?

These questions will be answered by evaluating the proposed system using three benchmark datasets. This will gain a better understanding of the methods and choices that are more valid in developing an automated system able to distinguish the degree of malignancy of glioma.

In this chapter, the same classifiers undertaken in previous work were used, namely LDA, SVM, KNN, DT, EBTree and ESDA, which were investigated with a comparison of the classification results. The feature selection was performed based on the proposed hybrid features selection algorithm (HFSA). The BRATS2013, BRATS2015 and BRATS2018 datasets were used to evaluate the classification system. The feature selection and classification based on the proposed features begin with applying the proposed automated system using the BRATS 2013 dataset. The full set of the proposed features, namely *Nec_M, Edm_M, tnC_M, tC_M, tC_R, tnC, Edm_R, and Nec_R,* were involved in the training and testing of all classifiers.

4.5 Evaluation of the Proposed Classification System Using the BRATS2013

In this experiment, the proposed automated classification system was evaluated using the BRATS2013 dataset and the results were then analysed. The experimental results obtained from testing the full set of the proposed features illustrate that the maximum classification performance in terms of classification accuracy, sensitivity, specificity, precision, and F-measure, was achieved by EBTree at 90%, 95%, 80%, 90.5%, and 92.68%, respectively (Table 4.2). While the next highest classification results, in terms of the same metrics, was obtained by both KNNF and KNNCOS classifiers at 86.67%, 90%, 80%, 90%, and 90%, respectively. This is followed by accuracies of 83.33% achieved by most of the other classifiers, such as LDA, SVML, SVMCUB, KNNW, and ESDA (Table 4.2). However, the outcome of DT classifiers shows a lower classification performance in terms of the evaluation metrics at 76.67%, 85%, 60%, 81%, and 82.92%, respectively (Table 4.2). In attempting to achieve the best possible improvement in the classification performance of the automated system, this work is extended by investigating the relevance analysis of the features involved in this experiment. Consequently, by using the ANOVA technique, the relevance analysis was applied to select the significant features in differentiating between low-grade glioma (I, II) and high-grade glioma (III, IV), while discarding others that are redundant. Results of the p-value using the ANOVA analysis (Table 4.1) indicate that the crucial features were *Nec_M, tnC_M, tC_M, tC_R, tnC_R,* and *Nec_R*. while *Edm_M and Edm_R* were discarded, as they did not meet the significance level at 0.05. This also indicates that the tumour edema (*Edm_M, and Edm_R*) showed no difference between the low and high-grade glioma and it has no role in differentiation between the two classes of glioma grades.

Table 4.1 P-value results of applying ANOVA to the full set of the proposed features extracted from the tumour descriptors using BRATS2013.

Feature	P-value	Feature	P-value
Nec_M	0.007344	*Nec_R*	0.02895
Edm_M	0.122108	*Edm_R*	0.16485
tnC_M	0.000292	*tnC_R*	0.00000
tC_M	0. 000012	*tC_R*	0.00016

Table 4.2 Comparative results of different classifiers incorporated with the full set of the proposed features associated with the tumour descriptors using the BRTAS2013 dataset. Class0 refers to low-grade glioma (I, II) and Class1 indicates high-grade glioma (III, IV).

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	6	4	60.00	66.70	63.15	76.67
	Class1	3	17	85.00	81.00	82.92	
LDA	Class0	7	3	70.00	77.80	73.68	83.33
	Class1	2	18	90.00	85.70	87.80	
SVML	Class0	7	3	70.00	77.80	73.68	83.33
	Class1	2	18	90.00	85.70	87.80	
SVMQ	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
SVMCUB	Class0	8	2	80.00	72.70	76.19	83.33
	Class1	3	17	85.00	89.50	87.17	
SVMG	Class0	6	4	60.00	75.00	66.66	80.00
	Class1	2	18	90.00	81.80	85.71	
KNNF	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
KNNM	Class0	7	3	70.00	77.80	73.68	83.33
	Class1	2	18	90.00	85.70	87.80	
KNNCOS	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
KNNCUB	Class0	7	3	70.00	77.80	73.68	83.33
	Class1	2	18	90.00	85.70	87.80	
KNNW	Class0	7	3	70.00	77.80	73.68	83.33
	Class1	2	18	90.00	85.70	87.80	
EBTree	Class0	8	2	80.00	88.90	84.21	**90.00**
	Class1	1	19	95.00	90.50	92.68	
ESDA	Class0	7	3	70.00	77.80	73.68	83.33
	Class1	2	18	90.00	85.70	87.80	

The aim of this experiment is to eliminate the redundancy in the feature space and only keep the significant features. Therefore, a relevance analysis was conducted based on the ANOVA method. It is possible to investigate the level of significance whereby important features can be selected and others discarded. This can potentially lead to either improving the classification accuracy of the automated system or reducing the dimensions of the feature space, leading to further enhancement in the classification performance. The selected features by the ANOVA method were then used to train and test all the classifiers. This is to investigate the behaviour of these features through the automated classification of glioma grades based on different machine learning algorithms.

The results obtained from examining the automated system based on the selected set of features by ANOVA illustrate that the KNNF classifier outperforms all other classification methods in terms of classification accuracy, sensitivity, specificity, precision, and F-measure at 93.33%, 95%, 90%, 95%, 95%, respectively (Table 4.3). The next highest classification accuracies were achieved at 90% by the classifiers: SVMCUB, KNNCOS, and KNNW. However, both KNNW and KNNCOS showed the same classification performance in terms of the evaluation metrics at 90%, 90%, 90%, 94.7%, and 92.3%, respectively (Table 4.3). Meanwhile, there were trade-offs in the sensitivity and specificity between SVMCUB and both KNNCOS and KNNW classifiers, where the outcome of SVMCUB in terms of the evaluation metrics was 90%, 95%, 80%, 90.5%, 92.68% respectively (Table 4.3). This is followed by a lower classification accuracy of 86.67% achieved by both LDA and SVMQ. Nevertheless, there is a trade-off in the sensitivity and specificity between LDA and SVMQ classifiers, where LDA achieved higher sensitivity at 90% and lower specificity at 80%, while SVMQ obtained lower sensitivity at 85% and higher specificity at 90% (Table 4.3). The results of testing both ensemble methods indicate that they achieved a lower classification accuracy at 83.33%. The outcomes of the other classifiers, such as KNNCUB and SVML, have shown the same level of classification performance in terms of classification accuracy at 83.33%. The results indicate that the DT classifier achieved the lowest classification accuracy at 76.67% (Table 4.3).

Table 4.3 Comparative results of different classifiers incorporated with the selected set of features extracted from the tumour descriptors by ANOVA using the BRTAS2013 dataset.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
DT	Class0	6	4	60.00	66.7	63.15	76.67
	Class1	3	17	85.00	81.0	82.92	
LDA	Class0	8	2	80.00	80.0	80.00	86.67
	Class1	2	18	90.00	90.0	90.00	
SVML	Class0	7	3	70.00	77.8	73.68	83.33
	Class1	2	18	90.00	85.7	87.80	
SVMQ	Class0	9	1	90.00	75.0	81.81	86.67
	Class1	3	17	85.00	94.4	89.47	
SVMCUB	Class0	8	2	80.00	88.9	84.21	90.00
	Class1	1	19	95.00	90.5	92.68	
SVMG	Class0	6	4	60.00	75.0	66.66	80.00
	Class1	2	18	90.00	81.8	85.71	
KNNF	Class0	9	1	90.00	90.0	90.00	**93.33**
	Class1	1	19	95.00	95.0	95.00	
KNNM	Class0	9	1	90.00	81.8	85.71	90.00
	Class1	2	18	90.00	94.7	92.30	
KNNCOS	Class0	9	1	90.00	81.8	85.71	90.00
	Class1	2	18	90.00	94.7	92.30	
KNNCUB	Class0	7	3	70.00	77.8	73.68	83.33
	Class1	2	18	90.00	85.7	87.80	
KNNW	Class0	9	1	90.00	81.8	85.71	90.00
	Class1	2	18	90.00	94.7	92.30	
EBTree	Class0	7	3	70.00	77.8	73.68	83.33
	Class1	2	18	90.00	85.7	87.80	
ESDA	Class0	7	3	70.00	77.8	73.68	83.33
	Class1	2	18	90.00	85.7	87.80	

To investigate the influence of each of the proposed features on the automated classification of glioma grades using machine learning algorithms, each one of the proposed features were individually integrated and tested with different machine learning algorithms. The results illustrate that the proposed features reflect varies performance in terms of classification accuracy when used with a different classification model (Figure 4.5). However, the best feature that showed the maximum classification accuracy for most of the classifiers was the *tC_R* predictor, which achieved 93.3% when examined with a single classifier system, including: LDA, SVML, KNNM, and KNNCUB. Furthermore, the same accuracy was also achieved when it was tested with the ensemble system by ESDA classifier (Figure 4.5). The predictor *tC_M* achieved the same classification accuracy of 93.3% when examined with the SVML classifier. This was followed with 90% obtained by *tC_M* when tested with both KNNM and KNNCUB, whilst tumour edema reflected the lowest behaviour compared to all other features. All the other predictors, including: *Nec_M, Edm_M, tnC_M, Edm_R*, and *Nec_R,* achieved lower accuracies below 90% when they were examined with all classifiers (Figure 4.5).

In comparison to the P-values obtained from applying the ANOVA method (Table 4.1), predictor *tC_R* appeared as a significant feature, but it was not the best one. Meanwhile, other features, such as *tnC_R,* showed a better significance level (0.0000020) in the P-value table (Table 4.1). However, an investigation into automated classification for glioma grades based on machine learning showed that the predictor *tC_R* achieved the best classification accuracy (Figure 4.5).

For further potential enhancement in the classification performance of the automated system, additional investigating into the features space was conducted to find the best subset of features that can achieve the optimal accuracy with the lowest possible dimension of features and thus, the proposed HFSA was applied and examined. It was noted that *tC_R* outperformed all other proposed features when it was examined by many classifiers; therefore, it was selected as a reference predictor. Then, the proposed HFSA was implemented to select the most crucial subset of features. The selected features chosen by ANOVA were then ranked against *tC_R*. The proposed method was implemented automatically based on tracking the output results of different classification methods, which are trained and tested for all possible feature subsets.
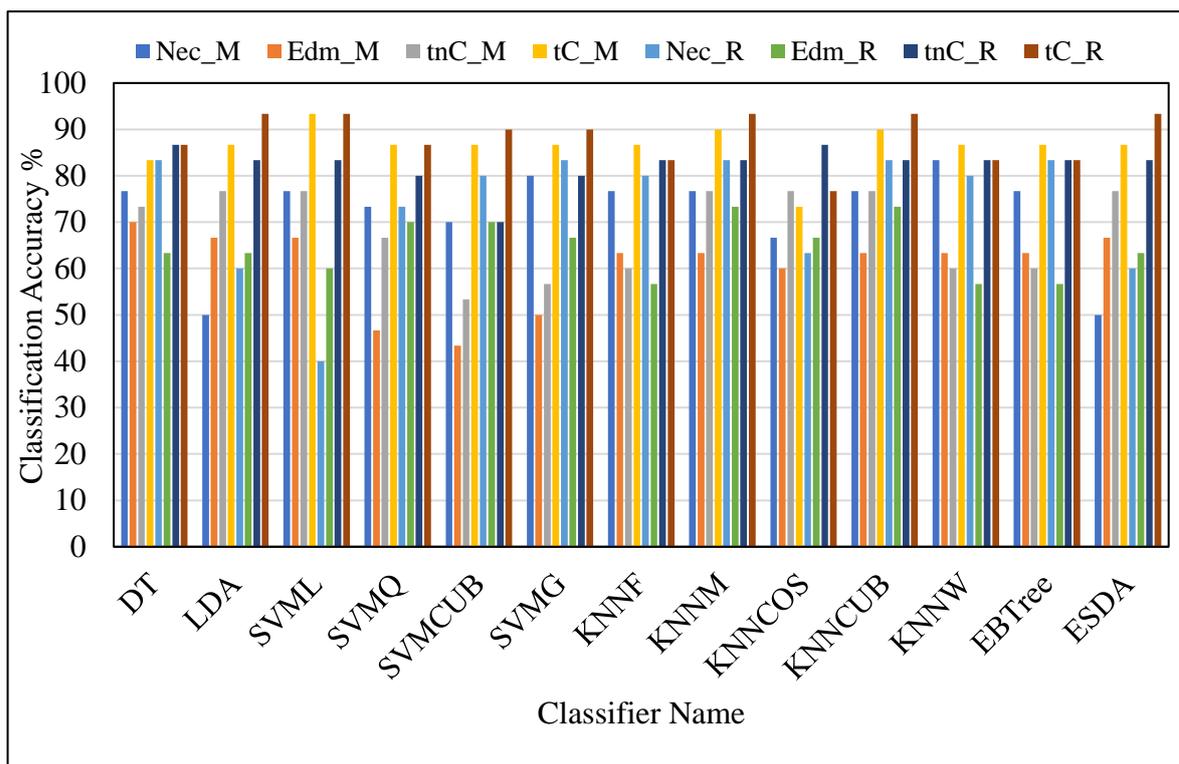
Figure 4.5 Comparative results in terms of classification accuracy for training and testing different classifiers incorporated individually with the proposed predictors. This test used BRATS2013 dataset.

According to the output of the proposed HFSA, the predictor *tC_R* was selected as the most significant feature, signifying that the predictor *tC_R* achieved the best classification accuracy (Table 4.4). This indicates that all subsets of features have not achieved any further improvement when examined by the proposed HFSA. The results obtained from applying the proposed HFSA illustrate that the best classification performance in terms of classification accuracy, sensitivity, specificity, precision, and F-measure were 93.33%, 95%, 90%, 95%, and 95% respectively, which were achieved by several classifiers, including LDA, SVML, KNNM, KNNCUB, KNNM, KNNCUB, and ESDA (Table 4.4). The next best results in terms of the evaluation metrics was achieved at 90%, 95%, 80%, 90.5%, 92.68% respectively, by both SVMCUB and SVMG. This is followed by a classification accuracy of 86.67%, achieved by both DT and SVMQ. However, KNNF, KNNW, and EBTree classifiers revealed lower classification accuracy at 83.33%. The lowest classification accuracy was achieved by KNNCOS at 76.67% (Table 4.4).

Table 4.4 Comparative results for training and testing different classifiers based on the selected set of features chosen by the proposed HFSA using the BRATS2013 dataset. The selected feature is *tC_R*, which represents the ratio of presence of the tumour enhancement compared to other tumour descriptors.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
DT	Class0	7	3	70.00	87.5	77.77	86.67
	Class1	1	19	95.00	86.4	90.47	
LDA	Class0	9	1	90.00	90.0	90.00	**93.33**
	Class1	1	19	95.00	95.0	95.00	
SVML	Class0	9	1	90.00	90.0	90.00	**93.33**
	Class1	1	19	95.00	95.0	95.00	
SVMQ	Class0	8	2	80.00	80.0	80.00	86.67
	Class1	2	18	90.00	90.0	90.00	
SVMCUB	Class0	8	2	80.00	88.9	84.21	90.00
	Class1	1	19	95.00	90.5	92.68	
SVMG	Class0	8	2	80.00	88.9	84.21	90.00
	Class1	1	19	95.00	90.5	92.68	
KNNF	Class0	8	2	80.00	72.7	76.19	83.33
	Class1	3	17	85.00	89.5	87.17	
KNNM	Class0	9	1	90.00	90.0	90.00	**93.33**
	Class1	1	19	95.00	95.0	95.00	
KNNCOS	Class0	9	1	90.00	60.0	72.00	76.67
	Class1	6	14	70.00	93.3	80.00	
KNNCUB	Class0	9	1	90.00	90.0	90.00	**93.33**
	Class1	1	19	95.00	95.0	95.00	
KNNW	Class0	8	2	80.00	72.7	76.19	83.33
	Class1	3	17	85.00	89.5	87.17	
EBTree	Class0	8	2	80.00	72.7	76.19	83.33
	Class1	3	17	85.00	89.5	87.17	
ESDA	Class0	9	1	90.00	90.0	90.00	**93.33**
	Class1	1	19	95.00	95.0	95.00	

Overall comparative results for the selected feature chosen by the proposed HFSA against both the full set of the features and the selected features chosen by ANOVA, indicates that the classification accuracy was improved using ANOVA as well as the proposed HFSA (Figure 4.6). For the ANOVA implementation, it was noted that the accuracies enhanced from 83.33% to 90% by both SVMCUB and the KNNM classifier, while the biggest improvement achieved with the use of the KNNF, where the classification accuracy went up from 86.66% to 93.33%. When the proposed HFSA was implemented, the selected feature integrated with the machine learning algorithm achieved an improvement in the classification accuracies for most classifiers (Figure 4.6). For example, compared to ANOVA, the classification accuracies were enhanced from 76.66% to 86.66% by DT, from 86.66% to 93.33% by LDA, from 83.33% to 93.33% by SVML, from 90% to 93.33% by KNNM, and from 83.33% to 93.33% by KNNCUB. While the classification accuracy in terms of the ensemble methods also achieved better results with the ESDA classifier, where the accuracy enhanced from 83.33% to 93.33%.

The results of evaluating the classification performance in terms of classification accuracy for the automated classification system using BRATS2013 indicate that both ANOVA and the proposed HFSA have shown the same maximum accuracies of 93.33%. However, the proposed HFSA achieved better improvement at 93.33% for most classifiers, while ANOVA showed this improvement for only one classifier, indicating that the selected feature by the proposed HFSA is more appropriate and clearer from redundant features, therefore it can achieve better enhanced results using a wide range of different classification models. The other advantage of the proposed HFSA is that it achieved a significant reduction in the features space where the dimension of the feature space was reduced from eight features to only one feature, namely *tC_R*. While the selected set of features by ANOVA were six features, namely *Nec_M, tnC_M, tC_M, tC_R, tnC_R,* and *Nec_R.* To summarise, although both the ANOVA and the proposed HFSA have shown the same enhancement in the classification accuracy based on the proposed features, the proposed HFSA achieved the best reduction in the feature space and maintain a good improvement in the classification accuracy for glioma grades.
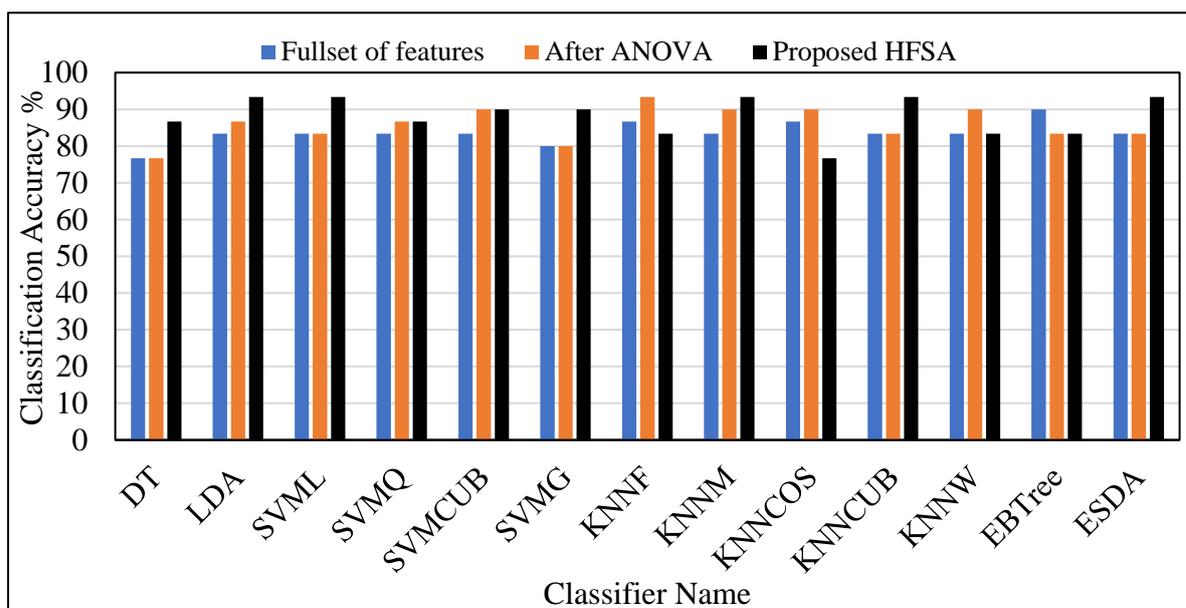
Figure 4.6 Overall comparative results showing the behaviour of the proposed features incorporated with different classifiers using BRATS2013. These features are *Nec_M, Edm_M, tnC_M, tC_M, tC_R, tnC, Edm_R*, and Nec_R. The results were compared with the selected set of features by ANOVA; these are *Nec_M, tnC_M, tC_M, tC_R, tnC_R, and Nec_R*, compared with the feature selected by the proposed HFSA; the selected feature was *tC_R*.

4.6 Evaluating the Proposed Classification System Using the BRATS2015

The automated system to classify glioma grades based on proposed features and incorporated machine learning algorithms was developed using the BRATS2015 dataset. The aim of this experiment is to add further and general evaluation to the automated classification system for glioma grades where the same previous framework implemented with BRATS2013 dataset was repeated using the BRATS2015 dataset.

When the full set of the proposed set of features were examined using the single classifier system, the results indicated that the highest classification performance in terms of classification accuracy, sensitivity, specificity, precision, and F-measure, was achieved by the SVML classifier at 89.42%, 98.64%, 51.85%, 89.3%, 93.73%, respectively compared to all other single classifiers (Table 4.5). This was followed by the KNNW classifier that achieved a slightly lower classification accuracy of 89.05%. The KNNM classifier also achieved the second highest accuracy of 87.59%. Whilst, when the proposed HFSA was tested with the ensemble system, the results indicated that the EBTree classifier achieved the highest classification result, outweighing all other classifiers in this experiment at 90.15%, with lower sensitivity at 95.45% and higher specificity at 68.52%, and with a very small difference in F-measure at 93.95% compared to SVML classifier (Table 4.5). The results of

all other classifiers in terms of classification accuracy ranged from 84.31% to 86.50% (Table 4.5).

Table 4.5 Comparative evaluation results showing the full set of the proposed features extracted from the tumour descriptors incorporating different machine learning algorithms using the BRATS2015 dataset.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
DT	Class0	21	33	38.89	84.0	53.16	86.50
	Class1	4	216	98.18	86.7	92.11	
LDA	Class0	33	21	61.11	75.0	67.34	88.32
	Class1	11	209	95.00	90.9	92.88	
SVML	Class0	28	26	51.85	90.3	65.88	89.42
	Class1	3	217	98.64	89.3	93.73	
SVMQ	Class0	27	27	50.00	75.0	60.00	86.86
	Class1	9	211	95.91	88.7	92.13	
SVMCUB	Class0	28	26	51.85	70.0	59.57	86.13
	Class1	12	208	94.55	88.9	91.63	
SVMG	Class0	18	36	33.33	90.0	48.64	86.13
	Class1	2	218	99.09	85.8	91.98	
KNNF	Class0	31	23	57.41	60.8	59.04	84.31
	Class1	20	200	90.91	89.7	90.29	
KNNM	Class0	27	27	50.00	79.4	61.36	87.59
	Class1	7	213	96.82	88.8	92.60	
KNNCOS	Class0	25	29	46.30	71.4	56.18	85.77
	Class1	10	210	95.45	87.9	91.50	
KNNCUB	Class0	25	29	46.30	75.8	57.47	86.50
	Class1	8	212	96.36	88.0	91.97	
KNNW	Class0	30	24	55.56	83.3	66.66	89.05
	Class1	6	214	97.27	89.9	93.44	
EBTree	Class0	37	17	68.52	78.7	73.26	90.15
	Class1	10	210	95.45	92.5	93.95	
ESDA	Class0	26	28	48.15	78.8	59.77	87.23
	Class1	7	213	96.82	88.4	92.40	

To investigate the relevance analysis of the proposed features, the ANOVA method is applied. The significant levels obtained after applying the ANOVA technique indicates that only four features were significant, namely *tnC_M, tC_M, tnC_R and tC_R,* while the others were discarded (Table 4.6). It was noted that the power of the significance level is increased where the predictor *tC_M* shows the best significance level at 1.6311×10^{-24}, followed by *tC_R* at 3.8146×10^{-18}. This is due to the large number of samples of datasets used in this experiment, supporting the ANOVA method to show higher values of significance P-levels.

Table 4.6 Significance levels for features extracted from brain tumour descriptors after applying the ANOVA method using the BRATS2015 dataset.

Feature	P-value	Feature	P-value
Nec_M	0.168299	*Nec_R*	0.095622
Edm_M	0.058754	*Edm_R*	0.125198
tnC_M	9.9578×10^{-8}	*tnC_R*	2.1213×10^{-9}
tC_M	1.6311×10^{-24}	*tC_R*	3.8146×10^{-18}

After feeding the feature selected by the ANOVA method to the classification system based on training and testing different classifiers, the results illustrate that both SVML and EBTree classifiers achieved the best results in terms of classification accuracy at 89.05% (Table 4.7). However, there is trade-off in the sensitivity and specificity between them, where SVML achieved a higher sensitivity at 98.18% and lower specificity at 51.85%, while EBTree gained a lower sensitivity at 95.4%, with higher specificity at 62.96%. However, the SVML classifier showed a slightly better accuracy than the EBTree classifier in terms of F-measure where they achieved 93.50% and 93.33%, respectively. The KNNW classifier achieved the second-best classification accuracy of 88.69%.

Table 4.7 Comparative evaluation results for the selected set of features using the ANOVA method, incorporating different classifiers using the BRATS2015 dataset.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	21	33	38.89	84.0	53.16	86.50
	Class1	4	216	98.18	86.7	92.11	
LDA	Class0	33	21	61.11	70.2	65.34	87.23
	Class1	14	206	93.64	90.7	92.17	
SVML	Class0	28	26	51.85	87.5	65.11	89.05
	Class1	4	216	98.18	89.3	93.50	
SVMQ	Class0	27	27	50.00	84.4	62.79	88.32
	Class1	5	215	97.73	88.8	93.07	
SVMCUB	Class0	28	26	51.85	71.8	60.21	86.50
	Class1	11	209	95.00	88.9	91.86	
SVMG	Class0	21	33	38.89	91.3	54.54	87.23
	Class1	2	218	99.09	86.9	92.56	
KNNF	Class0	36	18	66.67	72.0	69.23	88.32
	Class1	14	206	93.64	92.0	92.79	
KNNM	Class0	31	23	57.41	66.0	61.38	85.77
	Class1	16	204	92.73	89.9	91.27	
KNNCOS	Class0	32	22	59.26	61.5	60.37	84.67
	Class1	20	200	90.91	90.1	90.49	
KNNCUB	Class0	32	22	59.26	68.1	63.36	86.50
	Class1	15	205	93.18	90.3	91.72	
KNNW	Class0	34	20	62.96	75.6	68.68	88.69
	Class1	11	209	95.00	91.3	93.09	
EBTree	Class0	34	20	62.96	77.3	69.38	89.05
	Class1	10	210	95.45	91.3	93.33	
ESDA	Class0	25	29	46.30	78.1	58.14	86.86
	Class1	7	213	96.82	88.0	92.20	

To investigate the effectiveness of each one of the proposed predictors on the classification performance of glioma grades, as well as to identify which one of the proposed features shows the highest classification accuracy, a comparative evaluation is conducted using the BRATS2015 dataset. This is performed by examining the outcome of each predictor incorporated with the machine learning algorithms undertaken in this work (Figure 4.7). As a result, the predictor *tnC_M* reflects the best classification accuracy of 86.66% when tested with most of the classifiers, including LDA, KNNM, KNNCUB, and ESDA (Figure 4.7). The predictor *tC_R* has also achieved a competitive accuracy of 85.76% when examined with both the KNNM and KNNCUB classifiers. The predictor *tnC_R* has also shown the next dominant accuracy of 85.03% when evaluated with the same classifiers. Following these results, the highest classification accuracy achieved by the predictor *tC_M* is 82.11%, with both the KNNM and KNNCUB classifiers. However, the features related to tumour necrosis have reflected lower accuracies whereby the best accuracies achieved were with DT at 80.65% and LDA at 80.29%. The ESDA classifier also achieved the same accuracy of 80.29%, using necrosis features. Likewise, all the proposed features associated with tumour edema have shown lower accuracies whereby the averages of their results are around 80% (Figure 4.7). The classifiers that show the best accuracies using features from tumour edema are LDA, SVML, and EBTree, where their results are around 80.29%, while all other classification models show lower accuracies below 80%.

Consequently, it is evident that the best predictors are *tnC_M* and *tC_R* due to their superior accuracy when compared to all other features associated with tumour descriptors when they are integrated with the machine learning algorithm. Accordingly, it can draw the inference that the automated system based on presence of the statistical predictors related to the tumour descriptors can be a valuable prediction system for glioma grades when it is integrated with machine learning algorithms. However, the automated systems based on either the feature associated with non-enhancement or the ratio related to contrast enhancement, are the most important predictors due to their remarkable results that outweigh all other tumour descriptors in the automated classification of glioma grades based on machine learning strategies.
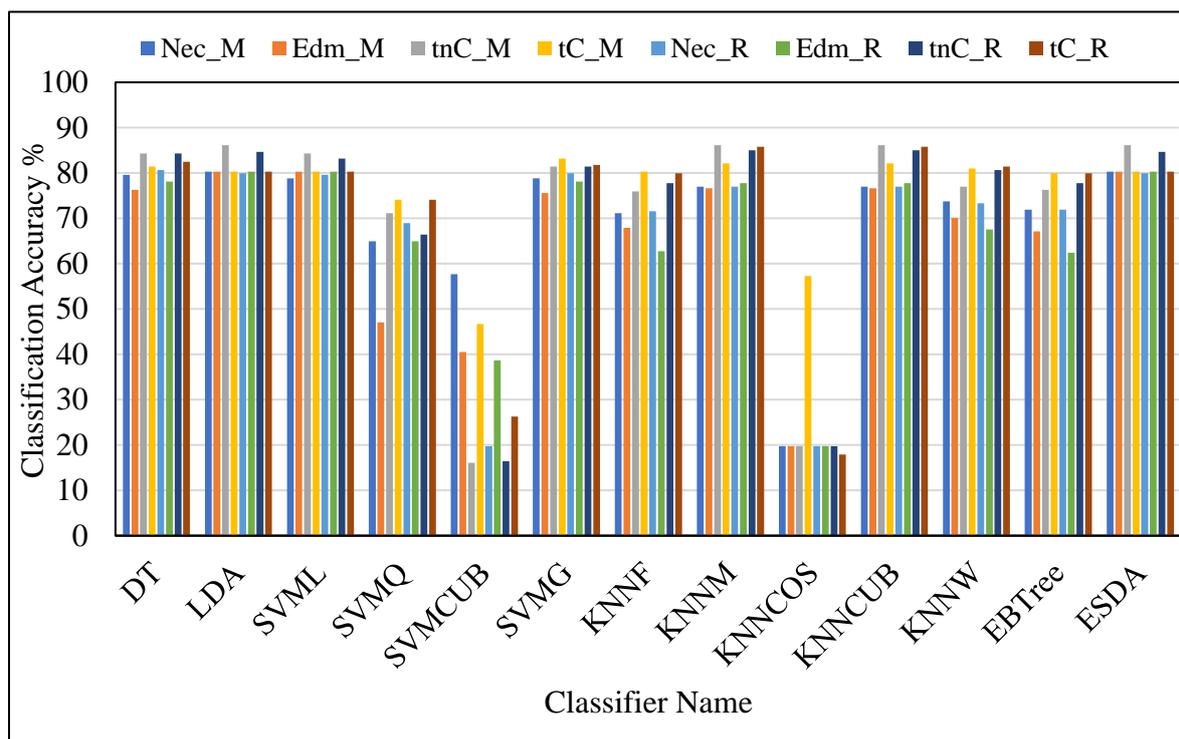
Figure 4.7 Comparative results in terms of classification accuracy for training and testing different classifiers using the proposed predicators using the BRATS2015 dataset.

Notably, the predictor *tnC_M* achieved the best classification accuracy for glioma grades in this experiment. Therefore, it has been chosen to be the reference feature for the initialisation of the proposed HFSA. The selected four features by ANOVA were *tnC_M, tC_M, tnC_R, and tC_R*. After demonstrating the proposed HFSA, the selected set of features were only two: *tnC_M, and tC_R*. The best classification performance in terms of classification accuracy, sensitivity, specificity, precision, and F-measure was found by the KNNW classifier at 90.51%, 96.86%, 64.82%, 91.8%, 94.24% (Table 4.8). The second-best results were achieved by the EBTree classifier at 89.42%, 96.36%, 61.11%, 91.0%, and 93.59%, respectively. The maximum sensitivity achieved in this experiment, however, was obtained by ESDA at 99.09%, with a lower classification accuracy at 87.59%. The results of all other classifiers in terms of classification accuracy ranged between 83.58% (achieved by KNNCOS) and 89.05% (achieved by different classifiers, such as DT, LDA, and SVML).

Table 4.8 Comparative evaluation results for the selected set of features by the proposed HFSA, incorporating different classifiers using the BRATS2015 dataset. The selected features were *tnC_M and tC_R.*

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	27	27	50.00	90.0	64.28	89.05
	Class1	3	217	98.64	88.9	93.53	
LDA	Class0	29	25	53.70	85.3	65.90	89.05
	Class1	5	215	97.73	89.6	93.47	
SVML	Class0	28	26	51.85	87.5	65.11	89.05
	Class1	4	216	98.18	89.3	93.50	
SVMQ	Class0	26	28	48.15	81.3	60.46	87.59
	Class1	6	214	97.27	88.4	92.64	
SVMCUB	Class0	28	26	51.85	73.7	60.86	86.86
	Class1	10	210	95.45	89.0	92.10	
SVMG	Class0	22	32	40.74	88.0	55.69	87.23
	Class1	3	217	98.64	87.1	92.53	
KNNF	Class0	39	15	72.22	67.2	69.64	87.59
	Class1	19	201	91.36	93.1	92.20	
KNNM	Class0	30	24	55.56	73.2	63.15	87.23
	Class1	11	209	95.00	89.7	92.27	
KNNCOS	Class0	29	25	53.70	59.2	56.31	83.58
	Class1	20	200	90.91	88.9	89.88	
KNNCUB	Class0	32	22	59.26	76.2	66.66	88.32
	Class1	10	210	95.45	90.5	92.92	
KNNW	Class0	35	19	64.81	83.3	72.91	90.51
	Class1	7	213	96.82	91.8	94.24	
EBTree	Class0	33	21	61.11	80.5	69.47	89.42
	Class1	8	212	96.36	91.0	93.59	
ESDA	Class0	22	32	40.74	91.7	56.41	87.59
	Class1	2	218	99.09	87.2	92.76	

An overall comparative evaluation has been conducted for all the classifiers based on the selected features by the proposed HFSA against both the use of the full set of features and the selected set by ANOVA (Figure 4.8). The experimental results indicate that the classification performance in terms of classification accuracy was improved using the ANOVA method compared to the use of the full set of features. For example, the accuracy was enhanced from 86.86% to 88.32% by SVMQ and from 86.13% to 87.22% by SVMG (Figure 4.8). It is noted that the highest improvement using ANOVA was obtained by the KNNF classifier where the accuracy enhanced from 84.3% to 88.32%. However, there are some other classifiers that show lower accuracies when examined with the selected set of features by ANOVA. For example, in terms of a single classifier system, the outcome of the LDA and KNNM classifiers illustrates a small reduction in the classification accuracy, where they reduced from 88.32% to 87.22% and from 87.59% to 85.76%, respectively. In terms of the ensemble classification method, there is also a minor reduction in classification accuracy shown by both EBTree and ESDA classifiers as their classification accuracies decreased from 90.14% to 89.05% and from 87.22% to 86.86%, respectively. Nevertheless, the results obtained from applying the proposed HFSA reveals that the classification accuracy improved when compared to both the use of the full set of features and the use of the selected set by the ANOVA method. Notably, many classifiers achieved an improvement using the selected set of features by the proposed HFSA. For example, the accuracies of DT, LDA SVMQ, and KNNF were enhanced from 86.49% to 89.05%, from 88.32% to 89.05%, from 86.86% to 88.32%, and from 84.3% to 87.59%, respectively. However, the results illustrate that the best classification accuracy among all methods and choices was obtained by the selected set of features by the proposed HFSA at 90.5% by the KNNW classifier (Figure 4.8).

The other advantage of applying the proposed HFSA is achieving a significant reduction in the features space where the dimensions of the input features were reduced from eight to only two features, namely tnC_M and tC_R, while ANOVA applied the input features that were reduced into four features, namely tnC_M, tC_M, tnC_R and tC_R. Therefore, a significant enhancement is achieved in developing the classification system based on the integration between the proposed features and the use of the proposed HFSA. The results reveal that this method gains a better reduction in features space as well as higher classification accuracy for glioma grades.
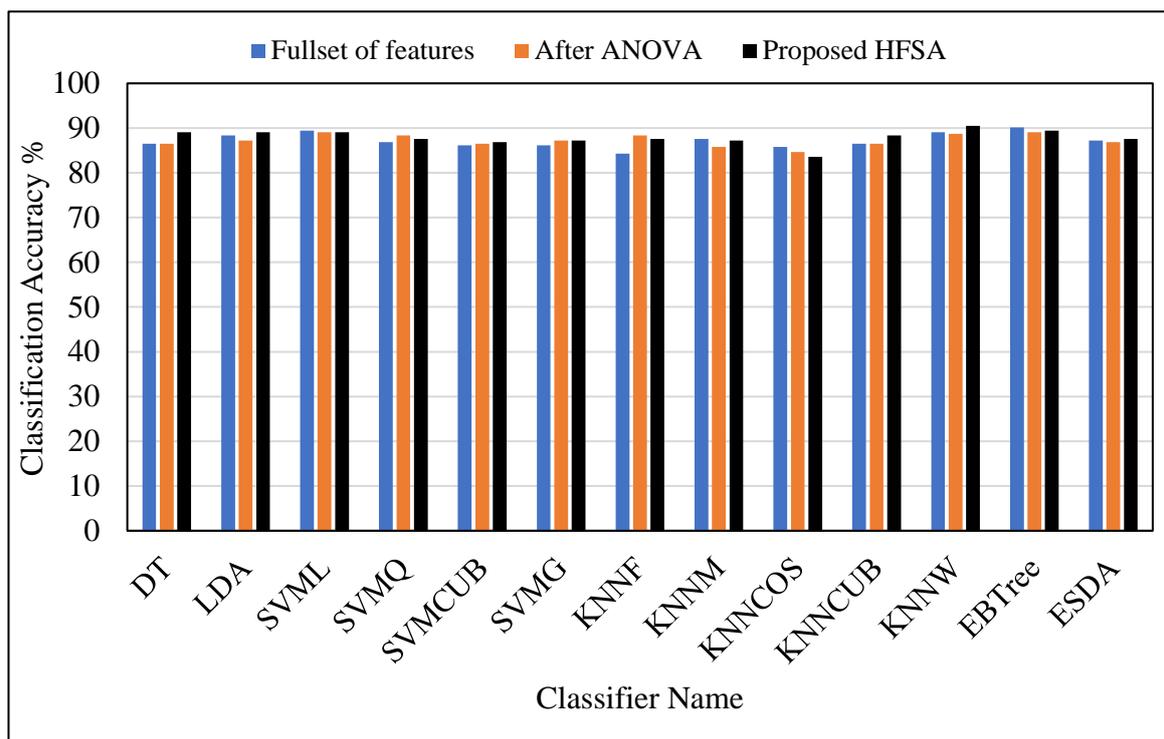
Figure 4.8 Overall comparative results to show the behaviour of the automated classification system based on three cases as follows: the first case is based on the full set of the proposed features, the second case is based on the features selected by ANOVA, and the final case is on using the selected set of features by the proposed HFSA. This experiment is developed using the BRATS2015 dataset.

4.7 Evaluating the Proposed Classification System Using the BRATS2018

The automated classification system for glioma grades based on the proposed features and incorporated different machine learning algorithms is implemented using the BRATS2018 dataset. The rationale for this experiment is to add further evaluation and validation to the classification system.

After extract the proposed features associated with tumour descriptors from BRATS2018 dataset, it was found experimentally that the features associated with non-contrast enhancement reveal zero-values. This indicates that the contrast enhancement in this dataset has covered all the active portion of the brain tumour area, which also reflects different tumour structure compared to the other BRATS datasets. The absence of the non-enhanced area from the tumour is due to the nature of the brain tumour, which indicates greater infiltration of the malignant brain tumour to the surrounding parts, and the large leakage of

blood-brain barrios compared to other BRATS datasets. Accordingly, two features are found to be zeros, these are *TnC_M*, and *TnC_R* and therefore they were discarded from the model design in this dataset.

The implementation starts by measuring the proposed features including *Nec_M*, *Edm_M*, *tC_M*, *Nec_R*, *Edm_R*, tC_R from ROI of MR brain tumour images for all patients in the dataset then the same framework that is used in this work is developed for this dataset starting with splitting the dataset into training and testing sets. The training set is used to train the classification models that include single and ensemble classifiers then the trained models are used to test the testing samples based on LOO-cross validation technique. As a result, the classification performances of all models are measured.

Three cases are implemented to investigate the importance of the proposed features incorporating machine learning on the classification accuracy of glioma grades. The first case is conducted by using the full set of the proposed features mentioned above to train and test the classification models. The second case is implemented based on investigating the impact of ANOVA technique incorporating the proposed features on the classification accuracy of glioma grades. The third case is developed based on examining the impact of the proposed HFSA integrating with the proposed features on glioma grading. The results obtained from the first case illustrated that the best classification performance in term of classification accuracy is achieved by SVML classifiers at 91.58%. This was followed by KNNM classifier that achieved 90.53%. Ensemble systems have shown lowest classification accuracy at 90.88% and 90.18% by EBTree and ESDA respectively (APPENDIX C, Table C.5). Considering the significance level of P-value that achieved by applying ANOVA technique (Table 4.9) and after discarding features less than 0.05, *Edm_R* was discarded, and only five features were significant, namely *Nec_M*, *tC_M*, *Nec_R*, *Edm_R*, *tC_R*.

Table 4.9 Significance levels for features extracted from brain tumour descriptors after applying the ANOVA method using the BRATS2018 dataset.

Feature	P-value	Feature	P-value
Nec_M	5.5905×10^{-12}	*Nec_R*	1.2404×10^{-15}
Edm_M	0.1439	*Edm_R*	0.0001
tC_M	1.6720×10^{-25}	*tC_R*	9.0504×10^{-26}

After feeding the feature selected by the ANOVA method to the classification system based on training and testing different classifiers, the results showed that the highest classification accuracy achieved by SVMQ classifier at 91.93%, followed by SVMCUB classifier with a slightly lower accuracy at 91.23% (APPENDIX C Table C.6).

To investigate the impact of each one of the proposed predictors on the classification performance of glioma grades, as well as to identify which one of the proposed features reveals the highest classification accuracy, a comparative evaluation is determined using the BRATS2018 dataset. This is performed by examining the outcome of each predictor incorporated with the machine learning algorithms undertaken in this work (Figure 4.9). Consequently, the predictor tC_R shows the highest classification accuracy when tested with most of the classification models. For instance, the classifiers LDA, SVML and ESDA have achieved the best classification accuracy at 90.87% based on the tC_R. The predictor tC_M has also achieved a competitive accuracy of 90.167% when examined with both the KNNM and KNNCUB classifiers. The predictor Nec_R has shown the next dominant accuracy of 87.36% when evaluated with LDA, ESDA classifiers. Following these results, the predictor Nec_M has shown lower accuracy at 83.15% using the same classifiers. While, all the proposed features associated with tumour edema have shown the lowest accuracies whereby the averages of their results are around 67 %. The classifiers that show the best accuracies using features from tumour edema are DT, LDA and SVML.

Remarkably, the predictor tC_R achieved the best classification accuracy for glioma grades in this experiment compared to all other predictors associated with tumour descriptors when they are integrated with the machine learning algorithm. Therefore, this predictor has been selected to be the reference feature for the initialisation of the proposed HFSA. The selected set of the five features by ANOVA were Nec_M, tC_M, Nec_R, Edm_R, tC_R. After implementing the proposed HFSA, the selected set of features were four features: tC_R, Edm_R, Nec_M, and Nec_R. The best classification performance in terms of classification accuracy, sensitivity, specificity, precision, and F-measure were achieved by the SVMCUB classifier at 93.33%, 99.05%, 77.33%, 92.44% and 95.63% This is followed by SVMQ classifier that achieved accuracy of 91.58%, 96.67%, 77.33%, 92.27% and 94.41% respectively (APPENDIX C, Table C.7).
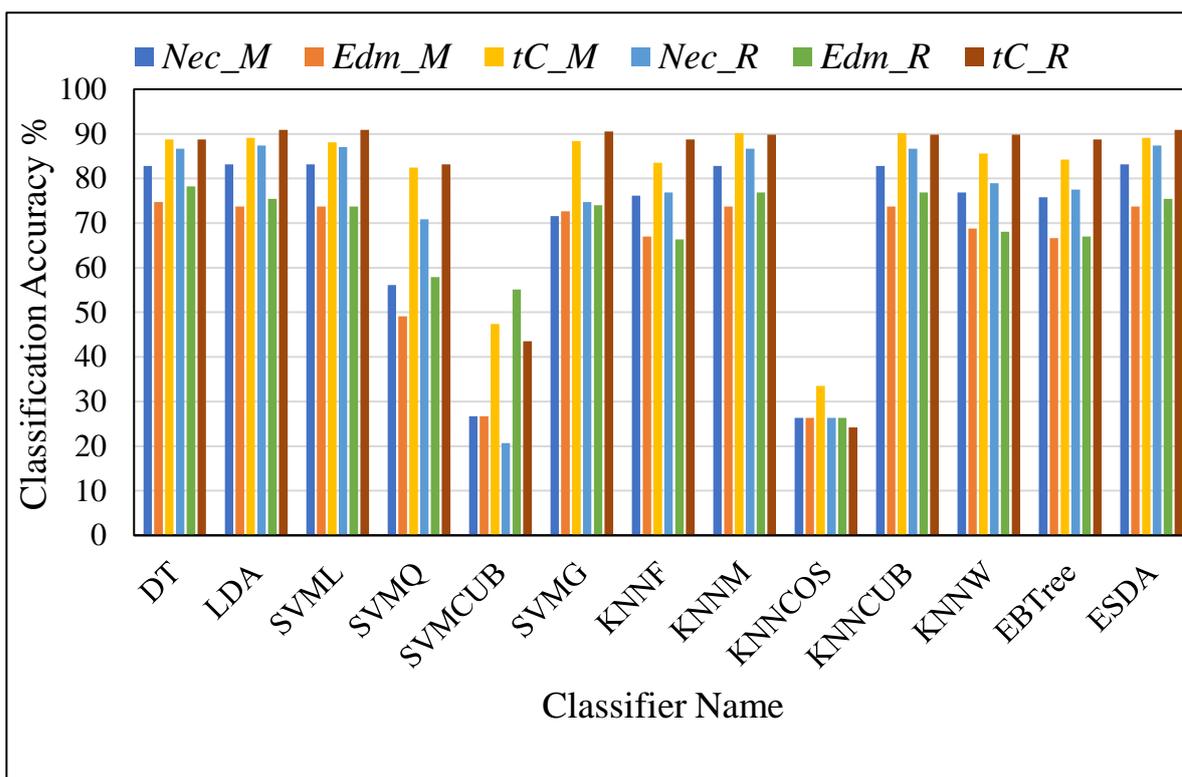
Figure 4.9 Comparative results in terms of classification accuracy for training and testing different classifiers based on the proposed predicators using the BRATS2018 dataset.

An overall comparative evaluation has been established for all the classifiers based on the three cases of different input feature subsets to the classification model (Figure 4.10). The experimental results illustrated that the classification performance in terms of classification accuracy was enhanced using the ANOVA compared to the use of the full set of features. For instance, the accuracy was improved from 89.12% to 91.22% by SVMCUB and from 89.12% to 90.17% by KNNW (Figure 4.10). It is observed that the best improvement using ANOVA was gained by the SVMQ classifier where the accuracy enhanced from 90.52% to 91.92%. However, some other classifiers showed lower accuracies when examined with the selected set of features by both ANOVA and the proposed HFSA. Nevertheless, the results obtained from demonstrating the proposed HFSA reveals that the classification accuracy improved when compared to both the use of the full set of features and the use of the selected set by the ANOVA method. Significantly, different classifiers achieved an improvement using the selected set of features by the proposed HFSA. For example, the classification accuracies were enhanced from 90.52% to 91.57 by SVMQ and from 86.66% 87.715 by KNNF classifier. However, the maximum classification accuracy that outperforms all other choices were achieved using the selected set of features by the proposed HFSA at 93.33% by SVMCUB classifier (Figure 4.10). Another advantageous is achieved by the development

of the classification system based on the incorporation between the proposed features and the use of the proposed HFSA as the results showed that this method gains a better reduction in features space as well as higher classification accuracy for glioma grades.
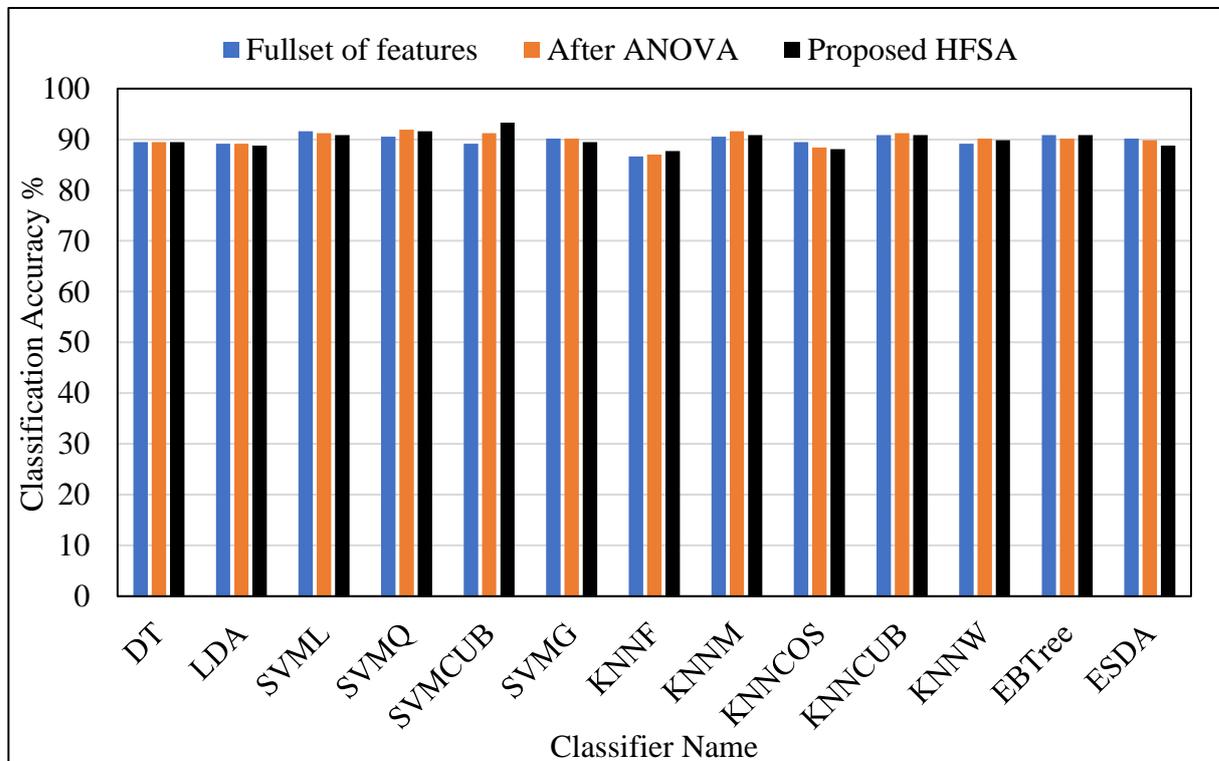
Figure 4.10 Overall comparative results to show the behaviour of the automated classification system for the proposed HFSA versus ANOVA and the full set of features using the BRATS2018 dataset.

4.8 Results Comparison and Discussion

The overall summarised results for the datasets, BRATS2013, BRATS2015 and BRATS2018 (Table 4.10), show that the automated classification system based on the proposed features achieved an improvement for the datasets when the proposed HFSA was involved in selecting the most significant features (Table 4.10). This table also shows that the final maximum classification accuracy achieved was 93.33% when examined with the BRATS2013 dataset, 90.5% when tested with BRATS2015, and 93.33% when tested with BRATS2018 dataset. In terms of sensitivity, the classification performance using the BRATS2015 dataset exhibits a slightly reduced result. However, specificity, precision and F-measure show better enhanced results thereby making better improvement in the classification accuracy at 90.5%. A broad scope of the evaluation on the classification performance is conducted using the three benchmark datasets, which incorporated different popular classification models. This is to assess the automated classification system for

glioma grades based on the proposed features and machine learning algorithm. The results indicated that the selected set from the proposed feature can achieve notable and competitive results and therefore these features are efficient predictors able to determine the grading of glioma through developing an automated design based on those features and the machine learning approach.

Table 4.10 Results Summary of the classification performance obtained by applying the automated classification system based on the proposed features associated with the tumour descriptors using BRATS2013, BRATS2015 and BRATS2018 datasets.

	BRATS2013				BRATS2015				BRATS2018			
	Accuracy %	Sensitivity %	Specificity %	Precision %	Accuracy %	Sensitivity %	Specificity %	Precision %	Accuracy %	Sensitivity %	Specificity %	Precision %
Before feature reduction	**83.33**	90.00	70.00	85.70	**89.05**	97.27	55.56	89.91	**89.12**	94.76	73.33	90.87
After feature reduction	**93.33**	95.00	90.00	95.00	**90.51**	96.82	64.81	91.80	**93.33**	99.05	77.33	92.44

Highlighting the difference in the findings of this chapter compared to the results of Chapter 3, results indicate a level of competitiveness obtained between the tumour heterogeneity investigated in Chapter 3 and the proposed features evaluated in this chapter. To illustrate, it was noted that the classification accuracy obtained from developing the classification system based on the texture features showed 87.9% when examined with the BRATS2015 dataset and 88.07% when examined with the BRATS2018 dataset (sections 3.7.4 and 3.7.5). While better accuracy of 90.5% for BRATS2015 and 93.33% for BRATS2018 were achieved when the new features were examined. However, the textural image features showed optimal results when evaluated using BRATS2013 dataset at 100%, while the classification system based on the proposed features in this chapter achieved 93.33% using the same dataset. For further elaboration, Texture feature depends on detecting the variance in the image intensities to differentiate between different dominant patterns associated with diverse grades of malignant tumours. Therefore, it has a high sensitivity to the level of homogeneity in the image's patterns, thereby showing a lower outcome when it is tested with samples from both the BRATS2015 and BRATS2018 datasets. This is because MR images from both BRATS2015 and BRATS2018 datasets have a higher level of homogeneity compared to the BRATS2013 dataset. On the other hand, the proposed features extracted from the presence of tumour descriptors are independent to the image intensities

and are therefore able to play a vital role in overcoming this limitation, indicating more robustness in detecting tumour grades of glioma. Consequently, they achieved better results compared to tumour heterogeneity when tested with the BRATS2015 and the BRATS2018 dataset. Nevertheless, the proposed features did not achieve the optimal result compared to tumour heterogeneity when tested with the BRATS2013 dataset. The reason behind this is the proposed features are dependent on the nature of the presence area of tumour descriptors, for example, the tumour enhancement, which is affected by the consequences of the leakage in the blood-brain-barrier caused by the growth of a malignant tumour and therefore the classification outcome can be degraded and affected. It can also be argued according to the experimental findings in this work that the tumour enhancement developed by samples from BRATS2013 and BRATS2018 shows a dominant pattern and thereby plays a significant role in the differentiation between low-grades and high grades of glioma. While samples from the BRATS2015 dataset behave differently as large areas of the low and high-grade tumours have not developed contrast enhancement, the non-contrast enhancement was the most dominant pattern. This can be evidence that the variability in response to the breakdown in the blood-brain-barrier leads to an absence of the contrast enhancement, which consequently leads to a reduction in the classification accuracy using the proposed features compared to the tumour heterogeneity.

The proposed predictors were analysed statistically before starting the classification process, through investigating the relevance analysis using the ANOVA technique. It was found that some of these predictors are significantly different between low and high-grade glioma. Further assessment was accomplished by incorporating machine learning algorithms to evaluate the predictive power of the proposed features in the classification of glioma grades. Many experiments were performed, and the resultant performance evaluation proved that the proposed MRI features incorporating machine learning algorithms offered a significant discrimination between low-grade glioma (I, II) and high-grade glioma (III, IV).

It is noted from the performance evaluation of the classification system using three benchmarks that the proposed features associated with the tumour descriptors datasets are significant features. However, the proposed ratio between tumour enhancement and the other tumour descriptors, namely tC_R is found to be the most significant indicator to predict the accurate grade of glioma. Therefore, the development of an automated classification system based on the proposed predictors and machine learning algorithm can achieve an efficient, objective and accurate differentiation for glioma grades.

4.9   Conclusion

In this chapter, eight MRI generic features are proposed and extracted from the presence of tumour descriptors, namely tumour necrosis, edema, non-enhancement and enhancement. These are utilised to develop an automated classification system for glioma grades. The proposed predictors are used to train and evaluate different popular machine learning algorithms, including DT, LDA and SVM with four different kernels, and KNN with five different designs and two common ensemble systems, namely EBTree, and ESDA. Accordingly, thirteen common classifiers were trained, tested, and validated using the leave-one-out-cross validation technique. The automated system is evaluated and validated based on various evaluation metrics. This evaluation includes a comprehensive range of analysis and compares the classification performance in terms of classification accuracy, sensitivity, specificity, precision, and F-measure. Three benchmark datasets are used, namely, BRATS2013, BRATS2015 and BRATS2018, to assess the predictive power of the automated classification system based on the proposed features integrated with the machine learning algorithm. This is to achieve the aim of this work of determining the accurate grade of malignancy of an unknown brain tumour. Several brain tumour descriptors were examined using a quantitative method to determine the malignant grade of unknown brain tumours. Furthermore, the development of the machine learning algorithm based on the proposed MRI predictors will offer a significant aid to assist clinicians in clinical diagnosis and may further reduce effort and unnecessary invasive procedures like biopsies through the confirmation process for the malignancy grade of a brain tumour. Ultimately, the proposed features associated with the tumour descriptor have shown a significant and robust classification outcome when evaluated with different common machine learning algorithms using three benchmark datasets. However, it is noted that there are various behaviours of classification models obtained by using different subsets of features and input data, and it is essential to also take into account eliminating redundant features so that the performance of the classification system can be enhanced. Therefore, to seek a better solution that can overcome the limitation mentioned above, the next chapter will include further development through the fusion of different machine learning algorithms based on all tumour heterogeneity and all other tumour descriptors, attempting to achieve better improvement in the classification accuracy for glioma grades.

CHAPTER 5 : An Automated Classification System for Glioma Grades Based on Multiple Classifier Schemes and Deep Neural Networks in MR images

Overview

This work proposes a novel method for developing a multiple classifier systems for glioma grades, which uses a deep neural network. This is performed based on developing two stages of multiple classifier systems. The first stage includes training and testing of eleven classifiers individually namely, DT, LDA, SVM (with four different kernels) and KNN (with five different designs). The second stage includes the establishment of the meta-trainable design based on Back-Propagation Neural Network (BPNN) and incorporating DINN. This chapter presents the details of the design of the proposed DINN. The purpose of this chapter is to assess the discrimination ability of the proposed meta-trainable multiple classifier systems (MTMCS) to classify malignant glioma grades with more enhanced classification accuracy. This will lead to more development in the automated classification system of glioma grades. This chapter also presents the implementation, results analysis and performance evaluation of the proposed system to classify glioma grades between low and high grades using four benchmark datasets.

5.1 Introduction

The development of machine learning methods that can accurately evaluate glioma grades is of great interest since it can potentially lead to more a repeatable and reliable diagnostic procedure (Zacharaki et al., 2009). The rapid development of machine learning has played an essential role in the classification and prediction of many cancer types and grades. Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Linear and Discriminant Analysis (LDA) and Decision Tree (DT) are common machine learning algorithms that are widely used for the classification of medical data.

Deep learning is basically an end to end machine learning approach where it can be applied directly to images, and there is no need to make an effort in features extraction or selection. However, the deep learning approach requires more advanced hardware and a more sophisticated design. Establishing the model also needs a large training dataset set (Dara and Tumma, 2018). The sample size is always the key issue to machine learning especially when

using the deep learning approach (Pan et al., 2015). For example, the work demonstrated by Google Incorporation to establish a deep learning model for face detector uses 10 million images downloaded from the internet, and the networks are trained on a cluster with 1,000 machines (16,000 cores), and the training takes three days (Le, 2013).

Many studies have developed classifications of brain tumours using a single classification model. For instance, a comparative study was conducted on different classification methods for glioma grading based on a single classification approach (Zhang et al., 2017). However, an approach that takes advantage of the combination of multiple machine learning algorithms would lead to an improvement in the classification accuracy (Woźniak et al., 2014). Therefore, to achieve further improvement in the classification accuracy for glioma grades, the use of a multiple classifier systems is investigated, and several machine learning algorithms are integrated into one automated grading system.

Enabling the diversity in the output decisions of the multiple classifier systems is one of the significant factors that can lead to enhance the quality of the classification performance (Thomas et al., 2018, Kuncheva, 2014). Furthermore, unlike the existing approach that relies on random subspace of samples or features that lead to sacrificing some of the significant information in the learning phase (Ashour et al., 2018a). This limitation is addressed by approaching the diversity by demonstrating different predefined design choices developed with different classification models, leading to avoid the random generation to enable the diversity in feature space (Table 3.1).

The majority voting has been applied widely to fuse multi-classification models (Xu et al., 1992, Bashir et al., 2016), used with Multiple Classifier Systems (MCS). However, the majority vote has a limited ability to sense the complex relationships of information among different classifiers. Using a learning strategy in the fusion stage of MCS is a far more powerful method. Neural Networks (NNs) is an efficient approach and is of great interest due to its ability to automatically uncover the nonlinear relationships of different data distributions. Deep learning can be designed based on Deep Neural Networks (DNN) with multiple hidden layers of nonlinear information processing that can learn complex data patterns (Mamoshina et al., 2016). Deep neural network (DNN) is recommended in the literature as a more efficient approach than convolution neural networks (convnet) (Mohsen et al., 2018) where convnet is considered the most common and successful method dedicated to deep learning (Litjens et al., 2017). DNN has been proven to be an effective alternative method to traditional deep learning approach, achieved promising results, for example, it

was used successfully in the classification of MRI brain tumour types into four classes, these are, normal, sarcoma, glioblastoma and metastatic bronchogenic carcinoma tumours (Mohsen et al., 2018). Consequently, to obtain a better representation of the relationships among many classifiers, DNN-based MCS was developed to stack multiple machine learning algorithms in the second stage of the MCS. It is necessary with the application of back-propagation neural networks to find the optimal convergence to the global minimum, which maximises its accuracy. Indeed, at present, no such method gives a general solution to this issue. Therefore, in this chapter, a deep systematic iteration NN (DINN) was developed in attempting to tackle this issue, and to further improve the classification accuracy of DNN-based MCS for glioma grading. A novel strategy for developing multiple classifier systems was proposed to improve the classification accuracy for glioma grades. This was based on the development of two stages in the classification system. The first stage was determined using eleven classifiers, namely, DT, LDA, SVM (with four different kernels), and KNN (with five different designs). These classifiers were trained individually based on different features; including the proposed features extracted from the brain tumour descriptors, which are associated with tumour necrosis, edema, non-enhancement, enhancement, and co-occurrence textural features extracted from T2-weighted MRI modality. Then, in the second stage, an efficient method of combining all these classifiers was designed, where the fusion stage was developed based on deep neural networks incorporating an extensive iteration of DNN. The difference in classification performance between the proposed MTMCS and the single classification models was analysed in terms of classification accuracy, sensitivity, specificity, precision and F-measure. The performance was also evaluated as compared to other common and current MCS including the majority voting, EBTree and ESDA. The classification performance of the proposed design is further evaluated using four benchmark datasets. For the Cancer dataset, two experiments were performed, the first one tests the discrimination between glioma grade IV and the lower grades (II, and III), and the second experiment covers the differentiation between the low-grade II against high-grade glioma (III, and IV). For the BRATS2013, the BRATS2015 and the BRATS2018, the experiments include the classification between low-grade glioma (I, II), against high-grade glioma (III, IV). The classification system is evaluated using the LOO cross-validation technique in all stages, to add more generalisation to the results of the classification system's reliability in unseen cases. Performance evaluation of the classification is measured using many evaluation metrics derived from the confusion matrix; these are the classification accuracy, the sensitivity of high grades, the sensitivity of low

grades (specificity), precision of high and low grades and F-measure of high and low grades. The details of the development of the two stages of the proposed MTMCS are explored in the following sections.

5.2 First Stage: Single Classifier System

The first stage of the proposed MTMCS is providing the MR image-textural features for training a different classification model individually. Eleven popular classifiers were utilised, these classifiers are DT, LDA, SVM that was developed with four kernels namely linear, quadratic, cubic and Gaussian, and KNN that was implemented with five different designs namely fine, medium, cubic, cosine, and weighted. The rationale behind applying different kernels and designs with SVM and KNN is to increase the diversity of the output decisions produced at this stage. The final step of this stage is to build the output decision matrix (ODM) where the output decision vectors produced from testing each sample at this stage will include a binary stream of ones (positives) and zeros (negatives), to construct the final form of the output decision matrix (Eq. 5.1). Leave-one-out cross-validation technique is employed in all stages to avoid the problem of overfitting and to add more generalisation to the classification system. All samples (N samples) are divided into training (N-1) and testing samples. Training samples are used to construct the model, in the learning phase, and then this model is used to predict the class label of an unknown test sample. The evaluation of the output results is conducted using a confusion matrix, and their metrics such as classification accuracy, sensitivity, precision, and F-measure for both classes, are utilised to assess the classification performance for every single classifier (Deepa and Devi, 2011). The general overview structure of the proposed classification system based on MCS that includes two stages of learning is depicted in (Figure 5.1).

The final output of the first stage is the output decisions matrix (ODM) which is constructed based on the decisions produced from the single classifiers trained and tested at the first stage. The mathematical representation of the output decision matrix produced from the output response for each classifier individually is defined by Eq. 5.1.

$$M_{ij=} \begin{cases} True & if\ sample\ i\ is\ classified\ by\ classifier\ j\ correctly \\ False & otherwise \end{cases} \qquad 5.1$$

Where *True* and *False* in Eq. 5.1 holds either 0 or 1 according to the index label of the test sample and the conditions of the equation at the right side, *i* indicates the index of the tested sample, where *i* < the total number of samples, j refers to the index of the classifier.
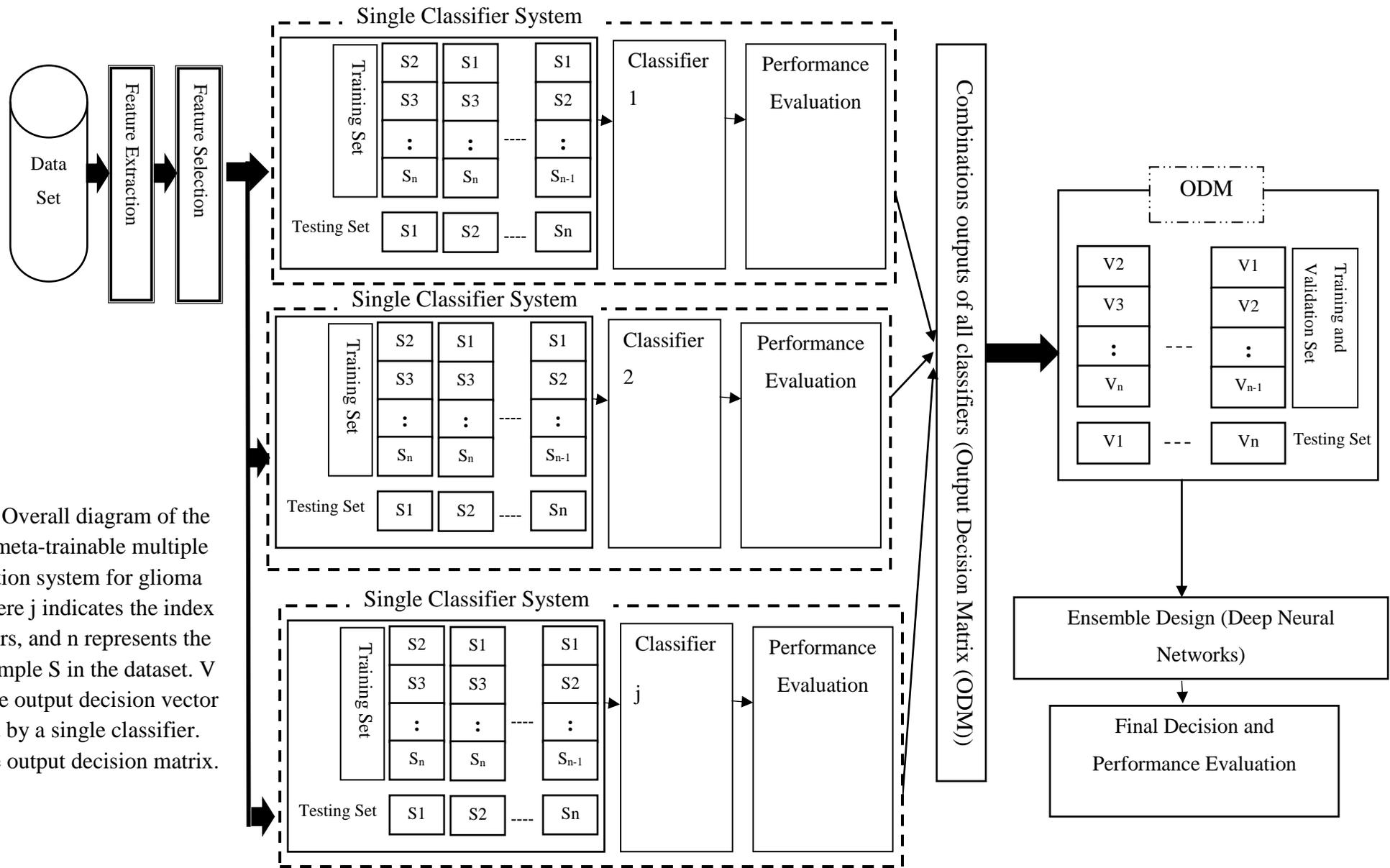
Figure 5.1 Overall diagram of the proposed meta-trainable multiple classification system for glioma grades, where j indicates the index of classifiers, and n represents the index of sample S in the dataset. V indicates the output decision vector generated by a single classifier. ODM is the output decision matrix.

143

5.3   Second Stage: Multiple Classifier System

The proposed automated system based on the MTMCS includes several stages and steps starting with features extraction, followed by the feature selection, the single classifier stage, and then the multiple classifier stage, and finally it ends with the evaluation of the classification performance (Figure 5.1). The second stage in the classification system is the ensemble of the classifiers trained previously at the first stage. The result of the first stage is ODM (Figure 5.1), which is established based on the vectors associated with the decisions of each classifier undertaken in this test. For example, given a matrix, let the classifiers be located column wise and the samples row-wise. The intersect between rows and columns represent the output decisions, and thus each constructed row will represent an output vector associated with a sample (Figure 5.1). After that, a binary classification to the output decisions matrix is conducted. This is performed by supplying the output decision vector produced from each single classification model to the DNN, where all samples in this vector are passed through the training and testing phase using the LOO cross-validation technique. The input feature vector to this stage includes streams of ones and zeros, where they represent the True and False decisions, reflecting the test of each sample in the dataset. The proposed design of incorporating DINN in the second stage is explained thoroughly in the next section.

5.4   Proposed Ensemble Design

The proposed fusion design is built based on the integration of multiple classification models using the back-propagation neural network incorporating a deep iteration NNs. All classifiers are integrated into one MCS system. The decisions vector produced from the first stage is used to learn DNN (Figure 5.2). The hyperbolic tangent sigmoid transfer function (tansig) is commonly used due to its full range output between the two classes (Graupe, 2013), it is therefore selected to be the activation function for all neurons in the NNs design.

A back-propagation strategy is used to optimise the NNs performance. Scaled conjugate gradient back-propagation (trainscg) is widely applied to functions with NNs and produces a fast response and efficient results (Baptista et al., 2013, Ashour et al., 2018b). Therefore, this technique is chosen to be the learning function of the design. DINN is used to evaluate the most significant design that reveals the maximum classification accuracy, leading to optimise the overall accuracy of the proposed system. In this design, a wide range of different numbers of neurons, iterations and hidden layers are examined. Consequently, extensive back-

propagation neural networks (BNN) were examined and tested in each iteration based on testing a broad range of different values of weights and biases of NNs.
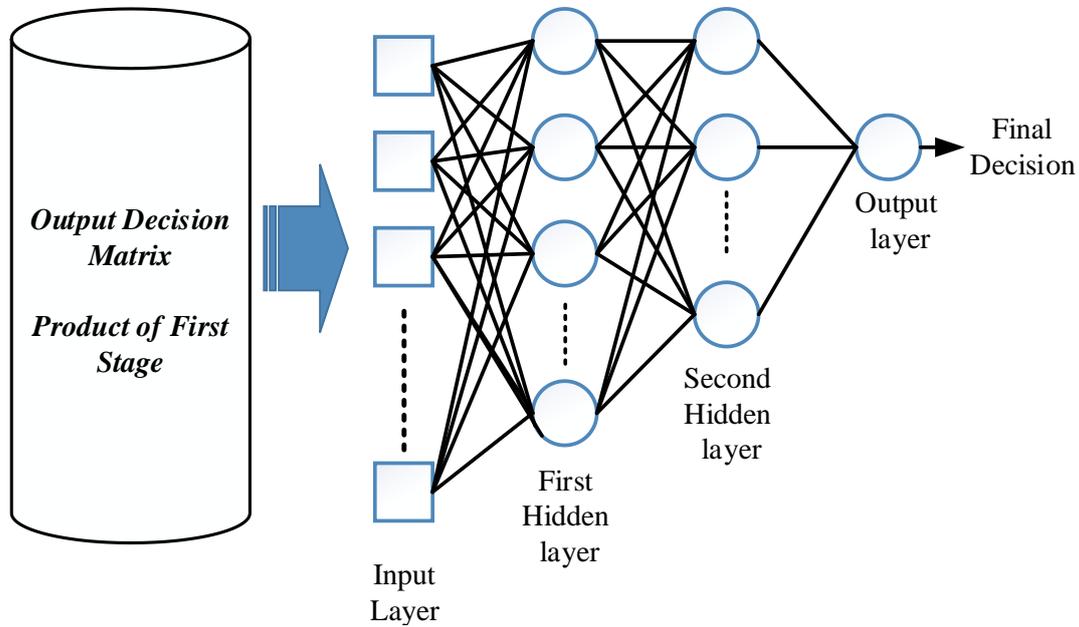
Figure 5.2 The general overview diagram of Deep Neural Networks with two hidden layers.

Let I_i represent iterations of NNs, where $i = 1, 2, \ldots, E$; and E is a selected endpoint of iterations, R_j represents rounds of NNs, where $j = 1, 2, \ldots, N$; and N is the total number of samples. The round of NNs is similar to the process of the LOO method, which is applied to avoid overfitting and get a response for every sample in the dataset. In each round, the input samples are divided into training (N-1) samples where one sample is used for testing, and this process will continue to iterate all the samples one after the other without repetition or randomly selecting any of the samples in the dataset. The deep iteration neural networks (DINN) is defined by Eq. 5.2 (AlZurfi et al., 2018).

$$A(i,j) = P((i,j) \mid n, L, R_j, I_i)) \qquad 5.2$$

Where n is the number of neurons in layer L, and P is a probability which represents the output of back-propagation NNs (BNN) measured based on the parameters (n, L, R_j, I_i). Implementing all possible design of NNs based on these parameters through the training, validation and testing phases of NNs led to the construction of the DINN that includes all the possible results of NNs using these parameters.

To further clarify the implementation of the proposed ensemble design, which incorporates the DINN, an example is presented as follows, given $N = 30$ samples in the experiment, the parameters of Eq. 5.2 become as follows; thirty rounds R_j based on the LOO technique is conducted to cover all N samples given to the classification system. In each round, the dataset is separated into different sets of training, validation and testing sets, and then the NNs design is trained, validated, and tested. In the implementation of the DINN, The BPNN uses different validation sets that are selected randomly from the original dataset and not included in neither the training phase nor the testing phase. Each generated design of NNs should be self-optimised using the learning function. In this example, twenty-five patients are selected to be in the training phase, four samples in the validation set and one sample in the testing phase. In general, one sample through the leave-one-out procedure is selected for the testing phase; then the remaining data is divided; 0.85% used for the training phase and 0.15% for the validation phase. The order of the samples is re-arranged in each iteration of NNs to examine the behaviour of each unique design of BNN. When implementing the DINN, it is necessary to ensure that the cross-validation is fully controlled by applying a completely different dataset in each of the training, validation, and testing phases. In this example, n is in the range of (1, 2, … 30). After the calculation of DINN is completed, the confusion matrix is measured by comparing the output results of each possible design generated by the DINN with the true class label. Considering a general threshold for the output probability of the NNs of 0.5 where if the output sample is greater than 0.5, it will be assigned to one class, otherwise it will be considered as belong to the other class. The results are then ranked to select the best model that shows the highest classification accuracy. The total number of experiments required to complete the implementation of DINN can be calculated by $R_j \times I_i \times n \; in \; L$. For example, if the DNN is trained based on 30 rounds (when 30 samples are used), and 20 iterations, and 30 neurons in the layer of NNs, as a result, 18,000 experiments are conducted. Similarly, if two layers of NN are used, the overall number of experiments required is as follows: 30 rounds multiplied by 20 iterations multiplied by 30 neurons in the first layer multiply by 30 neurons in the second layer, which produces 540,000 the number of experiments needed.

5.5 Algorithm for Redundancy Analysis and Selection of Classifier

In this section, a novel method is used to investigate redundancy in the first stage of the proposed MTMCS in order to remove the most redundant classifiers and keep only the significant set that shows a significant contribution in the classification accuracy. This is also advantageous in reducing the complexity of the proposed system and for further

optimisation in the classification accuracy. This is performed using several steps started with the initialisation process where the input to the algorithm will be sorted classifiers in descending order according to the classification accuracy. This algorithm is developed to select the best set of classifiers (SC) as well as maintaining the classification accuracy of the proposed system. More details about this algorithm (Figure 5.3) are illustrated below.

In the initialisation of the SC algorithm, let the single classifiers in the first stage of the proposed MTMCS are $Y = (Y_1,..., Y_n)$, n is the total number of classifiers. Each single classifier Y is tested individually through the training and testing phase. All classifiers Y are ranked from the largest to the smallest values according to *ACC* that is measured at the first stage of the proposed MTMCS. Let counter i =1. Let *Aci* represent the initial accuracy, which is set experimentally based on the result obtained from the first run of the MTMCS based on all classification models in the experiment. In the following step, an iterated process is started with removing one classifier from the input set Y where this elimination starts with the classifier associated with the lowest accuracy at Y indexed by (n-i+1) then the proposed MTMCS is implemented without this classifier, and the output classification accuracy *ACCnew* is measured. If the resultant accuracy *ACCnew* becomes less than the previous maximum state, or if the classification accuracy is reduced further after removing this classifier, meaning that this classifier is important and should be kept. The next step is to keep this classifier and insert it into the important set of classifiers. Otherwise, if there is no difference between *ACCnew* and *Aci* or the *ACCnew* goes higher, this indicates that this classifier is redundant and should be removed. The next step is to remove the classifier in ascending order based on the classifier accuracy *ACC*, and the same process is repeated in a loop until all classifiers are completely examined and the output classification accuracy of the proposed MTMCS is evaluated. Note that *ACC* is the classification accuracy evaluated for every single classifier at the first stage of the proposed system, while *ACCnew* is the overall classification accuracy measured for the proposed MTMCS. This algorithm can produce different cases where each case can have different choices of classifiers with their contribution in the final classification accuracy of the proposed system where it is possible to choose the case that has the best contribution to the classification performance. It also has the flexibility to select different cases where there is a trade-off between the classification accuracy and classifier dimensions, providing the ability to choose the case that satisfies a problem solution for an application. However, in this work, the most concentration is given the classification accuracy as the work deals with a critical medical field. Therefore, the

selection is conducted based on the case that shows the highest classification accuracy. The ultimate output of this algorithm is the best set of classifiers with the maximum classification accuracy.
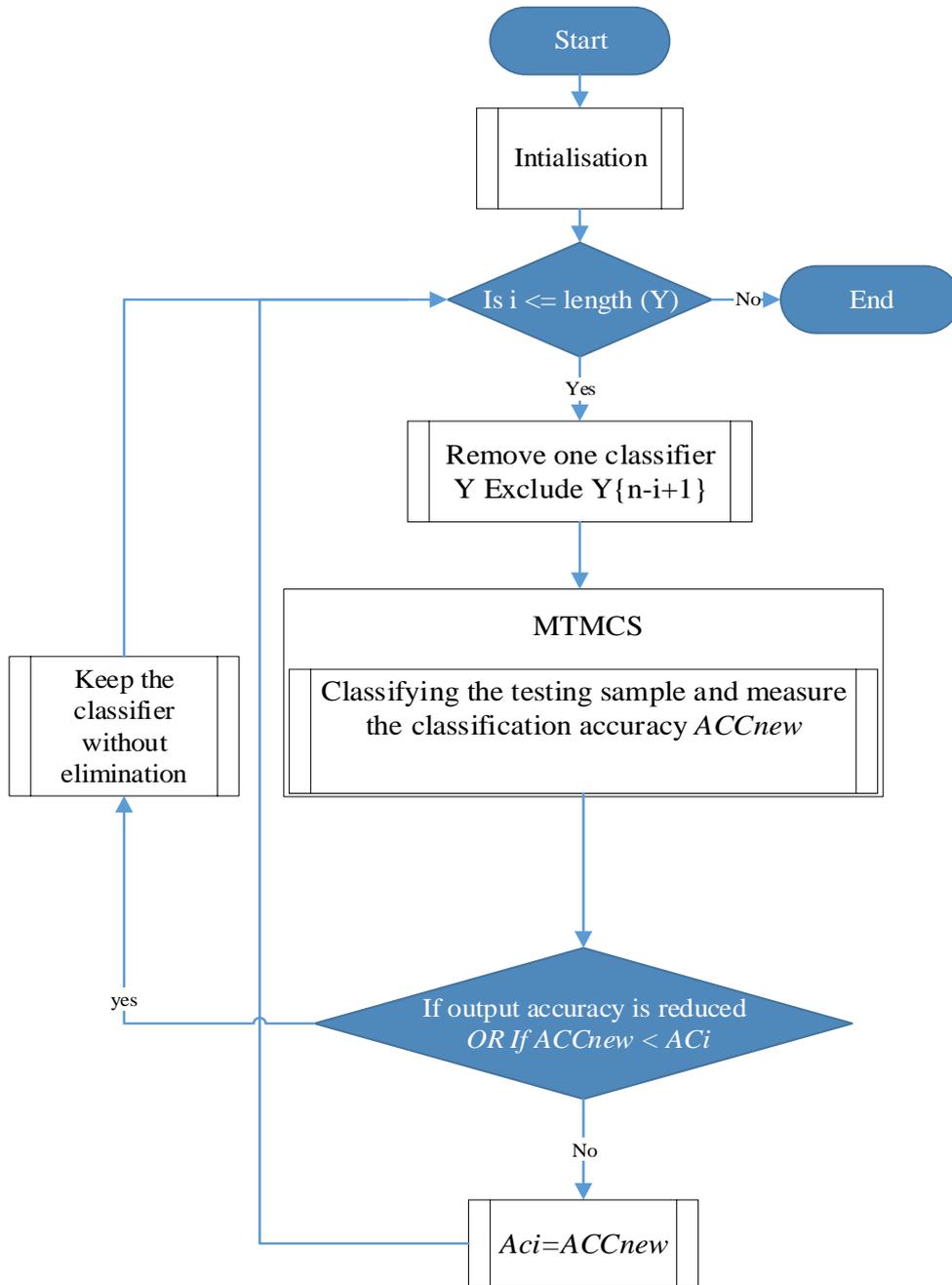
Figure 5.3 The flow chart of the SC algorithm to select the best of classifiers based on the proposed MTMCS

5.6 Evaluating the Proposed MTMCS Using the BRATS2013 Dataset

The purpose of this experiment is to evaluate the proposed MTMCS for the discrimination between low grade (I, and II) against higher grades (III, and IV) using the BRATS2013 dataset. The classification of glioma grades is crucial to preoperatively determine whether the malignant brain tumour has a high glioma grade or a lower grade. This is useful for the management of suitable treatment and prognosis for a patient who develops a malignant brain tumour. The evaluation of the proposed MTMCS starts with the implementation of one-layer NNs, where a range of neurons from 1 to 30, and 50 iterations are examined to every single neuron. The input to the first stage of MCS is the T2-weighted MRI textural features extracted by the eighteen statistics measured from the 2DGLCM. The input to the second stage is the output decision matrix (ODM) developed based on the eleven classifiers trained and tested individually at the first stage of the proposed system.

Considering the confusion matrix obtained after applying one-layer of NNs of the proposed MTMCS (Table 5.1), where all samples of low and high grades glioma are correctly classified by the proposed classification system, indicates the full discrimination rate between low and high-grade glioma. The results obtained in terms of classification accuracy, sensitivity, specificity, precision and F-measure reached to the full classification rate at 100% between low-grade glioma (I, II) and high grades glioma (III, IV).

Table 5.1 Confusion matrix for the proposed system based on one-layer NNs for the discrimination between low (I, II) and high grades (III, IV) using BRATS2013

Actual	Predicted	
	Low (GI, and GII)	High (GIII, and GIV)
Low (GI, and GII)	10	0
High (GIII, and GIV)	0	20

The results obtained by applying the proposed MTMCS reveals that optimal classification accuracy at 100% achieved using 20 neurons in the layer at the 13th iteration (Figure 5.4 and Figure 5.5). There are also many other reliable results obtained by using a different number of neurons, for example, when using 17 neurons, the classification accuracy reached 96.7%, while a lower number of neurons such as five neurons reflects the same classification accuracy at 96.7%. It is noted that many different numbers of neurons enable the proposed system to achieve same classification accuracy at 93.33%, this is due to the low errors

produced by the majority of the classifiers involved in the experiment. This indicates that using strong classifiers in the first stages of the proposed system can contribute to increasing the chance of obtaining good results reducing dependence on the tuning of the number of neurons. This can raise the chance of finding a suitable number of neurons even with the use of random selection, which could be much easier than investigating every single number of neurons. However, the proposed MTMCS can select the highest possible number of neurons that reveals the best optimal classification accuracy (Figure 5.4 and Figure 5.5).

Notably, the results obtained by using only one neuron in the proposed system is 93.33%, which is the same accuracy of the single classifier system when the highest accuracy achieved by SVML is 93.33%. However, various outcomes can be obtained by using a different number of neurons in the design of NNs (Figure 5.4).

It is found experimentally that the use of 20 neurons in one-layer of NNs with a different number of iterations revels the optimal classification performance at 100%, for examples the use of 13th and 34th iterations. However, 13th iteration requires the lowest number of iterations to achieve the optimal results, and therefore it is a better choice. The classification accuracy for the first iteration is 80%, or it can also be considered the default implementation using 20 neurons in the NNs. The second-best classification accuracy is achieved at 93.33% by the iterations 18th and 33th (Figure 5.5).
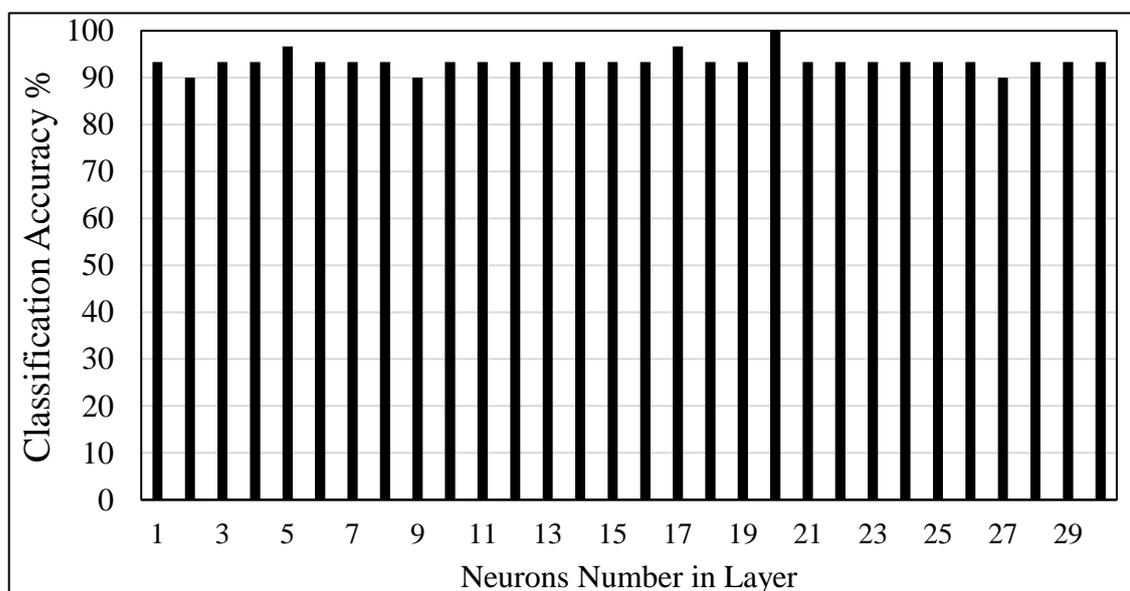
Figure 5.4 Classification accuracy results for applying the proposed system using one-layer NNs corresponding to the number of neurons in the layer to discriminate low grades (I, and II) against high glioma grades (III, and IV) using BRATS2013.
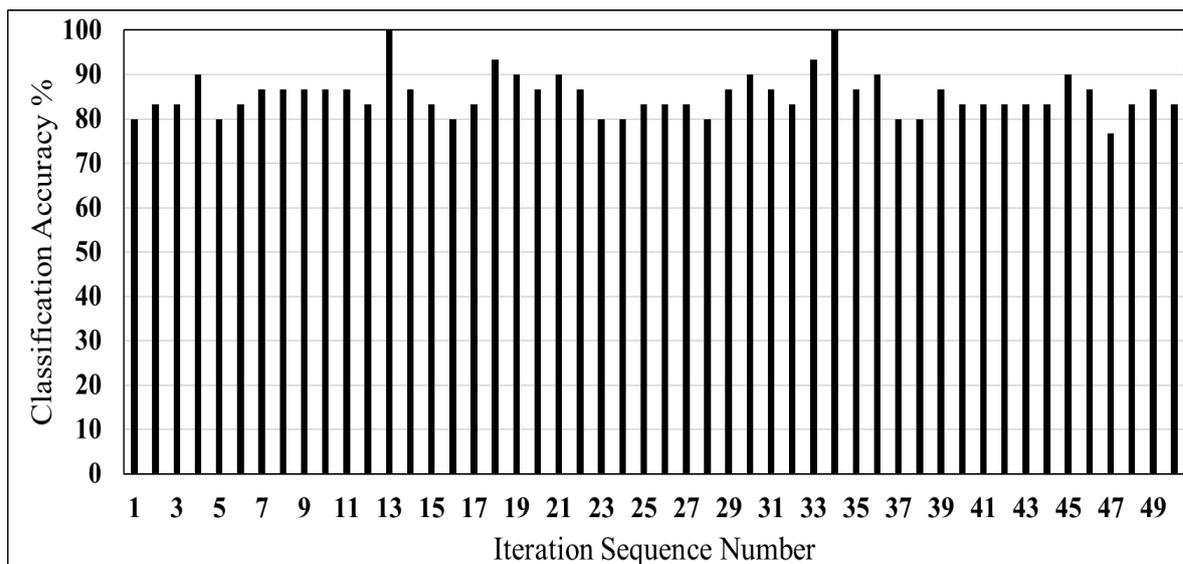
Figure 5.5 Classification accuracy results corresponding to the iteration sequence number based on the 20 neurons in the one-layer of NNs to discriminate low grades (I, and II) against high glioma grades (III, and IV) using BRATS2013.

5.6.1 Redundancy Analysis and Selection of Classifiers

The analysis of redundancy of the classifiers undertaken in this work is conducted based on the selection of the best set of classifier algorithm (SCA). In this algorithm, several runs are implemented by using a different set of classifiers, and the final classification accuracy of the proposed MTMCS is monitored. The range of these runs should cover the total number of classifiers. For example, for the eleven classifiers involved in the first stage of the proposed MTMCS, the number of required cases to implement is ranging from one to eleven. The best set of classifiers is selected through tracking the influence of every single classifier on the outcome of the proposed system. In this work every single classifier is examined twice; the first one is tested out of the stack of classifiers while the second one is examined in the stack with the other classifiers, and the best accuracy of classification is tracked (*ACCnew*) (Table 5.2). This table shows the process flow of the SCA to select the best set of the classifiers. The algorithm starts with sorting the classifiers in descending order according to the classification accuracy measured at the first stage of the proposed MTMCS. In this work, the classifiers are sorted from SVML classifier that has the accuracy of 93.33% to the KNNF classifier with an accuracy of 80% (Table 5.2). After that an initial implementation (Run Initial) is performed based on the fusion of all classifiers in the first stage of the proposed system. It is found that the initial classification accuracy of the proposed MTMCS and it is called *Aci* that is defined by the SC algorithm is 100%. The algorithm then starts an iterated process that by eliminating one classifier at a time and by

151

implementing the proposed MTMCS and by tracking the output accuracy of the proposed MTMCS the best set of classifiers will be selected. If the output classification accuracy remains at the same level or increases, it means that this classifier is redundant. Otherwise, if the new result shows a reduction in the accuracy indicating that this classifier is significant, and it needs to be maintained in the stack with the other classifiers. This process is repeated until finding the best result of accuracy with a minimum set of classifiers. Run1 shows the same accuracy at 100% after removing the KNNF classifier, and therefore the classifier KNNF is eliminated so it shows as 0 for all other Runs meaning that this classifier will not be included in any other Runs. In Run2, two classifiers are eliminated, these are KNNM, and KNNF, the results illustrate that the classification accuracy continued on the same level at 100% and therefore they are considered redundant classifiers, and both are eliminated and show as 0 for other cells (Table 5.2).

Table 5.2 Selection process conducted based on the SCA. The first column in the left represent the sorted classifiers according to their corresponding classification accuracy at the first stage of the proposed MTMCS. Table cells that include 1 and 0 refer to keep and removing actions respectively, which are determined to classifiers in different runs for the system (Run1 to Run11). *ACCnew* represents the final classification accuracy of the proposed MTMCS through the selection process using BRATS2013 dataset.

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run11
SVML	1	1	1	1	1	1	1	1	1	1	1	0
SVMG	1	1	1	1	1	1	1	1	1	1	0	0
KNNCOS	1	1	1	1	1	1	1	1	1	0	1	1
SVMQ	1	1	1	1	1	1	1	1	0	0	0	0
KNNCUB	1	1	1	1	1	1	1	0	0	0	0	0
KNNW	1	1	1	1	1	1	0	1	1	1	1	1
DT	1	1	1	1	1	0	1	1	1	1	1	1
LDA	1	1	1	1	0	0	0	0	0	0	0	0
SVMCUB	1	1	1	0	1	1	1	1	1	1	1	1
KNNM	1	1	0	0	0	0	0	0	0	0	0	0
KNNF	1	0	0	0	0	0	0	0	0	0	0	0
ACCnew	100	100	100	96.67	100	96.67	96.67	100	100	96.7	100	96.67

In Run3, when the SVMCUB, KNNF and KNNM classifier are eliminated, the results show a reduction in the classification accuracy from 100% to 96.67%. This indicates that the SVMCUB classifier is essential and it should not be removed and therefore this classifier is kept in the stack, which is presented by 1 for all other runs or cells of the table. This

procedure is repeated for all classifiers until the last run at Run11 is implemented. It is seen at Run11 that the classification accuracy is decreased and therefore the classifier SVML is considered as a significant classifier and is kept in the stack with other classifiers. Ultimately the finding of this work is that the best set of classifiers that have shown the optimal classification accuracy at 100% are namely SVML, KNNCOS, KNNW, DT and SVMCUB, which are highlighted in yellow in the table. These classifiers can be selected by different cases and can achieve the optimal classification accuracy, while the best choice with the lowest number of classifiers are shown at Run10 where only five classifiers are selected and achieved the best classification accuracy at 100%. The other advantage of this experiment is that the dimensions of the classifier are reduced from eleven classifiers to only five classifiers while maintaining the highest classification accuracy at 100%.

The classification performance in terms of the sensitivity, specificity, precision, and F-measure indicate that the proposed MTMCS achieved the optimal results at 100% and outperformed all other classification methods including the MCS based on the majority vote that showed a lower classification accuracy at 86.67% (Table 5.3).

Table 5.3 Evaluation of the classification performance for the proposed system against the majority vote for the discrimination between low grades glioma (I, and II) and the high grades (III, and IV) using BRATS2013 dataset. Class1 and Class0 refer to high (III, IV), and low grade (I, II) respectively.

| Classifier | Actual class | Confusion matrices | | Sensitivity % | Precision % | F-measure % | Accuracy % |
| | | Predicted class | | | | | |
		Class0	Class1				
Majority Vote	Class0	8	2	80.00	80.00	80.00	86.67
	Class1	2	18	90.00	90.00	90.00	
Proposed MTMCS	Class0	10	0	100.00	100.00	100.00	100.00
	Class1	0	20	100.00	100.00	100.00	

5.6.2   Comparison with Other Methods

The overall comparative results in term of classification accuracy (Figure 5.6) confirm that the proposed MTMCS based on one-layer of NNs, which achieved optimal results at 100%, outweighing all other single and ensemble classification models. The best next accuracy is obtained by the SVML classifier at 93.33% followed by 90% achieved by SVMG. While other classification models achieved lower accuracies of 86.67% namely KNNCOS, SVMQ,

KNNCUB, and KNNW, lower accuracies were achieved by DT, LDA, SVMCUB and KNNM at 83.33%. The lowest classification accuracy regarding the single classification model is obtained by KNNF classifier at 80%. While in term of the ensemble classification system, the highest classification accuracy achieved by the majority voting at 86.67% and the lowest result obtained by 76.67% classification accuracy.

This experiment based on BRATS2013 dataset does not extend further, for example, the investigation in the proposed HFSA and its impact on the proposed MTMCS. This is because there is no incitement to develop it further as the optimal classification accuracy was achieved at 100% by the single classifier system and all samples were undertaken in the experiment were correctly classified (Section 3.7.2).
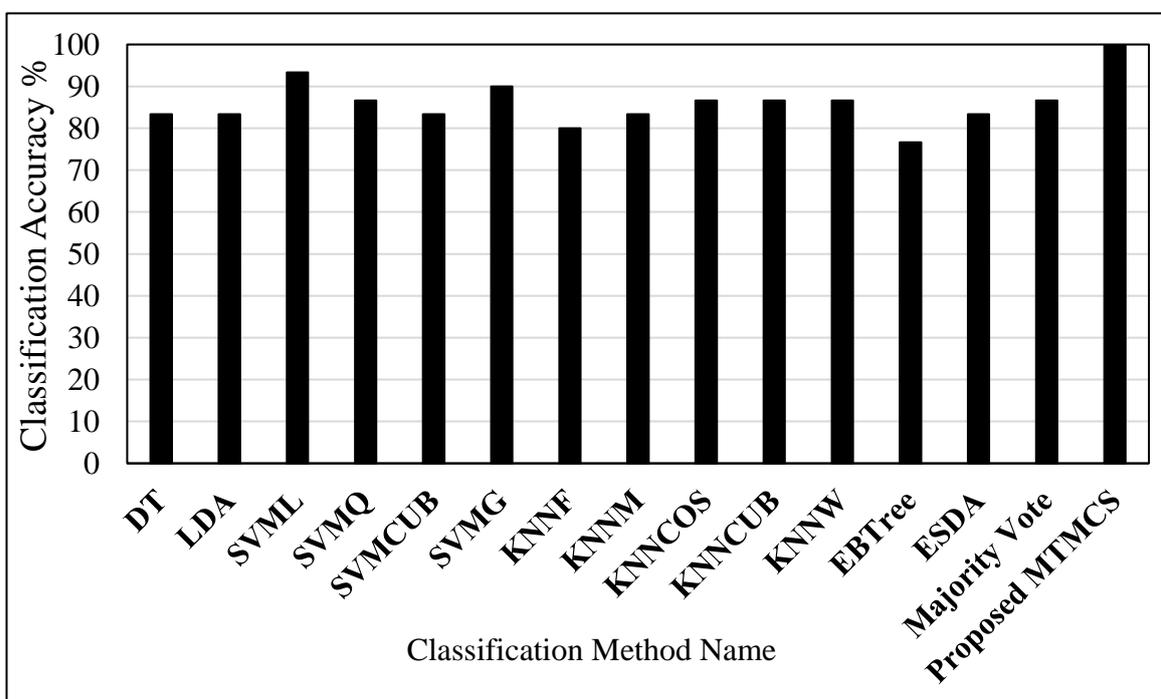
Figure 5.6 Comparative results in terms of the classification accuracy of the proposed system (one-layer NNs) versus all other single classification models and ensemble methods for the discrimination between low glioma grades (I, and II) and the high grades (III, and IV) using BRATS2013.

The total time utilised to implement the proposed MTMCS based on one-layer of NNs for BRATS2013 dataset through the process of finding the best parameters (number of neurons, and iteration sequence number) of NNs is 986.545 seconds, the average time measured for training and testing a sample is 0.155 ± 0.077 seconds. This time is measured using a personal computer with COR i7, and RAM 16 G, with enabling of parallel processes. The computation time that was required to train DNN varied and was highly dependent on the

structure of the NNs. The number of layers, the number of neurons, iterations and the time spent by the back-propagation process of NNs play a more significant role in determining the execution time required to implement the proposed MTMCS. Significantly, the execution time will be increased if a broad range of neurons and iterations are involved in the experiments. The other factor that dramatically increases the implementation time is the number of samples where more samples mean more time required for training and validating phases. Consequently, a large dataset essentially requires advanced hardware with the use of parallel computing.

5.7   Evaluating the Proposed MTMCS Using the Cancer Dataset

The proposed MTMCS is evaluated using the Cancer dataset to conduct a further assessment of the efficiency of the proposed system in the classification of glioma grades. This is achieved based on conducting two experiments; the first one is to classify the most malignant grade of brain tumours grade IV and distinguish it from the lowest glioma grades (II, and III) (Experiment 1), and the second one is to discriminate between low-grade II and high-grade glioma (III, and IV) (Experiment 2). The input image dataset to these experiments is the Cancer dataset that has thirty patients, each ten of which have different WHO glioma grades (II, III, and IV). The input features are the textural eighteen statistics extracted from the GLCM using T2- weighted MR images; these features are used to train and test the classification models.

5.7.1   Experiment 1: Discrimination between Grade IV and the Lower Glioma Grades (II and III)

The purpose of this experiment is to assess the ability of the proposed MTMCS in distinguishing grade IV from the lowest glioma grades (II, III), this is crucial because a brain tumour with grade IV is extremely malignant, and has a poor prognosis, and high mortality rate. The median survival rate for patients who develop a grade IV is usually one year (Moore and Kim, 2010). Furthermore, a brain tumour of grade IV requires early and aggressive treatment. Therefore, it is crucial to distinguish grade IV from the lower glioma grades preoperatively, as this impacts the prognosis and treatment of the patients (Theeler and Groves, 2011). Experimentally, it was noted that the 3DGLCM features enable most of the classifiers to achieve better classification accuracy as compared to the use of the 2DGLCM, and therefore the 3DGLCM is selected as the input features to the proposed MTMCS.

The evaluation process starts with implementing the proposed MTMCS based on one-layer of NNs; the results reflect various classification accuracies with respect to each number of neurons (Figure 5.7). The highest classification accuracy is achieved at 93.33%, by using 22 neurons (Figure 5.7). The same classification accuracy is achieved using 24 neurons (Figure 5.7). The lower number of neurons such as 5, 7, 10 and 11 reveals lower classification accuracy at 90%.

Fifty iterations are developed for every single number of neurons where the classification performance of the proposed MTMCS is evaluated in term of the classification accuracy (Figure 5.8) where the best classification accuracy of 93.33% is achieved using 22 neurons at the 3th iteration (Figure 5.7 and Figure 5.8). It is also noted that various numbers of iterations reflect different classification accuracies while the next best accuracy in this experiment is achieved at 90% by the 9th iteration while all other iterations reveal lower classification accuracies.
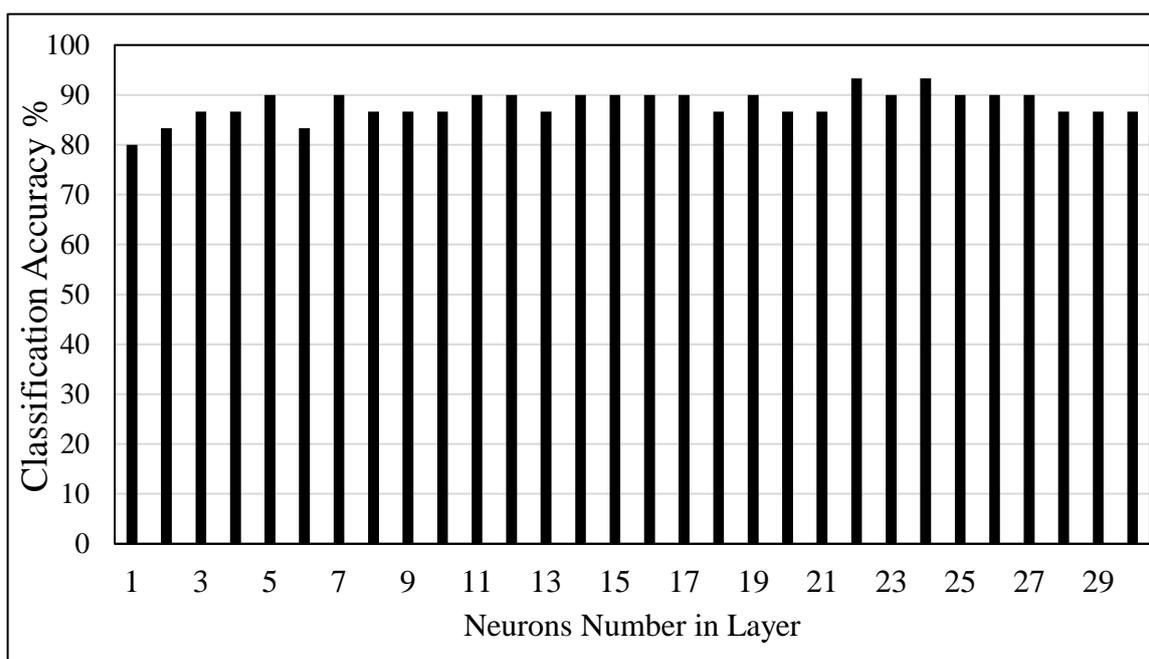
Figure 5.7 Classification accuracy results of the proposed MTMCS based on one-layer NN corresponding to the number of neurons per layer using the Cancer dataset. This test is to classify grade IV against lower grade (II, III).
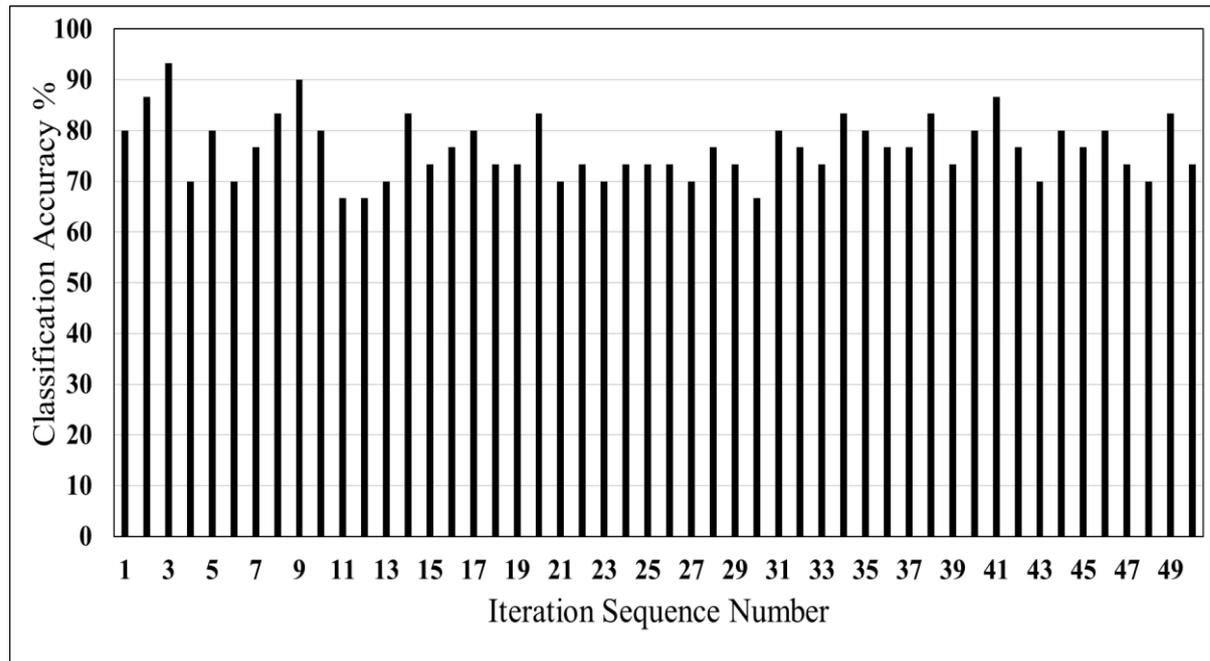
Figure 5.8 Classification accuracy results of the proposed MTMCS based on the one-layer NN corresponding to different iterations based on the 22 nodes per layer using the Cancer dataset. This test to classify grade IV against lower grade (II, III).

The evaluation of the proposed MTMCS is further extended to implement the proposed system using two-layer NNs. The result shows that the classification accuracy has improved from 93.33% that achieved by the one-layer NNs to 96.67% obtained by the two-layer NNs. The optimal design selected for the proposed MTMCS for this experiment was achieved at the 14th iteration, and with the use of 21 neurons in the first layer, and 19 neurons in the second layer NNs (Figure 5.9). Most of other examined designs of NNs in this experiment reflected accuracies of 93.33% or 90% (Figure 5.9).
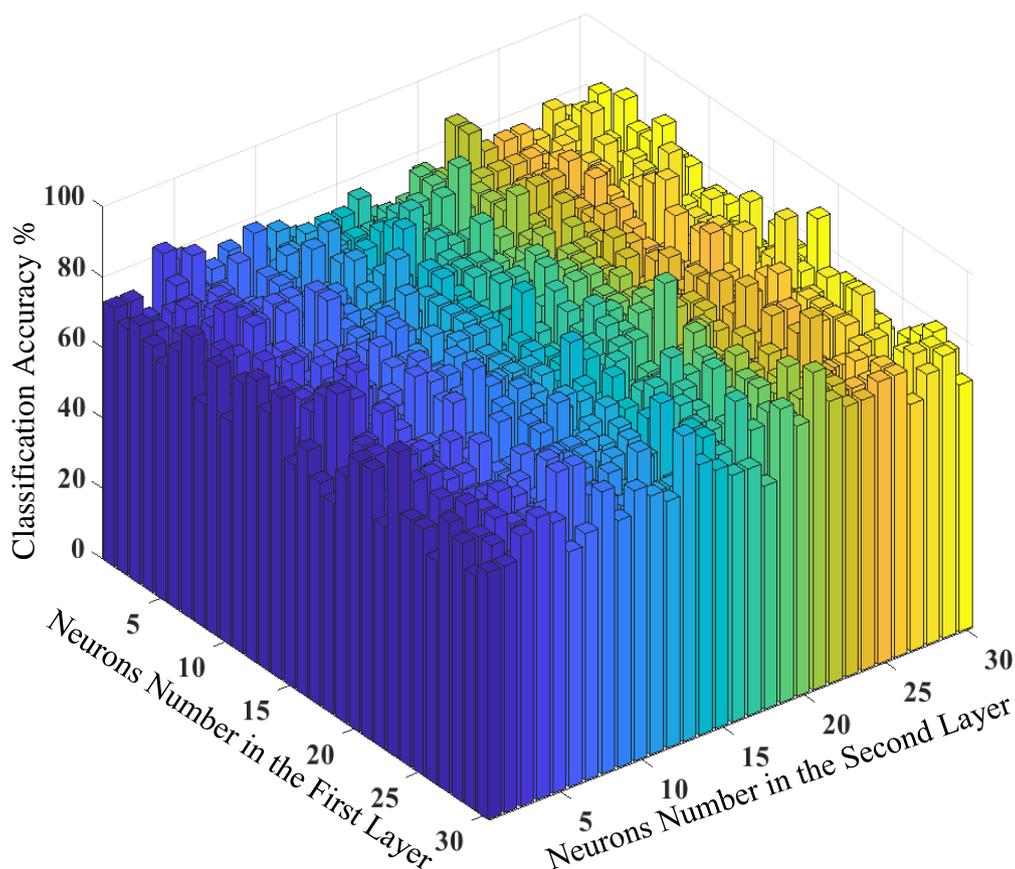
Figure 5.9 Classification accuracy results for the proposed MTMCS using two-layer NNs corresponding to number of nodes in the first and second layers in the left and right axis, respectively. This test is to classify grade IV against lower grade (II, III) using the Cancer dataset. The optimal selected design was found at the 14th iteration and 21 neurons in the first layer, and 19 at the second layer.

A comparative evaluation for the proposed MTMCS based on the two-layer of NNs versus all other single and ensemble classification systems is conducted (Table 5.4). The results illustrate that the best classification performance achieved by the proposed system in terms of classification accuracy, sensitivity, specificity, precision and F- measures compared to all the other methods is 96.67%, 100%, 95% 90.9%, 95.23% respectively. For the single classifier system, the highest accuracy achieved by both LDA, and KNNF at 83.33%. Concerning the ensemble approaches, the highest accuracy is achieved by the majority vote at 80%. The results obtained from testing all other classifiers show lower classification accuracy ranging from 80% obtained by both DT and SVML to 60% achieved by ESDA classifier.

Table 5.4 A comparative evaluation results of the proposed MTMCS (two-layer NNs) versus all other existing approaches - single and ensemble systems. This is for the differentiation between grade IV versus the lower-grade glioma, where Class1 refers to Grade IV, and Class0 indicates the lower grades (II, III), using the Cancer dataset.

Classifier	Actual class	Confusion		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	17	3	85.00	85.00	85.00	80.00
	Class1	3	7	70.00	70.00	70.00	
LDA	Class0	16	4	80.00	94.10	86.48	83.33
	Class1	1	9	90.00	69.20	78.26	
SVML	Class0	17	3	85.00	85.00	85.00	80.00
	Class1	3	7	70.00	70.00	70.00	
SVMQ	Class0	15	5	75.00	88.20	81.08	76.67
	Class1	2	8	80.00	61.50	69.56	
SVMCUB	Class0	15	5	75.00	83.30	78.94	73.33
	Class1	3	7	70.00	58.30	63.63	
SVMG	Class0	17	3	85.00	81.00	82.92	76.67
	Class1	4	6	60.00	66.70	63.15	
KNNF	Class0	17	3	85.00	89.50	87.17	83.33
	Class1	2	8	80.00	72.70	76.19	
KNNM	Class0	17	3	85.00	77.30	80.95	73.33
	Class1	5	5	50.00	62.50	55.55	
KNNCOS	Class0	16	4	80.00	80.00	80.00	73.33
	Class1	4	6	60.00	60.00	60.00	
KNNCUB	Class0	18	2	90.00	78.30	83.72	76.67
	Class1	5	5	50.00	71.40	58.82	
KNNW	Class0	17	3	85.00	81.00	82.92	76.67
	Class1	4	6	60.00	66.70	63.15	
EBTree	Class0	17	3	85.00	77.30	80.95	73.33
	Class1	5	5	50.00	62.50	55.55	
ESDA	Class0	12	8	60.00	75.00	66.66	60.00
	Class1	4	6	60.00	42.90	50.00	
Majority Vote	Class0	17	3	85.00	85.00	85.00	80.00
	Class1	3	7	70.00	70.00	70.00	
Proposed MTMCS	Class0	19	1	95.00	100.00	97.43	96.67
	Class1	0	10	100.00	90.90	95.23	

5.7.1.1 Redundancy Analysis and Selection of Classifiers

The selection algorithm for the classifiers (SCA) is applied for experiment 1 to select the significant classifiers and eliminate the others. This algorithm is applied based on one-layer NNs. The output results and the process flow of the algorithm are depicted in Table 5.5. The algorithm starts by sorting the classifiers from the largest accuracy achieved by LDA at 83.33% to the lowest accuracy at 73.33% obtained by the KNNM classifier. Then the iterated process of eliminating classifiers and monitoring the output results based on the SC algorithm is conducted.

It is found that the optimal classification accuracy is achieved of 100% at Run7, where the selected set of classifiers that achieved this optimal accuracy are namely LDA, KNNF, DT, SVML, SVMQ, SVMG, and KNNW, and they are highlighted in yellow in the first column on the left (Table 5.5). Incorporating the SC algorithm enabled the proposed MTMCS to achieve a notable reduction in dimensions of the classifier from eleven classifiers to only six classifiers as well as maintaining the optimal classification accuracy at 100% for the discrimination between grade IV and the lower grades of glioma (II, III).

Table 5.5 Selection process conducted based on the SCA. The first column in the left represents the sorted classifiers according to their corresponding classification accuracy at the first stage of the proposed MTMCS. Table cells that include 1 and 0 refer to keep and removing actions respectively, which are applied to classifiers in different runs of the system (Run1 to Run11). *ACCnew* represents the final classification accuracy of the proposed system through the selection process. This test is classifying grade IV versus grades (II, and III).

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run11
LDA	1	1	1	1	1	1	1	1	1	1	1	0
KNNF	1	1	1	1	1	1	1	1	1	1	0	1
DT	1	1	1	1	1	1	1	1	1	0	1	1
SVML	1	1	1	1	1	1	1	1	0	1	1	1
SVMQ	1	1	1	1	1	1	1	0	0	0	0	0
SVMG	1	1	1	1	1	1	0	1	1	1	1	1
KNNW	1	1	1	1	1	0	1	1	1	1	1	1
KNNCUB	1	1	1	1	0	0	0	0	0	0	0	0
SVMCUB	1	1	1	0	0	0	0	0	0	0	0	0
KNNCOS	1	1	0	0	0	0	0	0	0	0	0	0
KNNM	1	0	0	0	0	0	0	0	0	0	0	0
ACCnew	93.33	93.33	93.33	96.7	96.7	93.33	90	100	93.33	93.3	90	93.33

The results show that the optimal result at 100% of classification performance by this experiment is obtained when 22 neurons are used in the one-layer NNs (Figure 5.10). There are also different numbers of neurons with lower classification accuracy at 93.33% with the use of 16, or 25 neurons while many other numbers of neurons achieved lower accuracies at 90% such as 4, 6, 7 and 15 neurons (Figure 5.10).

It is noted that the optimal design of the proposed MTMCS based on one-layer NNs that achieve the best classification accuracy at 100% is based on the 50th iteration (Figure 5.11). Followed by 93.33% achieved at the 11th iterations. While other iterations including 27th and 46th iterations reflect the next best accuracies at 86.67%. Although the 50th iteration has more complexity than the 11th iteration, it achieves the optimal and full discrimination between grade IV and the lower grades glioma (II, III) and therefore it is selected for the optimal design of the proposed automated system for classification of glioma grades.
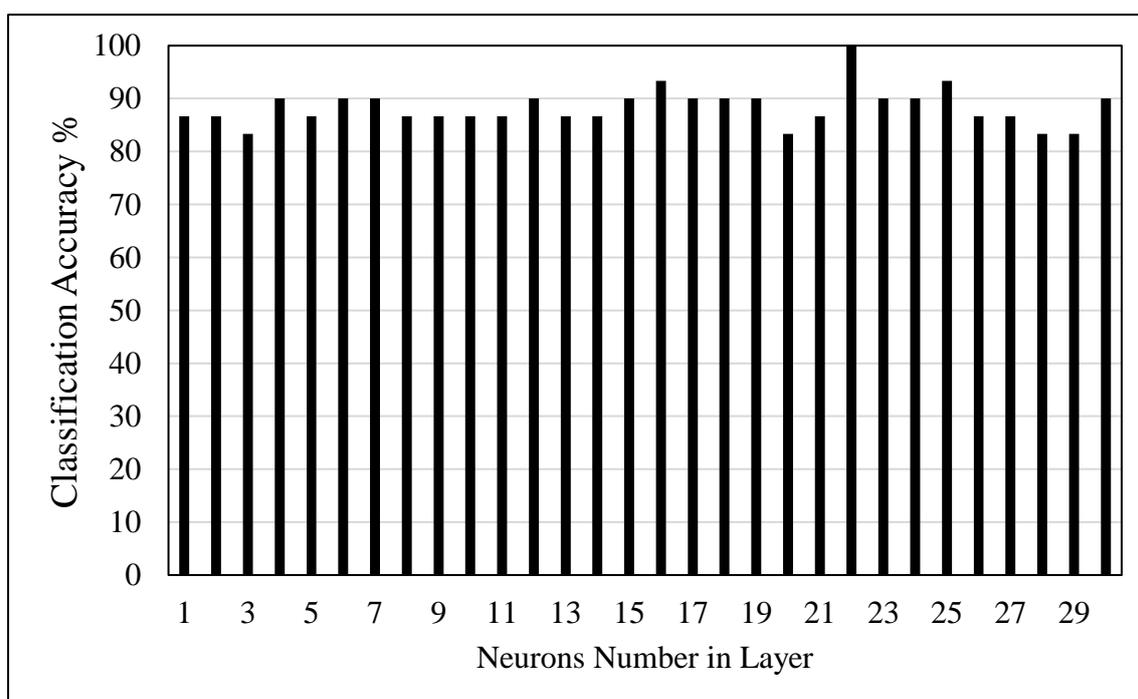
Figure 5.10 Classification accuracy results of the proposed MTMCS based on one-layer NN corresponding to the number of neurons, referring the optimal design achieved at Run7 by the SCA. This test is to classify grade IV against lower grade (II, III) using the Cancer dataset.
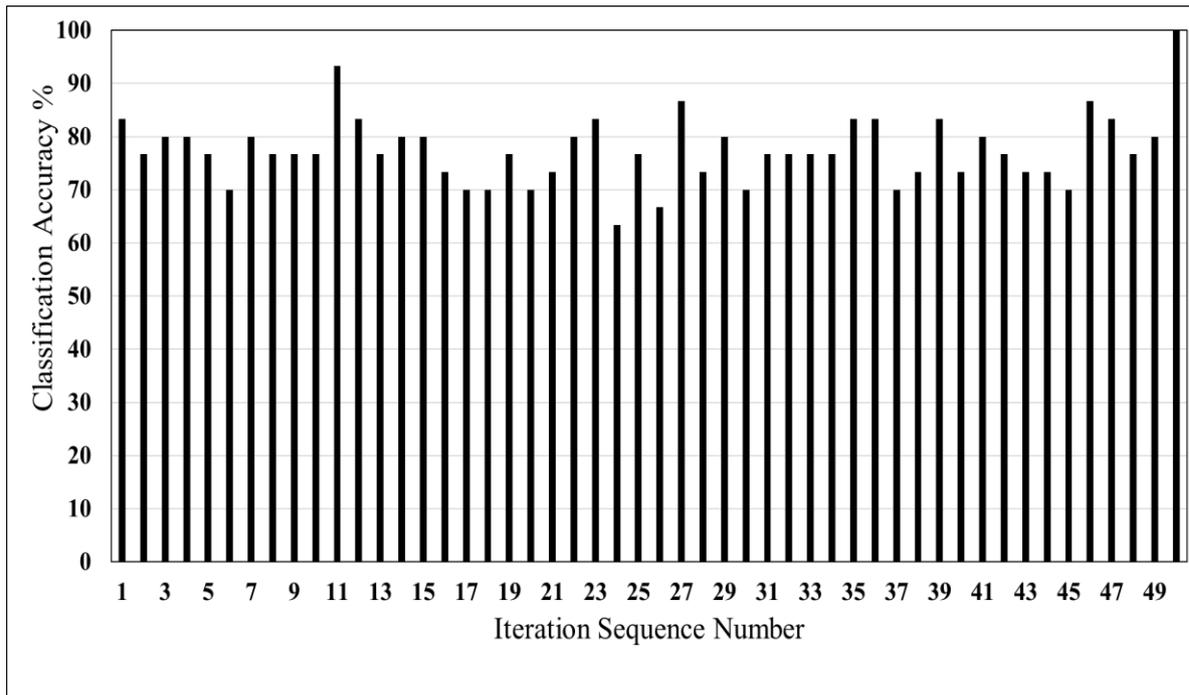
Figure 5.11 Classification accuracy results of the proposed MTMCS based on the one-layer NN corresponding to different iterations for the 22 neurons and referring the optimal design achieved at Run7 by the SCA. This test is to classify grade IV against lower grade (II, III).

5.7.1.2   Comparison with Other Methods

The overall comparative results of the classification performance of the proposed MTMCS based on one-layer NNs versus all other single and ensemble classification methods is shown (Figure 5.12). The results reveal that the proposed system outweighs all other methods and achieves the optimal classification rate at 100% for distinguishing of grade IV from the lower grades glioma (II, III). The next best classification accuracy is achieved at 83.33% by both LDA and KNNF classifiers, followed by the accuracy of 80% obtained by the DT and the majority voting classifiers. The average accuracy of the majority of other classifiers is around 75%.
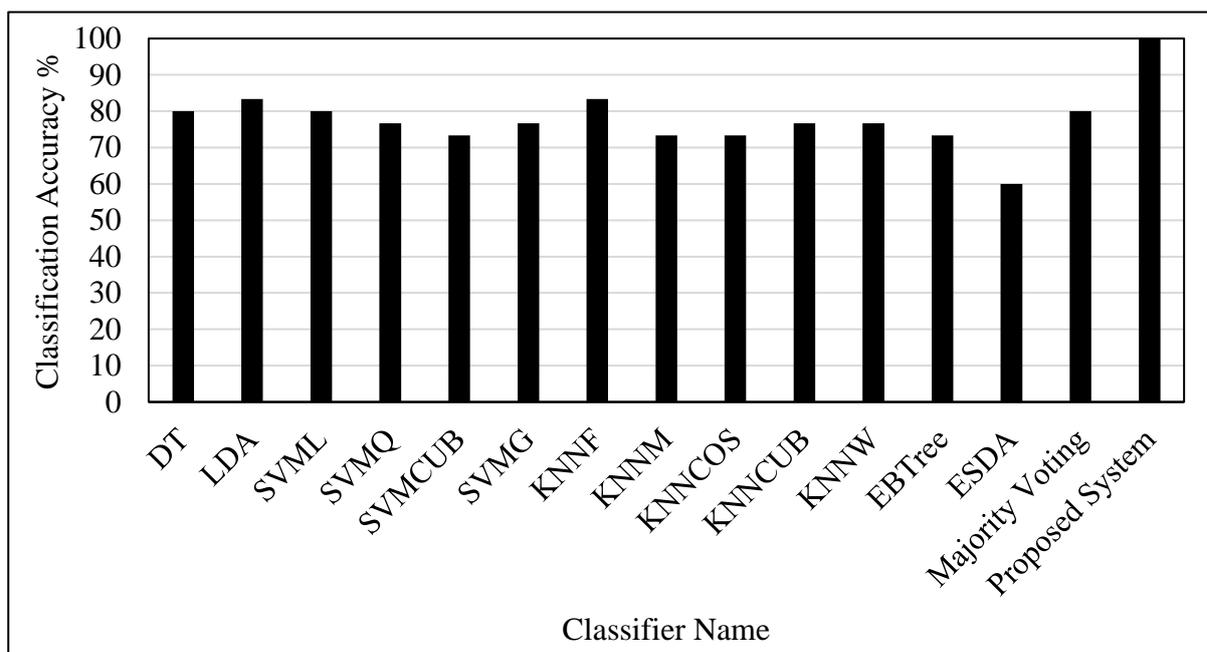
Figure 5.12 Comparative results in terms of the classification accuracy of the proposed system based on one-layer NN against all other single and ensemble classification models. This tested is to classify grade IV against lower grade glioma (II, III) using the Cancer dataset.

5.7.2 Experiment 2: Discrimination between Low Grade (II) Against the High Grades Glioma (III, and IV) Using the Cancer Dataset

The purpose of this experiment is to assess the proposed MTMCS for the discrimination between grade II against higher grades (III, and IV) using the Cancer dataset. It is advantageous to preoperatively determine whether the unknown brain tumour has a high glioma grade or a lower grade. This is essential for managing suitable treatment and prognosis for a patient with a malignant brain tumour. The classification performance of the proposed system is evaluated against different approaches - the single and ensemble classification systems. The input features are the eighteen statistical features extracted from the 2DGLCM using the T2-MR images.

The training and testing process are accomplished for both the first stage and the second stage of the MTMCS. The second stage is designed based on the one-layer NN. The classification performance is evaluated in terms of the classification accuracy by examining a different number of neurons in the layer (Figure 5.13). It is observed through this test that the best classification accuracy is achieved at 96.67% when 24 neurons are used in the layer (Figure 5.13). This optimal result is achieved at the thirteenth iteration (Figure 5.14). Considering the resultant confusion matrix of implementing the proposed MTMCS based on the design that shows maximum possible classification accuracy at 96.7% (Table 5.6), this

indicates that the propped system successfully recognises all samples of high grades while nine out of ten samples are correctly classified as low grades glioma.

Table 5.6 Confusion matrix for the proposed system using one-layer NN, associated with a design that shows the maximum accuracy for the discrimination between low (II) and high grades (III, IV).

		Predicted	
		III, and IV	II
Actual	III, and IV	20	0
	II	1	9

After implementing the proposed MTMCS based on one-layer NNs, the results indicate that a various number of neurons reflect different classification accuracy, for example, the use of 6 neurons enables the proposed system to reach an accuracy of 93.33%. While neurons 3, 4, 7, 9 and 10 show a lower accuracy of 90%. The other number of neurons reflects a lower classification accuracy at 86.67% such as 1, 6, 27, 28, 29, and 30 (Figure 5.13). The results also reveal that the maximum accuracy at 96.67% is obtained through implementing the proposed MTMCS based on the 13th iteration. While all other iterations conducted in this experiment achieved lower accuracies, for instance, the second iteration shows an accuracy of 86.67%, the first iteration shows 80% classification accuracy while a higher number of iterations such as 15, 17, and 23 reflect a lower classification accuracy at 70% (Figure 5.14).
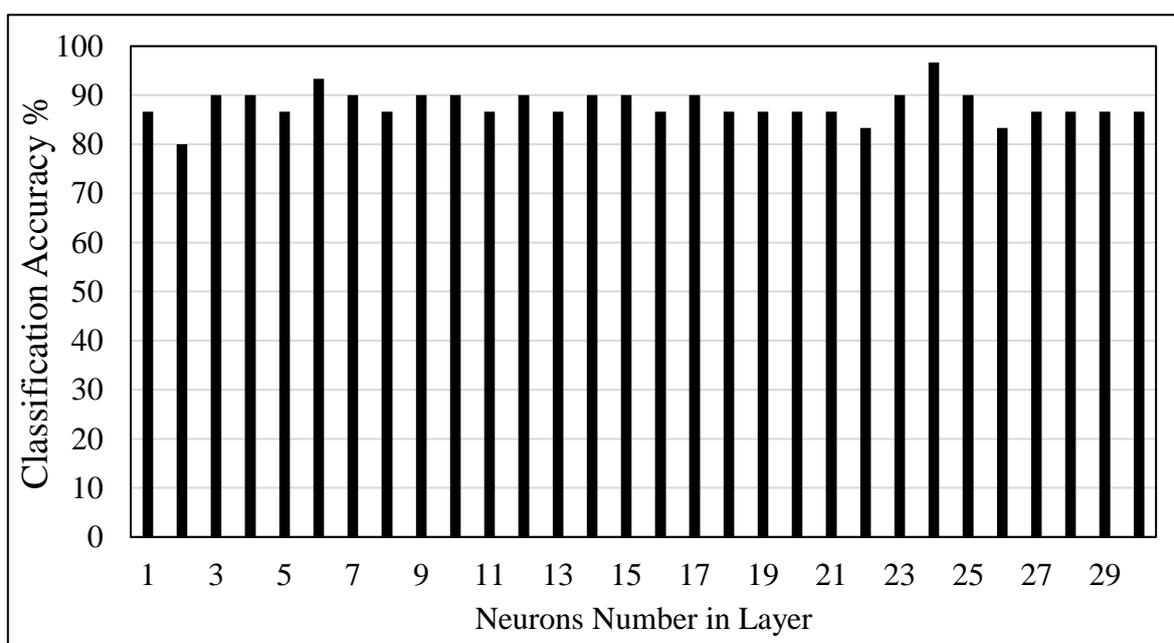
Figure 5.13 Classification accuracy results for applying the proposed system using one-layer NNs corresponding to the number of neurons in the layer to discriminate low grade(II) against high glioma grades (III, and IV) using the Cancer dataset.
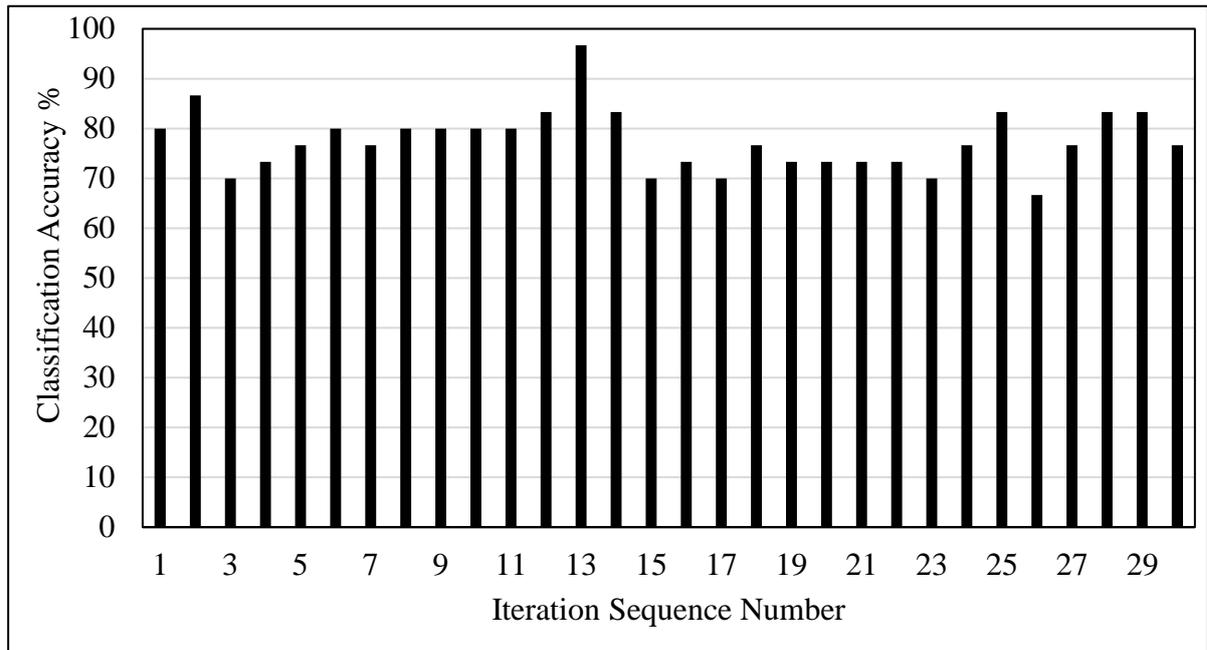
Figure 5.14 Classification accuracy results of applying the proposed system using one-layer NNs corresponding to the iteration sequence number using 24 neurons. This test is to discriminate grade II against the highest grades (III, IV) using the Cancer dataset.

5.7.2.1 Redundancy Analysis and Selection of Classifiers

The SC algorithm is applied in this experiment to investigate the redundancy of the classifiers to achieve further development of the proposed system and to select the best set of classifiers that have a significant impact on the final classification performance. It is advantageous to investigate the possibility of gaining a reduction in the dimensionality of the classifiers set. It is worthy to note that the objective function of the proposed classification system is the classification accuracy metric, therefore and based on the SC algorithm, any classifier that leads to decrease the classification accuracy or does not show any noted impact on the classification accuracy will be eliminated. After applying the SC algorithm, it is observed that the optimal results are achieved at 100% at Run5 (Table 5.7) where seven classifiers that are selected for this achievement namely DT, LDA, KNNCOS, KNNM, KNNCUB, SVML and KNNW (highlighted in the table). Several other choices for the use of a different selected set of classifiers are presented (Table 5.7).

Table 5.7 Selection process of the best set of classifiers according to SC algorithm with the corresponding classification accuracy *ACCnew* for each selected set of classifiers.   Where 1 and 0 refer to the keep and removing actions of a classifier respectively, Run represents running the process for each selected case. This test is classifying grade II versus higher grades (III, and IV) using the Cancer dataset.

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run11
DT	1	1	1	1	1	1	1	1	1	1	1	0
LDA	1	1	1	1	1	1	1	1	1	1	0	1
KNNCOS	1	1	1	1	1	1	1	1	1	0	1	1
KNNM	1	1	1	1	1	1	1	1	0	1	1	1
KNNCUB	1	1	1	1	1	1	1	0	1	1	1	1
SVML	1	1	1	1	1	1	0	1	1	1	1	1
SVMQ	1	1	1	1	1	0	0	0	0	0	0	0
SVMG	1	1	1	1	0	0	0	0	0	0	0	0
KNNF	1	1	1	0	0	0	0	0	0	0	0	0
KNNW	1	1	0	1	1	1	1	1	1	1	1	1
SVMCUB	1	0	0	0	0	0	0	0	0	0	0	0
ACCnew	96.66	96.66	93.33	96.66	96.66	100	93.33	93.3	96.66	96.7	93.33	93.33

It is noted that the proposed MTMCS based on one-layer of NNs incorporating the fusion of DT, LDA, KNNCOS, KNNM, KNNCUB, SVML, KNNW classifiers has achieved the optimal classification accuracy by Run5 at 100% in the discrimination between grade II and the higher grades (III, IV). The results show that this achievement is determined with the use of 4 neurons in the NNs design (Figure 5.15). The results also illustrate that most of the other neurons number including the choices: 2, 5, 7, 8, 9, 10, and 11 neurons exhibit high accuracies of either 90% or 93.33%. The achievement of the optimal results at full discrimination rate at 100% is determined using 40th iteration. While most of the other iterations reflect lower accuracies, 90% of classification accuracy is the next best result obtained by the 44th iteration (Figure 5.16).
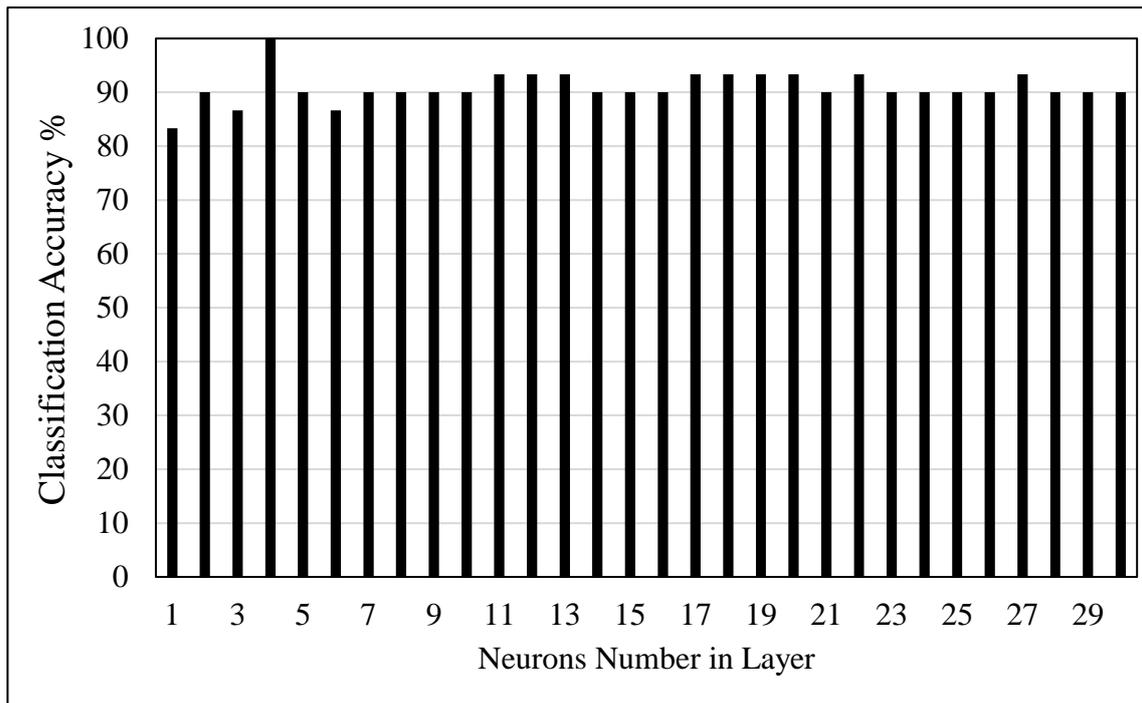
Figure 5.15 Classification accuracy results for the proposed MTMCS based on one-layer NNs according to neurons number, referring to the optimal design achieved at Run5 by SCA for the discrimination between grade II versus higher grade (III, and IV) using the Cancer dataset.
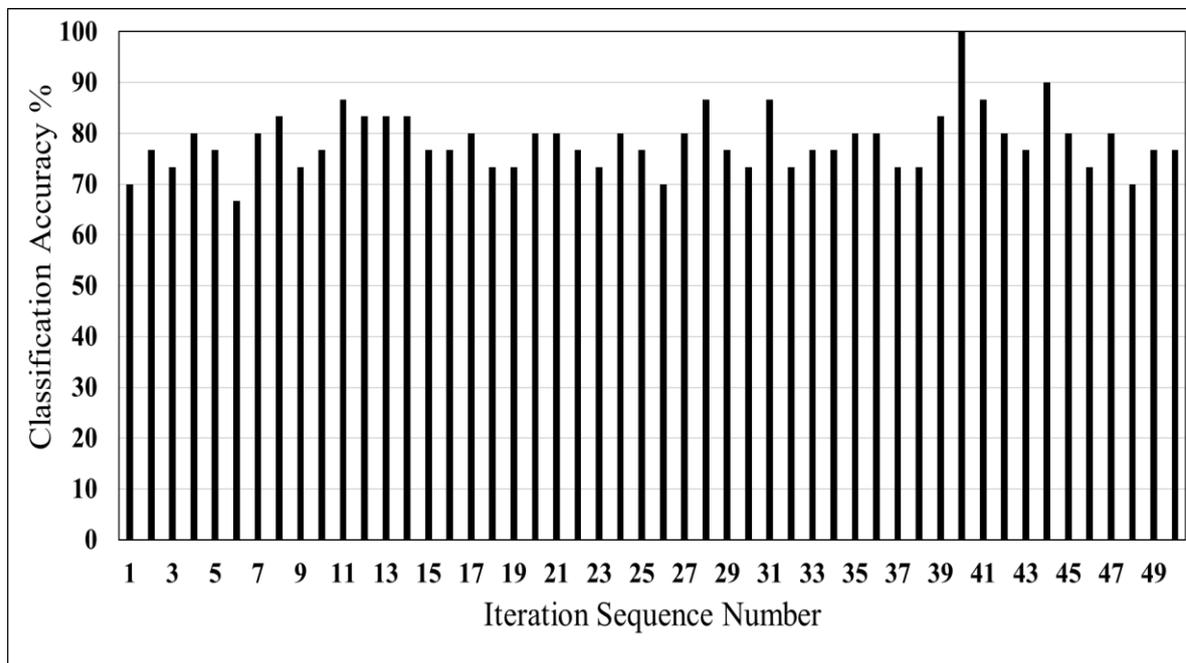
Figure 5.16 Classification accuracy results for the proposed MTMCS based on one-layer NNs with respect to number of iterations using 4 neurons, referring the optimal design achieved at Run5 by SCA for the discrimination between grade II versus higher grade (III, and IV) using the Cancer dataset.

5.7.2.2   Comparison with Other Methods

The overall performance evaluation of Experiment 2 is shown in Figure 5.17, in which a comparison of the classification accuracy between the proposed MTMCS based on the one-layer NNs and all other the single and ensemble classification models is illustrated. It is noted that the proposed system has achieved the full discrimination rate of 100% between grade II and higher grades (III, IV) outperforming all other classification methods. The next best classification accuracy is obtained by the DT classifier at 86.67%, followed by the LDA at 73.33%, with slightly lower accuracy at 70% the KNNCOS classifier that comes next. With respect to the ensemble systems, the EBTree classifier has achieved the highest accuracy at 66.67%, while the lowest accuracy at 53.33% is achieved by the ESDA classifier. The results obtained from the evaluation of different popular methods illustrates that the proposed MTMCS has the best optimal results compared to all other classification algorithms (Figure 5.17).
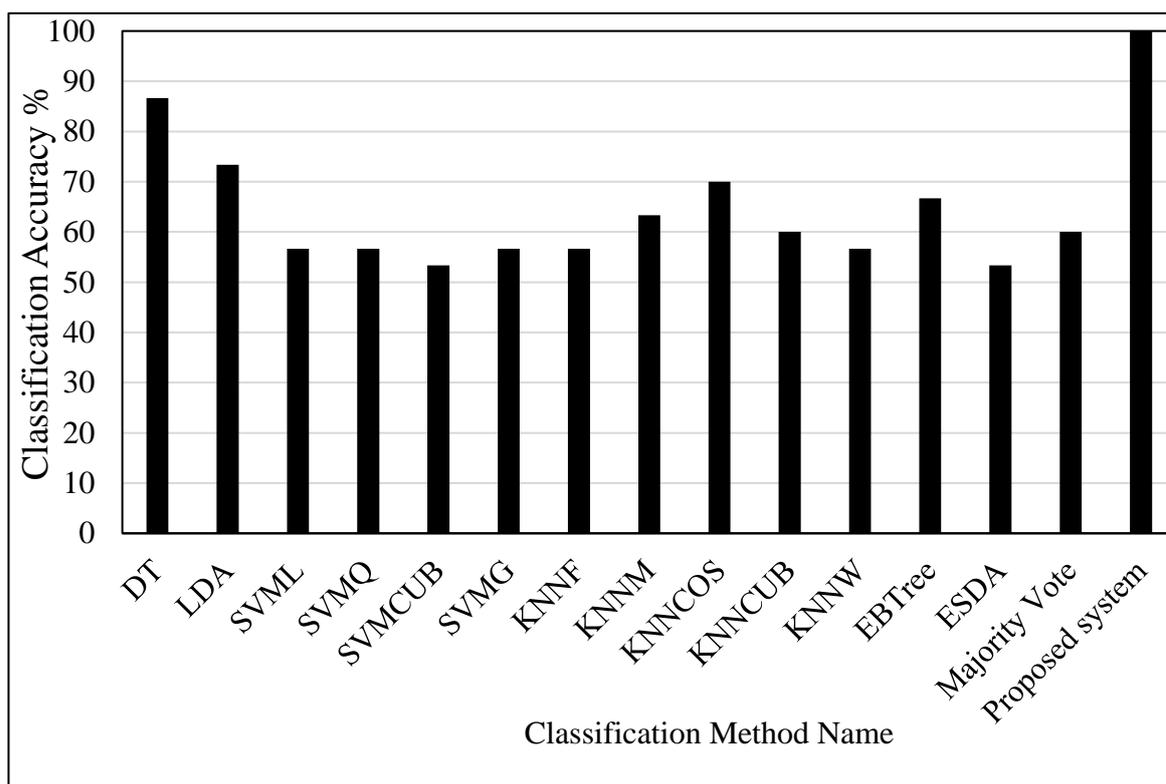
Figure 5.17 Comparative results in term of classification accuracy for implementing the proposed system versus the other classification system: single and ensemble classifiers, this test is based on one-layer NNs to discriminate glioma grade II against higher grades (III, IV) using the Cancer dataset.

5.7.3 Experiment 3: Evaluating the Proposed MTMCS Based on the Selected Set of Features in the First Stage of the System Using the Cancer Dataset

The purpose of this experiment is to evaluate the proposed MTMCS using the selected set of features chosen by the proposed HFSA. This is also to investigate the influence of the selected features on the performance of the proposed MTMCS. The outcome of this experiment is the discrimination between low glioma grades (II) and high grades (III, and IV). This experiment is performed through training and testing the classifiers based on the selected set of features that are fed to the first stage of the MTMCS. In the second stage of the proposed MTMCS, the ODM is trained and tested by applying the proposed methodology.

The proposed MTMCS is implemented based on one-layer NNs, with number of neurons ranging from 1 to 30 and 50 iterations are examined for each neuron. The dataset set used in this experiment includes thirty patients; with ten patients of grade II and twenty patients of a high grade (III, and IV). This experiment starts with allocating the target vector based on assigning the index 0 to the class label of samples with low grades (II) and index 1 to the samples with high grades (III, and IV). Then the proposed MTMCS is implemented.

The selected set of features chosen by the proposed HFSA are namely ($Autocorrelation$, 0^0), ($Homogeneity$, 90^0), and ($Homogeneity$, 0^0) which are utilised as an input to the first stage of MTMCS. These features have shown the highest classification accuracy achieved by the DT classifier at 93.3% compared to all other single and ensemble classification models. The input to the second stage of MTMCS is the output decision matrix (ODM) developed based on the eleven classifiers trained and tested based on these selected features. The results show that the performance evaluation of classification of the proposed system in terms of classification accuracy, sensitivity, specificity, precision, and F-measure are 96.67%, 100%, 90%, 95.2%, 97.56% respectively (Table 5.8). These results are obtained using one-layer NNs, and three neurons in the layer and by 10th iteration in the NNs design of the proposed system (Figure 5.18 and Figure 5.19). There is also a different design that includes different number of neurons which have shown the same highest classification accuracy at 96.6%, for examples, using 6, 7, 14, or 19 neurons in the proposed design based on one-layer NNs (Figure 5.18). However, to avoid adding more complexity to the design of the proposed system, the use of a small number of neurons is recommended as it will maintain the same classification accuracy.

Consequently, the best choice is to select the lowest number of neurons and this is achieved by using three neurons, this has shown the same high accuracy at 96.7%. In this experiment, two iterations have shown the highest classification accuracy at 96.6%, these are 10th, and 48th, where it is possible to use a design developed by any one of these iterations since both are enabling the highest classification accuracy. However, the 10th iteration is a better choice as it obviously requires a lower number of iterations and thus the use of this option is more recommended as it causes less complexity.

Table 5.8 The classification evaluation performance for the proposed system using the selected set of features chosen by the proposed HFSA. The input dataset for this experiment is the Cancer dataset. This test is to discriminate between glioma grade (II) indicated by Class0 and high grades (III, IV) assigned by Class1.

| Classifier | Actual class | Confusion matrices | | Sensitivity % | Precision % | F-measure % | Accuracy % |
| | | Predicted class | | | | | |
		Class0	Class1				
Proposed MTMCS	Class0	9	1	90.00	100.0	94.73	96.67
	Class1	0	20	100.00	95.2	97.56	
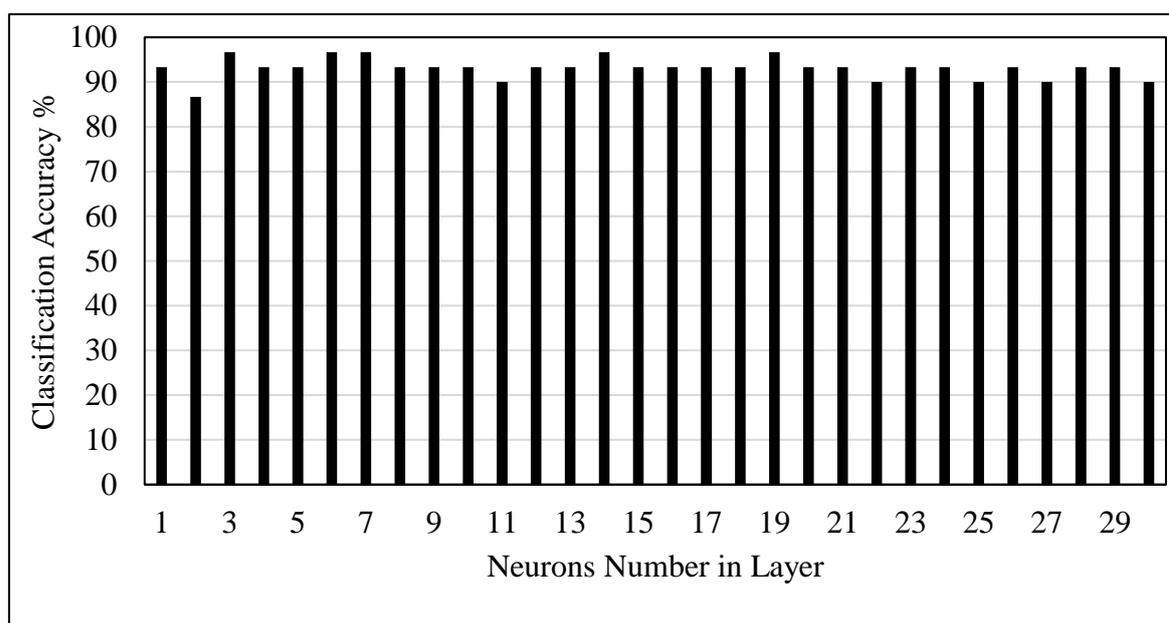
Figure 5.18 Classification accuracy results for applying the proposed system using one-layer NNs corresponding to the number of neurons in the layer. This test is to evaluate the selected set of features involved in the first stage of the proposed MTMCS for discriminating of low grades (II) against high glioma grades (III, and IV) using Cancer dataset.
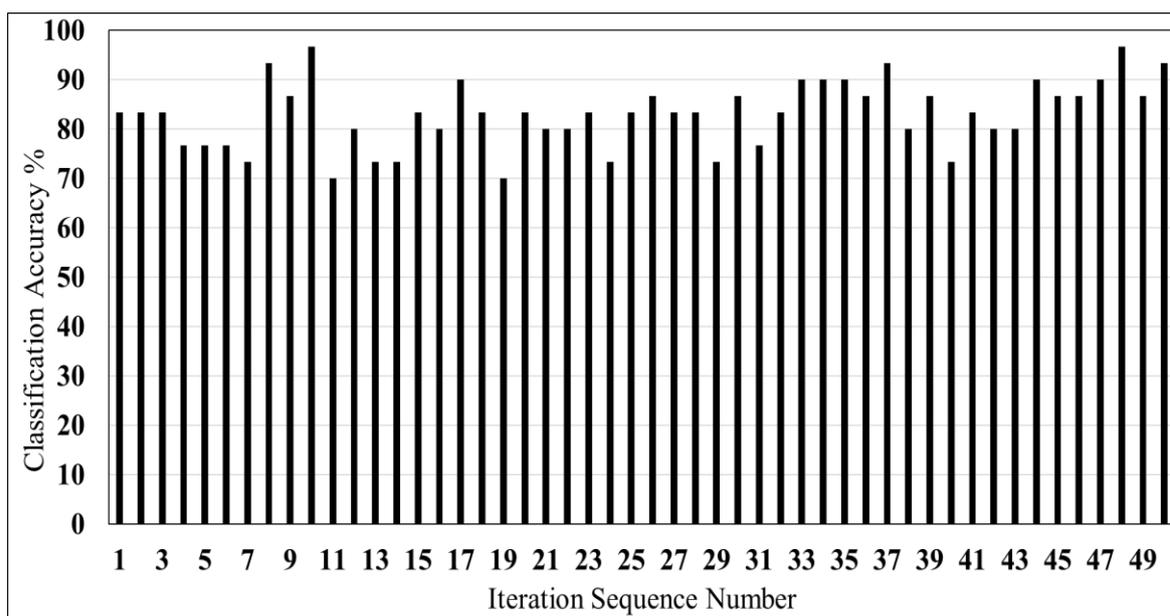
Figure 5.19 Classification accuracy results corresponding to the iteration sequence number using one-layer NNs to evaluate the selected set of features (in the first stage of the proposed MTMCS) in discriminating low grades (II) against high glioma grades (III, and IV) using the Cancer dataset.

5.7.3.1 Redundancy Analysis and Selection of Classifiers

In this experiment and after applying the SC algorithm, the results show that the best selection of classifiers is the DT alone, as it achieves a classification accuracy of 96.67%. It is observed that the elimination process through the SC algorithm of all other classifiers has not had a significant effect on the classification accuracy as it remained at the same level of accuracy at 96.67%. This indicates that the use of the proposed MTMCS based on one-layer NNs incorporates the SCA has the advantage of reducing the classifier dimensions from 11 to only one classifier while maintaining the same classification accuracy at 96.67%.

To highlight the difference between experiment 2 and 3, in experiment 2 when the full set of features is used, the optimal classification accuracy is obtained at 96.7% after which the result is improved to 100% by incorporating the SC method. While in experiment 3, when the selected set of features is used, the highest accuracy achieved by the proposed system is 96.7%, and the classification accuracy remains on the same level despite the use of the SC algorithm. The reason behind this is that experiment 3 uses the selected set of features where those features can enhance some single members involved in the first stage of the MTMCS and boost their accuracies while other classifiers do not show noted improvement in their accuracies. Consequently, the fusion of the outputs of these classifiers impacts the behaviour

of the output of the proposed MTMCS and therefore the results of experiment 3 have shown lower accuracy compared to experiment 2. However, the results of both experiments 2 and 3 show a superior classification accuracy compared to all other single and ensemble classification methods.

5.7.3.2   Comparison with Other Methods

Although the full classification accuracy has not been achieved in experiment 3, the proposed MTMCS based on those selected set of features has achieved a better classification accuracy than all other single and ensemble classification models at 96.67% (Figure 5.20). In term of single classification models, the results show that the highest classification accuracy is achieved by the DT classifier at 93.33%, and with respect to the ensemble methods, the EBTree classifier achieves the next accuracy at 90%. While most of all other single and ensemble classifiers reflect a lower classification accuracy, ranging from 63.33% to 80%.
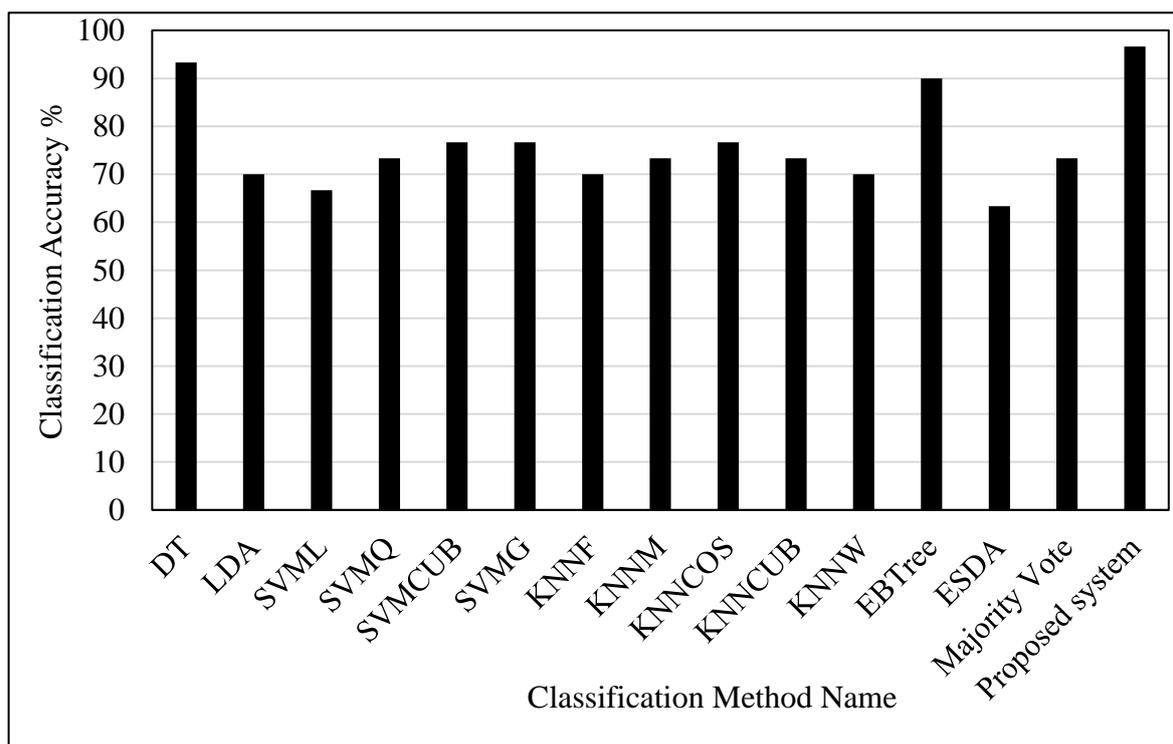
Figure 5.20 Comparative results in terms of the classification accuracy of the proposed system (one-layer NNs) against all other classifiers; single classification models and ensemble methods. This is to evaluate the proposed system based on the selected set of features in the first stage of the proposed MTMCS in discriminating between low glioma grades (II) and the high grades (III, and IV) using the Cancer dataset.

For the experiments conducted with the Cancer dataset, the total time required to implement the proposed MTMCS based on one-layer NN is 6787.887521 seconds that is practically reduced to 999.377 seconds by enabling parallel processing, while the time need for training and testing a sample is 0.150±0.036 seconds. Regarding the implementation of the proposed MTMCS based on two-layers NNs, the time required for finding the best optimal result is 253645.5425 seconds, which is decreased to 35693.57723 seconds using parallel computing, while the time required for training and testing a sample is 0.187 ± 0.022 seconds.

5.8 Evaluating the Proposed MTMCS Using the BRATS2015 Dataset

The purpose of the work is to assess further the ability of the proposed MTMCS in the classification of glioma grades using the BRATS2015 dataset; this dataset includes 274 patients, with 54 patients with low grades, and 220 patients with high grades. The input to this experiment is the textural features using the eighteen statistics measured from the 3DGLCM. The evaluation process begins with the implementation of the proposed system based one-layer NNs (Experiment 1). After that, the evaluation is conducted for the proposed system using the selected set of features chosen by the proposed HFSA (Experiment 2). Following this, the proposed system is developed based on the combination of all features including the texture features and the tumour descriptors (Experiment 3). After that, the proposed system is developed based on enabling the diversity in features space for the full set of features including the texture features and the tumour descriptors (Experiment 4). The main purpose of these experiments is to evaluate the proposed MTMCS in various conditions and scenarios to extract the behaviour of the proposed classification system through the findings and outcomes of these experiments using the BRTAS2015 dataset.

5.8.1 Experiment 1: Implementing the Proposed MTMCS Using the Full Set of Textural Features

In this experiment, the input features are the textural statistics based on the full set of features measured from 3DGLCM to classify glioma grades using the BRATS2015 dataset. This choice is made due to an empirical test that showed that the full set of 3DGLCM had shown higher results compared to using only 2DGLCM when both sets are investigated using the BRATS2015 dataset. The ODM is built based on the training and testing each classifier individually. The produced ODM will have the dimensions - eleven classifiers $\times$ 274 samples.

Considering the highest results that are achieved using the proposed MTMCS (Table 5.9), there are 36 samples out 54 of low-grades samples are correctly classified as low-grade glioma, while 212 out of 220 of high-grades samples are correctly recognised as high-grade glioma. The results show that the classification performance for the proposed system in terms of classification accuracy, sensitivity, specificity, precision, F-measure are 90.51%, 96.36%, 66.67%, 92.2%, 94.22% respectively.

Table 5.9 Evaluation results of the proposed system for the discrimination between low glioma grades (I, and II) and the high grades (III, and IV) using the BRATS2015 dataset. Where Class1 and Class0 refer to high, and low grade respectively.

| Classifier | Actual class | Confusion matrices | | Sensitivity % | Precision % | F-measure % | Accuracy % |
| | | Predicted class | | | | | |
		Class0	Class1				
Proposed MTMCS	Class0	36	18	66.67	81.82	73.46	90.51
	Class1	8	212	96.36	92.17	94.22	

After applying the proposed MTMCS based on one-layer NNs, the results indicate that the highest classification accuracy of 90.5% is achieved using 7 neurons in the layer with the 13th iteration (Figure 5.21 and Figure 5.22). There are also many other results which reflect the same classification accuracy with the use of different numbers of neurons, for example when using 18 or 23 neurons. However, with the aim of a reduction in the system complexity, the better choice would be 7 neurons and therefore the design based on 7 neurons was selected. There are other choices of neurons number including 14, 15 and 19 show a little lower classifications accuracy at 90.15%, while many other numbers of neurons illustrate various classification accuracies but mostly ranging between 89% and 89.7% (Figure 5.21).

The experimental results obtained by conducting the proposed MTMCS based on 7 neurons in the one-layer NNs indicate that 13th iteration reflects the best classification accuracy at 90.5% compared to all other iterations. Various iterations show different results of classification accuracy, while the next best accuracy at 89.42% is achieved by the 25th iteration followed by the accuracy of 89.05% achieved by the 24th iteration (Figure 5.22).
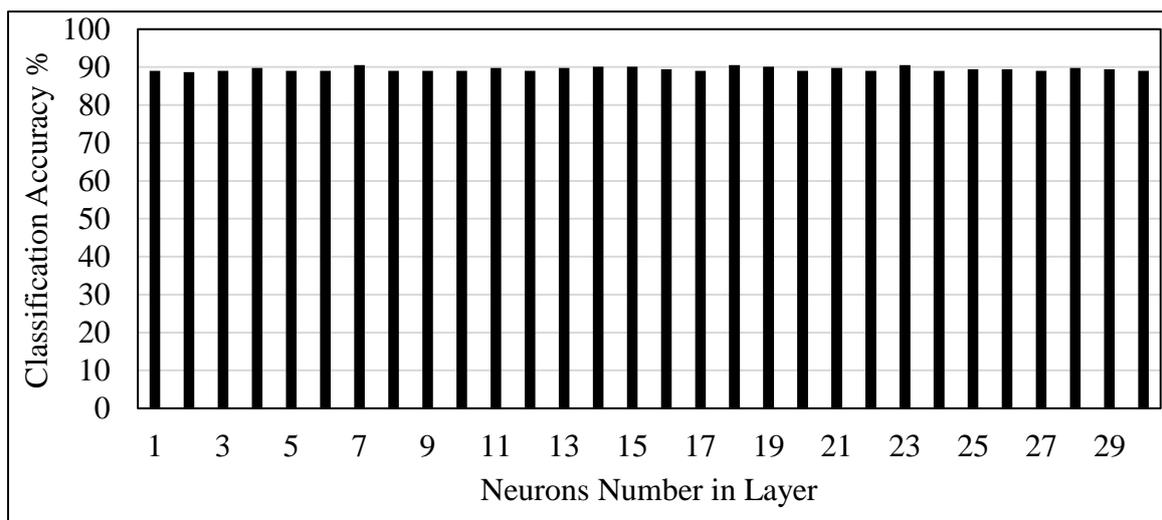
Figure 5.21 Classification accuracy results for applying the proposed system based on one-layer NNs corresponding to the number of neurons in the layer to discriminate low grades (I, and II) against high glioma grades (III, and IV) using BRATS2015 dataset.
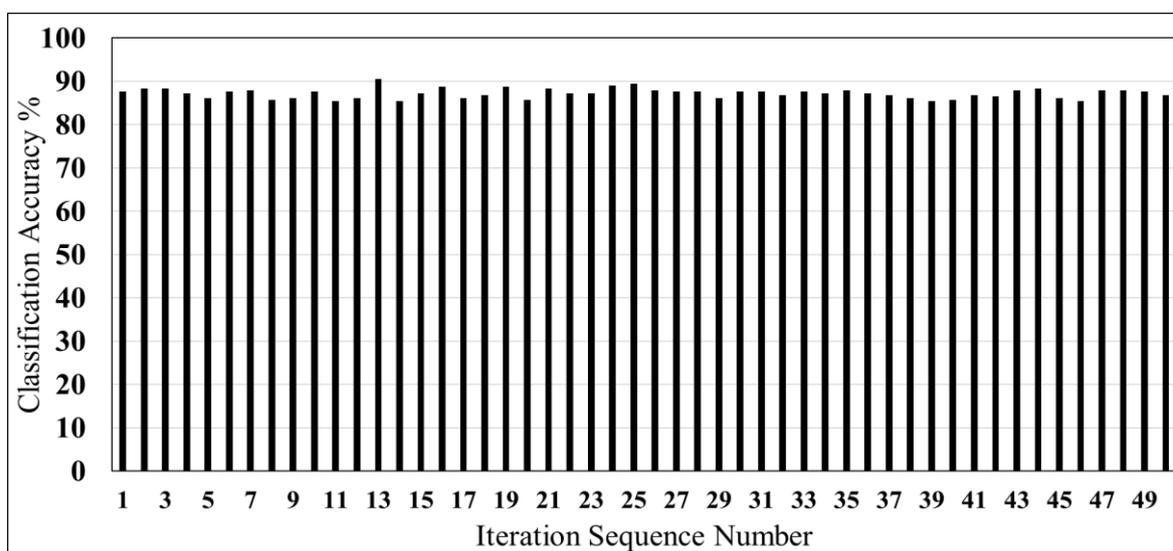
Figure 5.22 Classification accuracy results for applying the proposed system corresponding to the iteration sequence number using the seven neurons in the one-layer NNs to discriminate low grades (I, and II) against high glioma grades (III, and IV) using the BRATS2015 dataset.

5.8.1.1 Redundancy Analysis and Selection of Classifiers

The SC algorithm is applied in this experiment to select the best set of classifiers as well as to eliminate redundant classifiers. The results obtained from using the proposed MTMCS based on one-layer NNs reveal that the best classification accuracy is achieved at 91.24% by Run10 (Table 5.10), where the highlighted classifiers are the optimal selected set of classifiers (Table 5.10). There is also Run1 that shows a slightly lower accuracy at 90.51% where DT classifier is removed, and the classification accuracy continues at the same level

of accuracy at 90.51%. The results indicate that the SC algorithm introduces a significant reduction in the classifier dimensions that they reduced from eleven classifiers to seven classifiers with 91.24% of classification accuracy.

Table 5.10 Selection process of the best set of classifiers with the corresponding classification accuracy *ACCnew* for each selected set of classifiers.   Where 1 and 0 refer to the keep and removing actions of a classifier respectively, Run represents running the process for each selected case. This test is classifying between low grades (II, III) and high grades (III, IV) using the BRATS2015 dataset.

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run11
KNNF	1	1	1	1	1	1	1	1	1	1	1	0
SVMQ	1	1	1	1	1	1	1	1	1	1	0	0
SVML	1	1	1	1	1	1	1	1	1	0	1	1
SVMCUB	1	1	1	1	1	1	1	1	0	1	1	1
KNNW	1	1	1	1	1	1	1	0	1	1	1	1
SVMG	1	1	1	1	1	1	0	1	1	1	1	1
KNNCUB	1	1	1	1	1	0	0	0	0	0	0	0
KNNMED	1	1	1	1	0	0	0	0	0	0	0	0
KNNCOS	1	1	1	0	1	1	1	1	1	1	1	1
LDA	1	1	0	1	1	1	1	1	1	1	1	1
DT	1	0	0	0	0	0	0	0	0	0	0	0
ACCnew	90.51	90.51	89.42	89.42	90.88	90.88	90.51	90.51	90.51	90.51	91.24	88.69

It is noted that the classification accuracy of the proposed MTMCS is improved from 90.5% to 91.24% when the optimal set of classifiers is selected based on the SC method where the classification performance in terms of classification accuracy, sensitivity, specificity, precision, F-measure are 91.24%, 97.27%, 66.67%, 92.20%, 94.69% respectively. This achievement in the classification accuracy is obtained for applying the proposed MTMCS based on one-layer NNs using 9 neurons and the 48th iteration in the design of NNs.

5.8.1.2   Comparison with Other Methods

The overall classification performance in terms of classification accuracy shows that the proposed MTMCS based on one-layer NNs has achieved an accuracy of 91.24% outperforming all other single and ensemble classification models. The next highest classification accuracy compared to all other the traditional approaches were achieved at 85.77% by the KNNF classifier. The next lowest accuracy is achieved by SVMQ at 85.40%,

while the lowest classification accuracy is obtained by DT at 71.53%. With respect to the ensemble systems, the averaged classification accuracy is around 80%, where the majority voting was the one that shows the best accuracy at 84.67% compared to the other ensemble approaches (Figure 5.23).
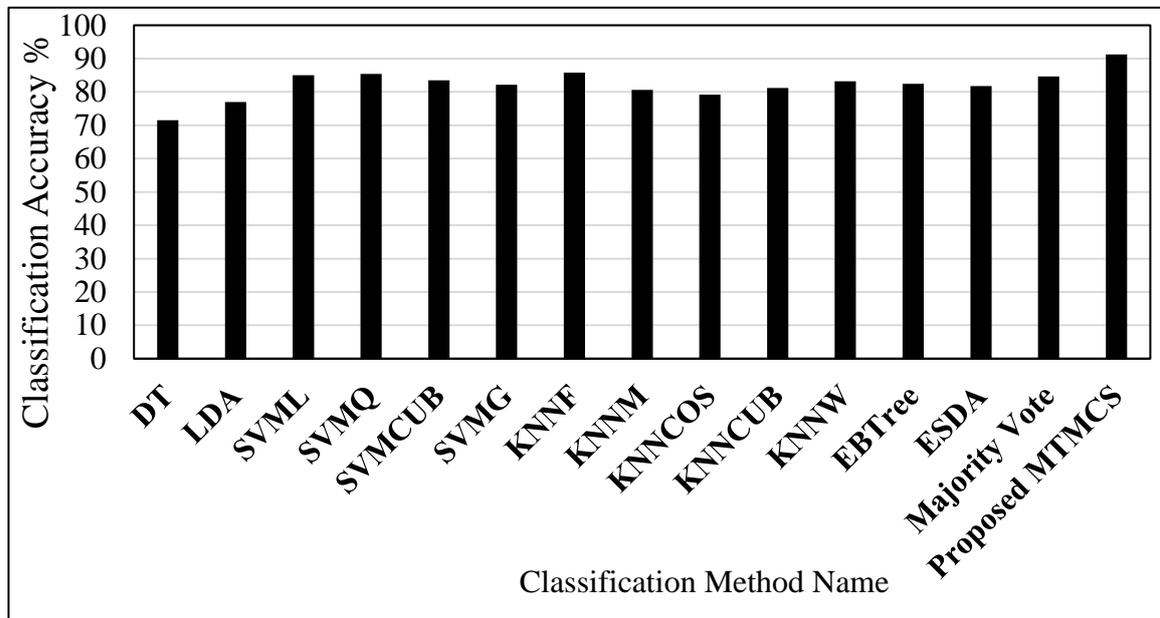
Figure 5.23 Comparative results in terms of the classification accuracy of the proposed MTMCS based on one-layer NNs against all other single classification models and ensemble methods, for the discrimination between low glioma grades (I, and II) and the high grades (III, and IV) using the BRATS2015 dataset.

5.8.2 Experiment 2: Implementing the Proposed MTMCS Based on the Selected Set of Features

The results obtained from experiment 1 show a significant accuracy at 90.5% for the proposed MTMCS using the full set of the textural features measured from 3DGLCM, after which, the classification accuracy has improved to 91.24% using the SC method. To investigate the influence of the selected set of features chosen by the proposed HFSA on the classification performance of the proposed MTMCS, the selected features were fed to the first stage of the proposed system and then the classification system is implemented and evaluated. The evaluation starts with performing the proposed MTMCS based on one-layer NNs. This was conducted by training and testing the classifier members using the selected set of features, then the outcome of these classifiers was combined using the proposed methodology. Considering the resultant confusion matrix of the proposed MTMCS, the results indicate that the best classification performance in terms of the classification

accuracy, sensitivity, specificity, precision, F-measure are 89.42%, 96.36%, 61.11, 91, 93.59% respectively (Table 5.11).

Table 5.11 Evaluation of the classification performance for the proposed system based on one-layer NNs. This is for the discrimination between low grades glioma (I, and II) and the high grades (III, and IV) using the BRATS2015 dataset where Class1 and Class0 refer to high (III, IV), and low grade (I, II) respectively.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
Proposed MTMCS	Class0	33	21	61.11	80.50	69.47	89.42
	Class1	8	212	96.36	91.00	93.59	

It was noted experimentally that the design that achieved the best classification accuracy at 89.42% is based on using 21 neurons in the layer of the NNs, and at the 5th iteration (Figure 5.24 and Figure 5.25). There are also other numbers of neurons that lead to slightly lower classification accuracy, at 89% with the use of 11 or 17 neurons in the layer. With respect the investigation in the iterations, the results show that different iterations such as at 16th reflect the next best classification accuracy at 88%. While all other iterations show lower classification accuracy and most of them reflect accuracies averaged around 85% (Figure 5.25).
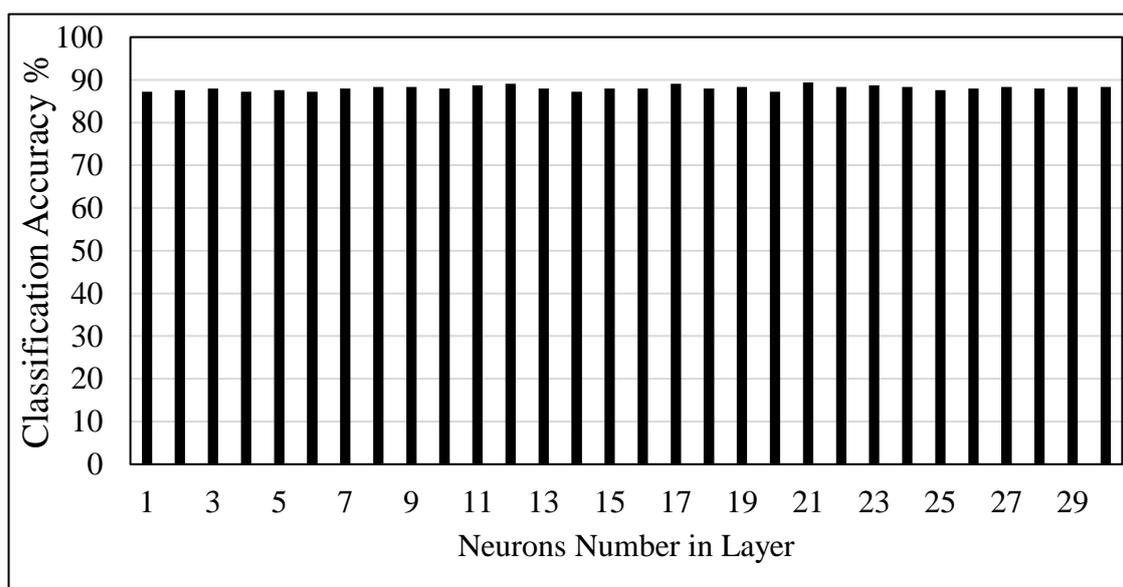
Figure 5.24 Classification accuracy results for applying the proposed system using one-layer NN corresponding to the number of neurons in the layer to discriminate low grades (I, and II) against high glioma grades (III, and IV) using BRATS2015.
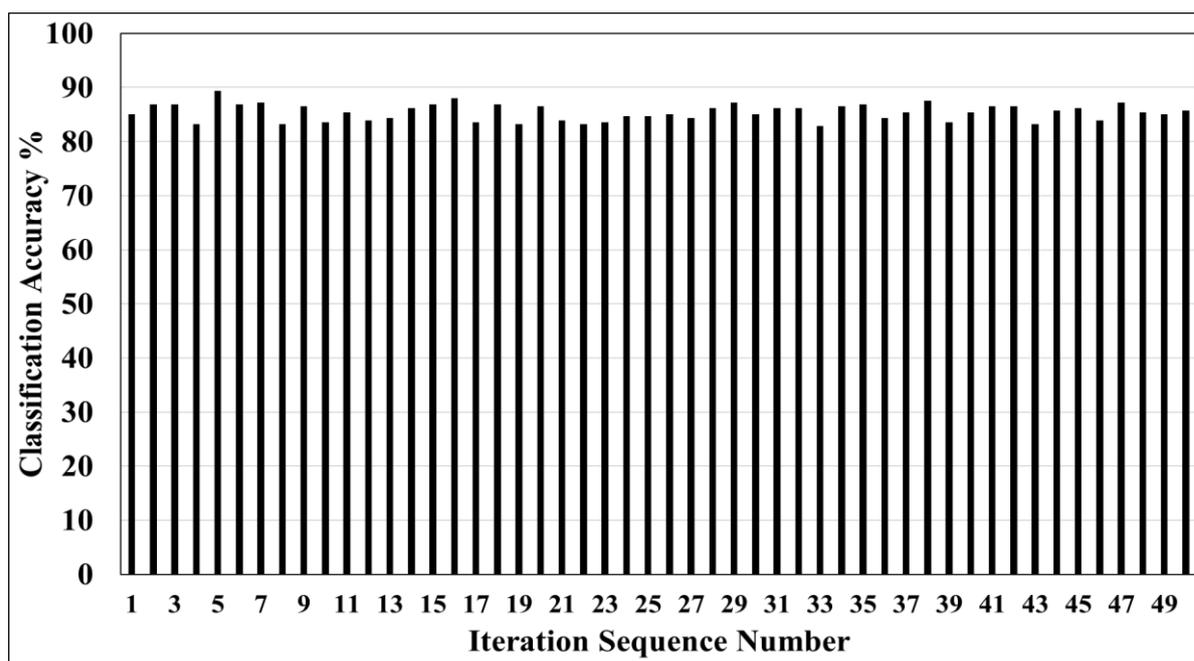
Figure 5.25 Classification accuracy results corresponding to the iteration sequence number using 21 neurons in the one-layer NNs to discriminate low grades (I, and II) against high glioma grades (III, and IV) using BRATS2015.

The results obtained by applying the proposed system based on the two-layer NNs indicate that that the best design achieved a classification accuracy of 90.51% where the sensitivity, specificity, precision, and F-measure of the proposed system are 96.82%, 64.81%, 91.8%, 94.24% respectively (Table 5.12).

Table 5.12 Evaluation of the classification performance for the proposed system based on the two-layer NNs. This is for the discrimination between low grades glioma (I, and II) and the high grades (III, and IV) using the BRATS2015 dataset where Class1 and Class0 refer to high (III, IV), and low grade (I, II) respectively.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
Proposed MTMCS	Class0	35	19	64.81	83.30	72.91	90.51
	Class1	7	213	96.82	91.80	94.24	

It is observed experimentally that the best achievement of 90.51% is obtained using the 9th iteration, which reflects the highest classification accuracy compared to all other iterations. It is also noted that the number of neurons of this achievement is 5, and 22 in the first and the second layer of NNs respectively (Figure 5.26). There are many other choices of neurons

number that revel close results, for example, the first, and second layers of NNs, when using 11 and 4 respectively achieved an accuracy of 89.78% and using 15 and 28 neurons respectively give an accuracy of 89.41%. Most of the other choices of number of neurons present similar accuracies around 88% and 89%.
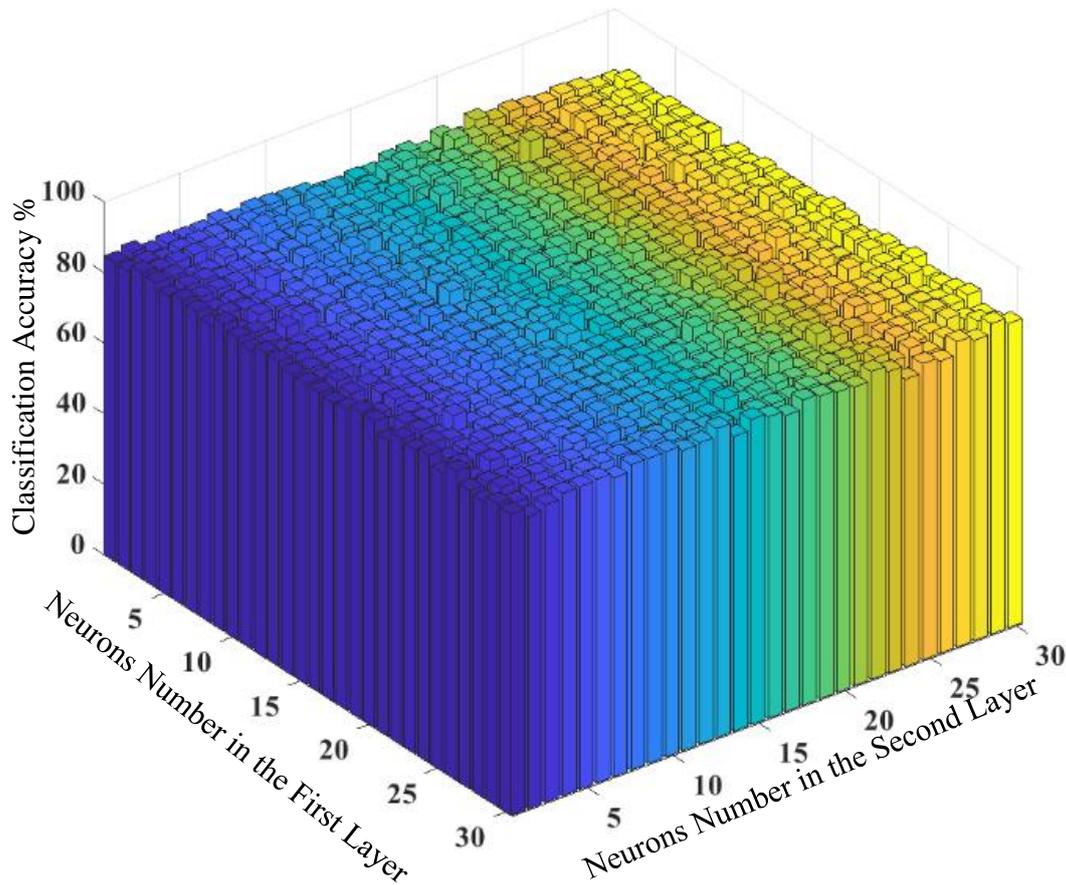
Figure 5.26 Classification accuracy results for the proposed MTMCS using two-layer NNs corresponding to number of neurons per the first and second layers in the left and right axis respectively. This test is to classify low grade glioma (I, II) versus the high grades (III, IV) using the BRATS2015 dataset. The optimal selected design was found at the 9th iteration with 5 neurons in the first layer, and 22 neurons in the second layer.

5.8.3 Comparing the Results of Experiment 1 and Experiment 2

Experiment 1 includes the evaluation and implementation of the full set of features presented by the 3D texture features and these features were used in the training and testing of the all classifiers at the first stage of the proposed MTMCS. Experiment 2 covers the evaluation and implementation of the selected set of features chosen by the proposed HFSA. The classification performance of both experiment 1 and 2 have shown good accuracy outweighing all other single and ensemble classifiers, but there is a slight difference in the

classification accuracy between them - the obtained accuracy is 90.51 and 89.42 respectively. To explain this difference further, the results obtained by both experiments are compared, and the classification performances are analysed in terms of sensitivity of high grades (Figure 5.27), sensitivity of low grades (specificity) (Figure 5.28), F-measure of high grades and low grades (Figure 5.29, and Figure 5.30) respectively. The outcome of the integrated classifiers is used as an input to the second stage of the proposed MTMCS where it is observed that the outcome of the classifiers has various behaviour after using the proposed HFSA. Notably, there is an improvement in the classification accuracy in some classifiers including the KNNF classifier due to the use of the proposed HFSA, where the classification accuracy is enhanced from 85.77% to 87.96%. The selected set of features influence the classifiers differently, for example, the classifiers SVMCUB, KNNM, KNNW have shown a small reduction in the sensitivity of high grades (Figure 5.27), while the sensitivity of low grade is also reduced in some other classifiers such as the LDA and SVMQ classifiers (Figure 5.28). Performance evaluation in term of F-measure of both high and low grades between the two experiments also confirms that using the proposed feature selection method has a variable impact on different classifiers by improving the outcome of some classifiers while not others.

To sum up, using a different set of features enables a single classification system to improve the performance of some classifiers in term of classification accuracy and not others. Different set of features has a different impact on the performance of the classifiers. Due to the proposed MTMCS depending on the outcome of the fusion of all members in the ODM, the reduction in the sensitivity of low and high grades of some members has slightly degraded the final classification accuracy obtained by the proposed MTMCS. On the other hand, when the proposed system is implemented based on the two-layer the classification accuracy is increased to 90.5%. This shows that the proposed MTMCS has a stable accuracy but in some circumstances, it is necessary to investigate further layers that can lead to superior results in the classification accuracy. Hence, to gain further improvement in the classification accuracy, the proposed system has been developed based on utilising the full set of features that includes the textural features and the proposed features associated with tumour descriptors. Also, the development in the automated system will explore enabling diversity in features space, which can lead to achieve further improvement in the classification accuracy.
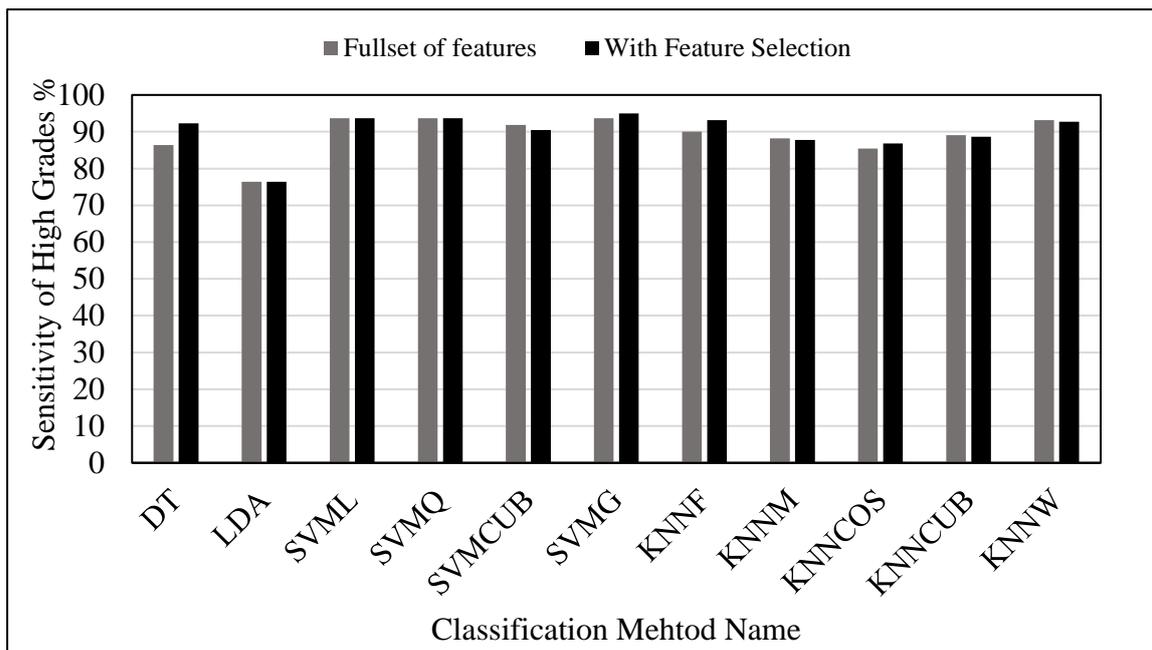
Figure 5.27 Classification results in term of the sensitivity of high grades to show the difference in behaviour between using the fullest of features (3D texture features) and after applying the selected features using the proposed HFSA for the discrimination between low Glioma grades (I, and II) and the high grades (III, and IV) using BRATS2015.
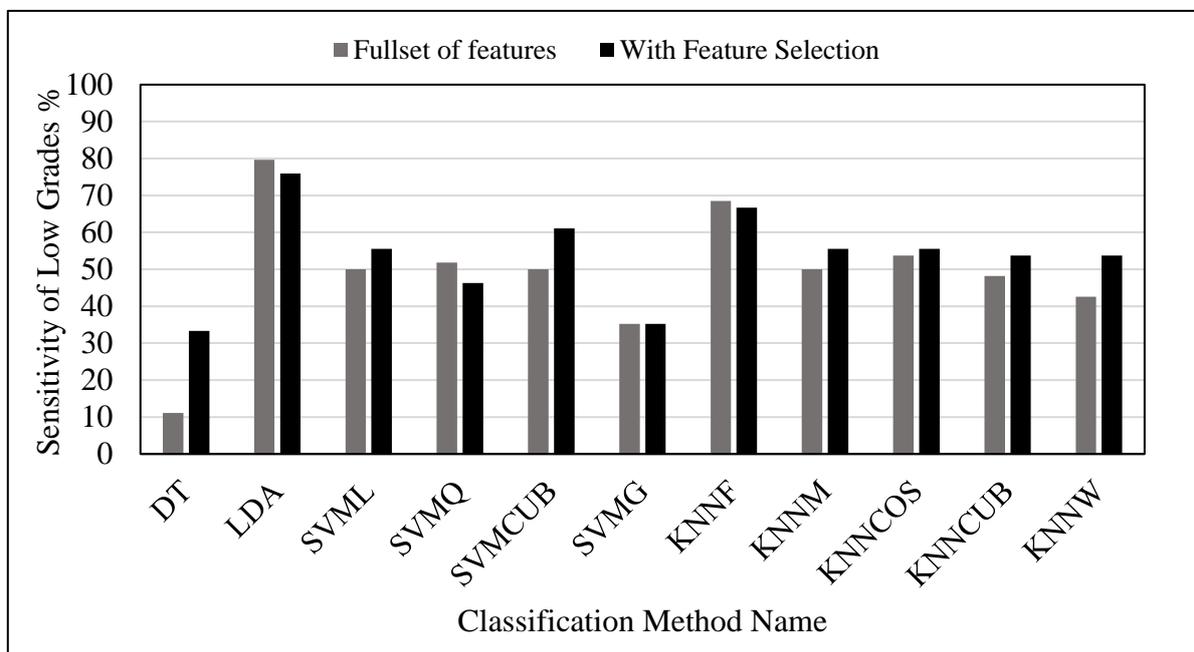
Figure 5.28 Classification results in term of the sensitivity of low grades to show the difference in behaviour between using the fullest of features (3D texture features) against applying the selected features using the proposed HFSA for the discrimination between low glioma grades (I, and II) and the high grades (III, and IV) using BRATS2015.
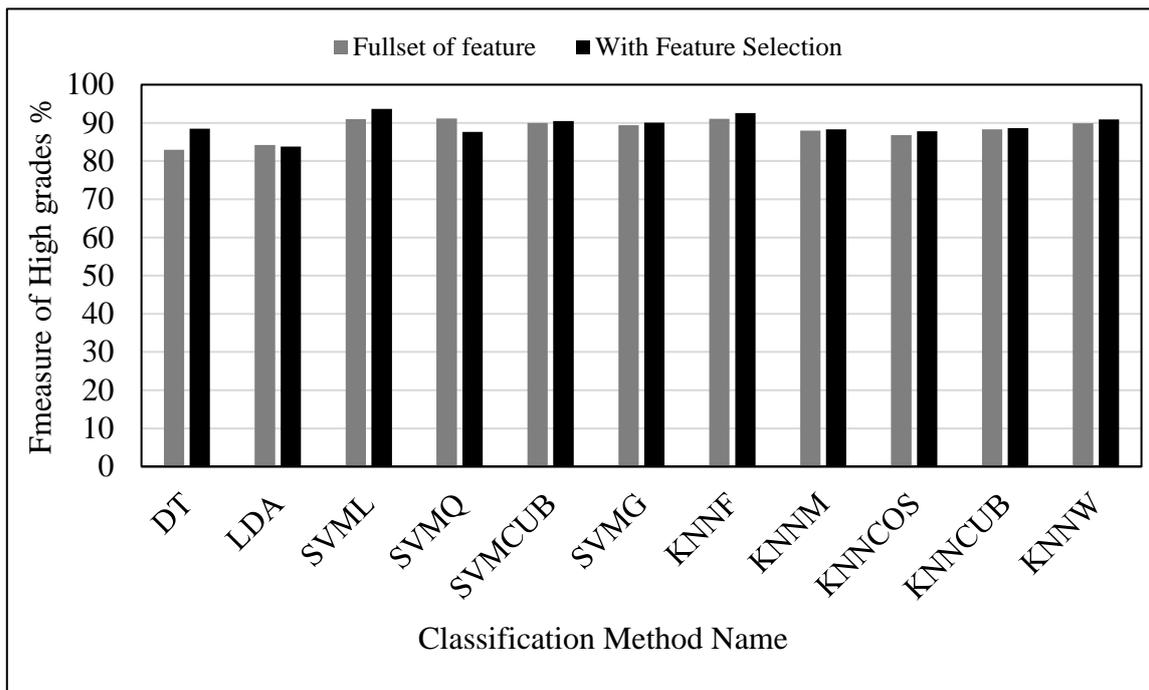
Figure 5.29 Classification results in term of the F-measure of high grades to show the difference in behaviour between using the fullest of features (3D texture features) and after applying the selected features using the proposed HFSA for the discrimination between low glioma grades (I, and II) and the high grades (III, and IV) using BRATS2015.
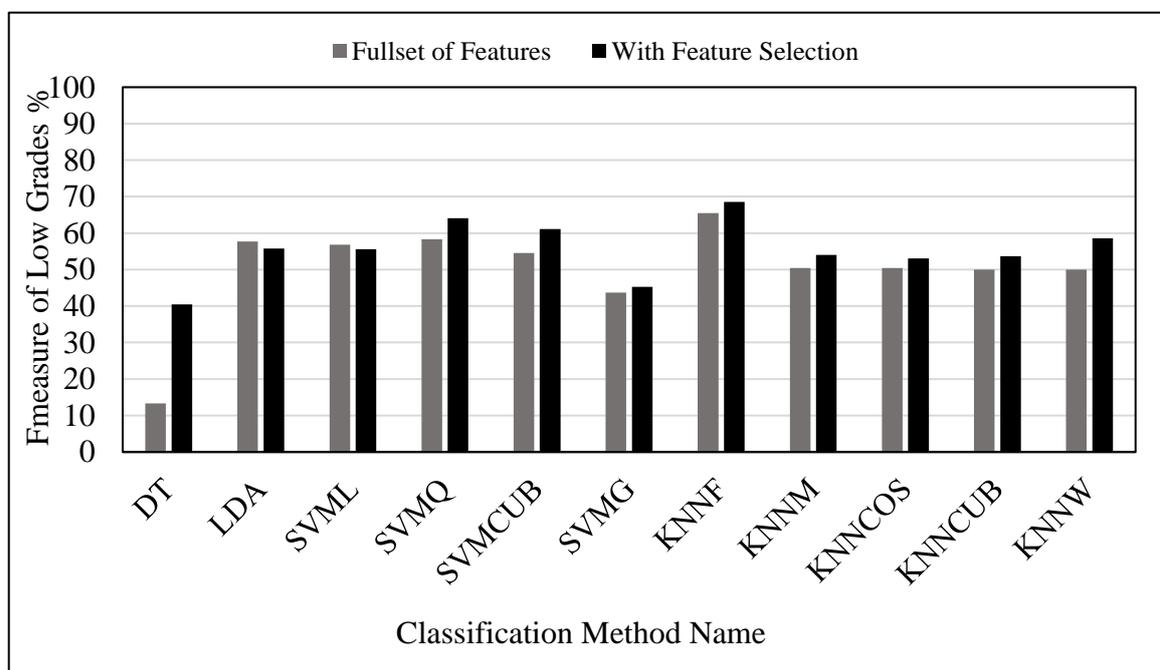
Figure 5.30 Classification results in term of the F measure of low grades to show the difference in behaviour between using the fullest of features (3D texture features) and after applying the selected features using the proposed HFSA for the discrimination between low Glioma grades (I, and II) and the high grades (III, and IV) using BRATS2015.

5.8.4 Experiment 3: Develop the Proposed System Using the Textual Feature and the Proposed Feature Associated with Tumour Descriptors

The proposed MTMCS is developed based on all the available features set extracted from 3DGLCM and features associated with tumour descriptors (FTD). The purpose of this experiment is to assess the ability of the proposed system to classify the glioma grades based on the combination of all the textural features and proposed features extracted from the tumour descriptors (TD).

The performance evaluation of the classification system starts with implementing the proposed MTMCS based on one-layer NNs. The results illustrate that the achieved classification accuracy is of 92.70% with the use of 7 neurons and second iteration (Figure 5.31 and Figure 5.32). While the next best classification accuracy is achieved at 92.34% using 17 neurons, other neurons number also shows good accuracies at 92% with such as 2, 3 and 4 neurons.
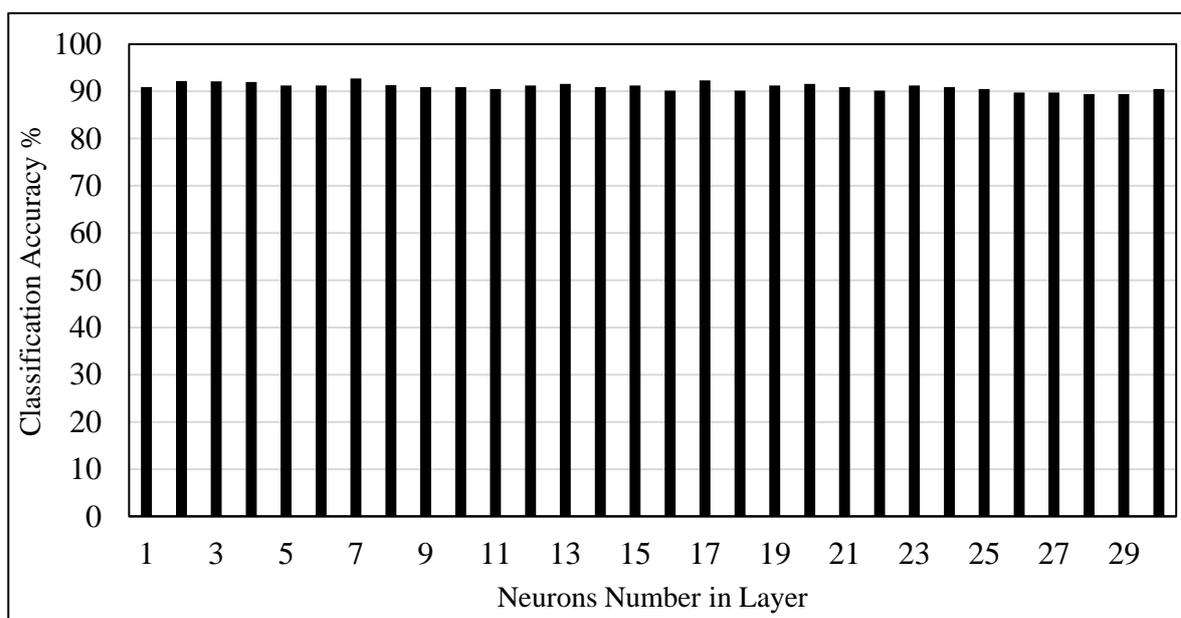
Figure 5.31 Classification accuracy results for demonstrating the proposed MTMCS using one-layer NNs corresponding to the number of neurons in the layer to discriminate between low grades (I, and II) and high glioma grades (III, and IV), based on the full set of features 3DGLCM and FTD derived from BRATS2015.
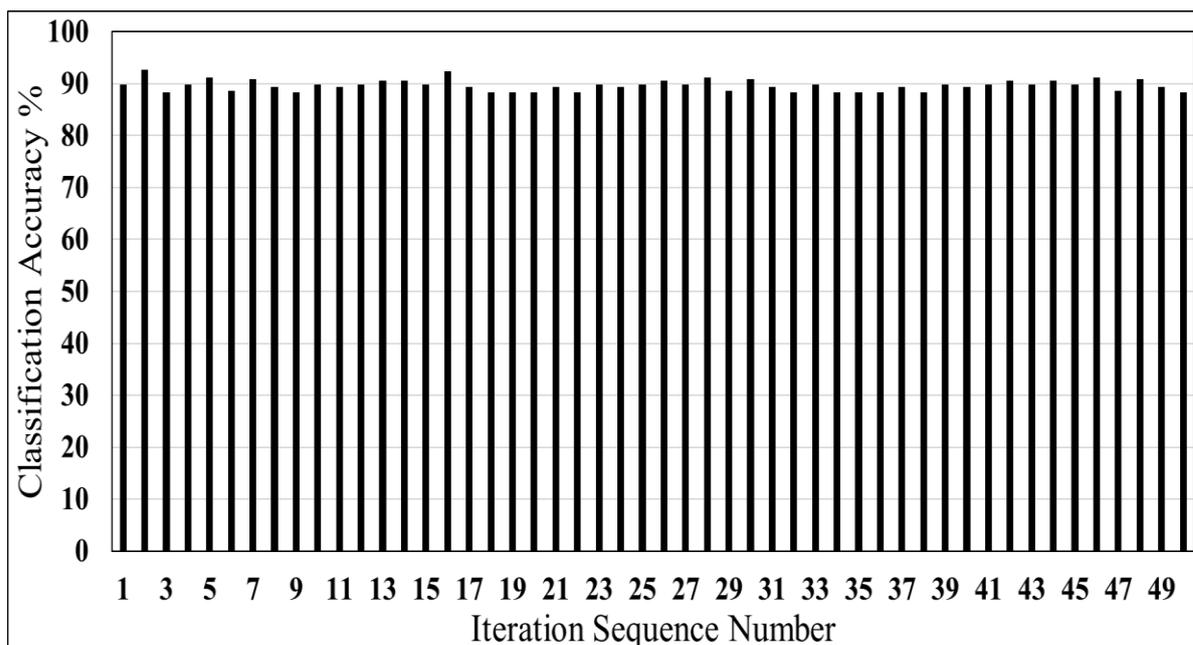
Figure 5.32 Classification accuracy results corresponding to the iteration sequence number using one-layer NNs based on the full set of features 3DGLCM and TD to classify low grades against high grade glioma using BRATS2015.

5.8.4.1 Redundancy Analysis and Select Classifiers

The results obtained from applying the SC algorithm to this experiment indicate that the classification accuracy has slightly improved from 92.7% to 93.07%, where this enhancement is achieved by both Run9 and Run10 (Table 5.13). However, Run10 includes a smaller number of classifiers. Therefore, it is a better choice, and thus it is selected for the proposed system. Run11 also shows a good classification accuracy at 92.7% where it has lower dimensions of classifiers as only five classifiers are needed for this accuracy.

The other significant achievement in this experiment is that the classifier dimensions have been reduced from eleven classifiers to six classifiers that are highlighted in Table 5.13. While maintaining a high classification accuracy at 93.07%, these classifiers are namely SVML, SVMG, KNNF, KNNM, KNNCOS and LDA. The classification performance for the proposed MTMCS for this achievement by Run10 in terms of sensitivity, specificity, precision and F-measure are 99.09%, 68.52%, 92.8%, 95.82% respectively (Table 5.14).

Table 5.13 Selection process of best set of classifiers with the corresponding classification accuracy *ACCnew* for each selected set of classifiers. Where 1 and 0 refer to the keep and removing actions of classifier respectively, Run represents running the process for each selected case. This test is classifying low grades (II, III) versus high grades (III, IV) using the BRATS2015 dataset based on the full set of features (3DGLCM and FTD).

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run11
SVML	1	1	1	1	1	1	1	1	1	1	1	0
SVMQ	1	1	1	1	1	1	1	1	1	1	0	0
KNNW	1	1	1	1	1	1	1	1	1	0	0	0
SVMCUB	1	1	1	1	1	1	1	1	0	0	0	0
SVMG	1	1	1	1	1	1	1	0	1	1	1	1
KNNCUB	1	1	1	1	1	1	0	0	0	0	0	0
KNNF	1	1	1	1	1	0	1	1	1	1	1	1
KNNM	1	1	1	1	0	1	1	1	1	1	1	1
DT	1	1	1	0	0	0	0	0	0	0	0	0
KNNCOS	1	1	0	1	1	1	1	1	1	1	1	1
LDA	1	0	1	1	1	1	1	1	1	1	1	1
ACCnew	92.7	91.97	92.34	92.7	92.34	91.97	92.7	92.34	92.7	93.07	93.07	92.7

Table 5.14 Evaluation of the classification performance for the proposed system based on the one-layer NNs. This is for the discrimination between low grades glioma (I, and II) and the high grades (III, and IV) using the BRATS2015 dataset where Class1 and Class0 refer to high (III, IV), and low grade (I, II) respectively.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
Proposed MTMCS	Class0	37	17	68.52	94.90	79.57	93.07
	Class1	2	218	99.09	92.80	95.82	

5.8.4.2   Comparison with Other Methods

It is observed from the comparison of the proposed MTMCS versus the other single and ensemble classifiers that the proposed system based on the combination of the textural feature and the proposed features derived from the tumour descriptors has achieved a superior classification accuracy at 93.07 % compared to all other classification approaches (Figure 5.33, and Table 5.15). Among the single classifiers, SVML achieved the best result at 90.15. Similar classification accuracy was obtained by using the MCS based on majority vote. This indicates that the integration of these features with the proposed methodology enables the majority vote to achieve a good classification accuracy at 90.15%. Following these results, the best accuracy was achieved by SVMQ at 89.78%. While all other classifiers result is between the lowest accuracy at 77% achieved by LDA to 88.3% by KNN classifiers (Figure 5.33, and Table 5.15).
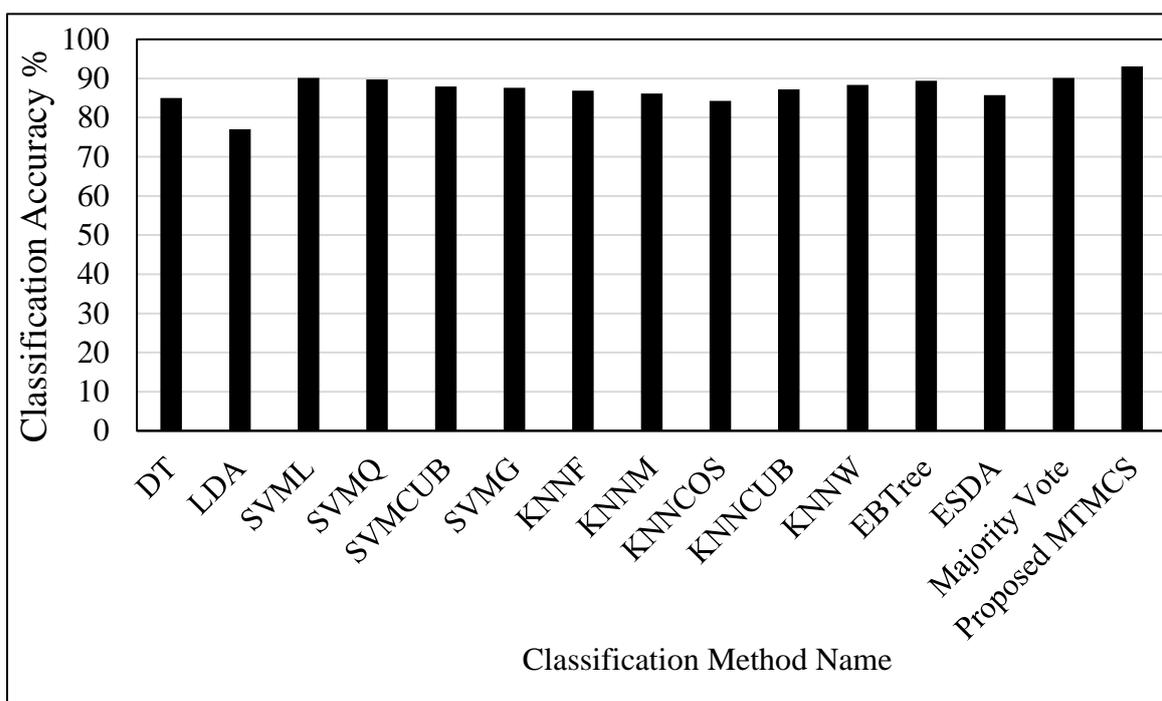
Figure 5.33 Comparative evaluation in term of classification accuracy for the proposed MTMCS system versus the other single and ensemble classifiers to classify low against high grade glioma using BRATS2015 based on the full set of features of 3DGLCM and proposed tumour features. The proposed MTMCS is performed based on the one-layer NNs

Table 5.15 A comparative evaluation results of the proposed MTMCS based on one-layer NNs versus all other classification approaches and the majority vote. This is for the classification of low-grade glioma (I, II) versus the low-grade (III, IV) based on the full set of features of 3DGLCM and FTD that derived from BRATS2015.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
DT	Class0	26	28	48.15	66.67	55.91	85.04
	Class1	13	207	94.09	88.09	90.98	
LDA	Class0	43	11	79.63	45.26	57.71	77.01
	Class1	52	168	76.36	93.85	84.21	
SVML	Class0	31	23	57.41	88.57	69.66	90.15
	Class1	4	216	98.18	90.38	94.11	
SVMQ	Class0	36	18	66.67	78.26	72.00	89.78
	Class1	10	210	95.45	92.11	93.75	
SVMCUB	Class0	35	19	64.81	71.43	67.96	87.96
	Class1	14	206	93.64	91.56	92.58	
SVMG	Class0	26	28	48.15	81.25	60.46	87.59
	Class1	6	214	97.27	88.43	92.64	
KNNF	Class0	36	18	66.67	66.67	66.66	86.86
	Class1	18	202	91.82	91.82	91.81	
KNNM	Class0	30	24	55.56	68.18	61.22	86.13
	Class1	14	206	93.64	89.57	91.55	
KNNCOS	Class0	33	21	61.11	60.00	60.55	84.31
	Class1	22	198	90.00	90.41	90.20	
KNNCUB	Class0	32	22	59.26	71.11	64.64	87.23
	Class1	13	207	94.09	90.39	92.20	
KNNW	Class0	30	24	55.56	78.95	65.21	88.32
	Class1	8	212	96.36	89.83	92.98	
EBTree	Class0	34	20	62.96	79.07	70.10	89.42
	Class1	9	211	95.91	91.34	93.56	
ESDA	Class0	36	18	66.67	63.16	64.86	85.77
	Class1	21	199	90.45	91.71	91.07	
Majority Vote	Class0	34	20	62.96	82.93	71.57	90.15
	Class1	7	213	96.82	91.42	94.03	
Proposed MTMCS	Class0	37	17	68.52	94.87	79.57	93.07
	Class1	2	218	99.09	92.77	95.82	

5.8.5   Experiment 4: Develop the Proposed System Based on the Diversity in Feature Space

In this experiment, the proposed system is developed by enabling the diversity in the output decision that comes from the classifiers, which can be generated by dividing the features space into a different subset of features. The produced subsets of features are utilised in the training and testing phases of the classifier members where the outcomes of these members are employed to enlarge the ODM. Consequently, each patient in the dataset will have various output decisions developed to improve the classification accuracy thus increasing the chance of producing correct decisions from different subsets of the feature. The rationale for this development in the proposed MTMCS is to investigate the influence of activating diversity in the features space, which can lead to further enhancement in the classification accuracy of glioma grades.

In this experiment, the BRATS2015 dataset has been used, which has 274 patients with 54 patients of low grades and 220 patients of high-grade glioma. In the first stage of this test, the proposed classification system is updated by generating further output decision vectors produced from a different subset of features based on the strategy of dividing the features set to a multi subset of features. This leads to increasing the diversity of the output decision vectors, which can overcome the misclassified samples that could occur due to weak classifiers. Several subsets of features are produced and used to train and test the classifiers (Figure 5.34). These subsets of features are generated as follows; the proposed features related to tumour descriptors (FTD) that includes 8 features, the 2DGLCM with 72 features, the 3rd part of GLCM with 162 features, the full set of 3DGLCM with 234 features, and the overall features set with 242 features (Figure 5.34). Each subset is utilised individually to train and test each one of the classifier members and the output decisions for each one of them are constructed and combined in the ODM. Consequently, the ODM will have the dimensions (11 output decision vectors $\times$5 sets of features $\times$ number of patients) and thus 55 classifiers are constructed to be used in the second stage of proposed MTMCS.

TD (8 features) → (11 Classifiers)

2DGLCM (4 angles × 18 features) → (11 Classifiers)

3rd GLCM (9 angles × 18 features) → (11 Classifiers)

2D, 3rD (13 angles × 18 features) → (11 Classifiers)

TD, 2D, 3rD (8+72+162) → (11 Classifiers)

(11 output decision vectors ×5 subsets × number of patients)
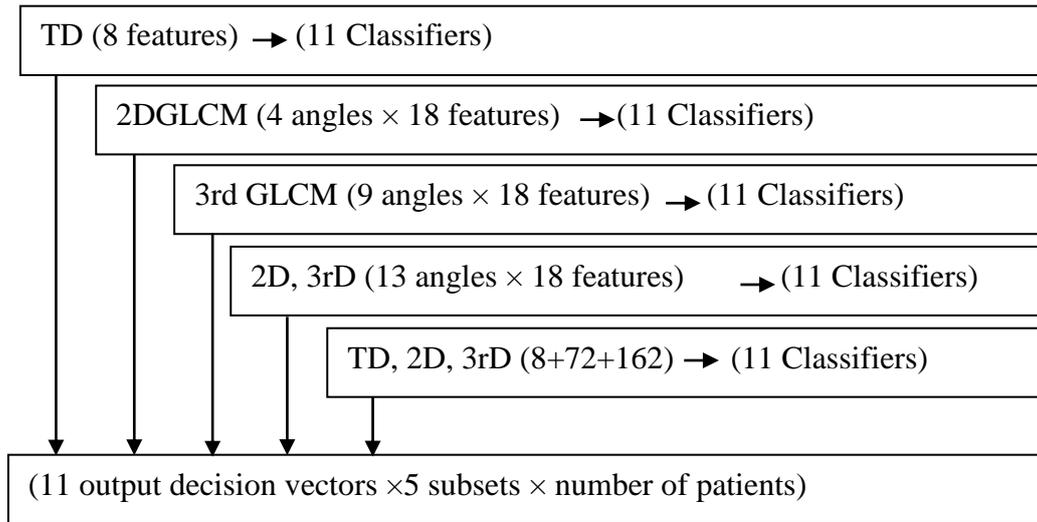
Figure 5.34 The diagram of the generation of the output decision vectors conducted based on the diversity in feature space. Eleven output decision vectors refer to the eleven classifiers involved in the first stage of the proposed system, which applied to five subsets of features resulting in fifty-five classifiers being built.

To further elaborate, the total dimensions of the output decisions vector are calculated as follows; number of classifiers used × number of features subset × number of patients. Accordingly, 274 samples × 55 classifiers equal 15070 tests that are performed in the first stage of the proposed MTMCS. The following stage is the training and testing of the ODM based on the proposed methodology of MTMCS.

In the second stage of the proposed MTMCS, for the implementation of one-layer NNs, 274 tests × 30 neurons × 50 iterations are investigated, as a result, in total, 411000 tests are implemented and evaluated. The hardware used in this experiment is Core i7, RAM 16, and MATLAB 2018. For the one-layer NNs, the total time spent for the implementation of the second stage of the proposed MTMCS, including the cross-validation, training and testing of 274 samples (each sample represented by 55 classifiers) × 50 iterations × 30 nodes is approximately 65375.523 seconds (18.159 hours), which is reduced practically using parallel processing and realistically becomes about 9387.871266 seconds (2.6 hours). The average execution time for training N-1 samples and testing one sample is 0.159 ± 0.033 seconds, where N refers to the total number of samples.

The results obtained from testing for the proposed MTMCS based on one-layer NNs exhibit that the best classification performance achieved in terms of the classification accuracy,

sensitivity, specificity (or sensitivity of low grades), precision, and F-measure are 93.07%, 98.64%, 70.37%, 93.1%, and 95.80% respectively (Table 5.16).

Table 5.16 the results of the performance evaluation of the proposed MTMCS based on one-layer NNs incorporated the diversity in features space using the BRATS2015 dataset. This test is to classify between low grades glioma (I, II) and high grades glioma (III, IV). Class0 and Class1 refer to low and high grades respectively.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
Proposed MTMCS	Class0	38	16	70.37	92.68	80.00	93.07
	Class1	3	217	98.64	93.13	95.80	

It is noted that the highest classification accuracy of 93.07% is achieved by the implementation of one-layer NN through the use of two neurons, and through using the 42nd iteration (Figure 5.35 and Figure 5.36). It is also observed that the classification accuracy gradually decreases with the increasing number of neurons in the one-layer NNs. The next best accuracy is achieved at 92.34% by 5, 8 and 9 neurons. There are also other numbers of neurons that show good results such as the use of 1 neuron that reflects 91.97% accuracy and 4 neurons that shows 91.97% accuracy. With respect to the investigation in conducting the iterations, it is observed that there are various numbers of iterations reflecting different classification accuracy, while the next best classification accuracy is obtained by the 24th iteration at 91.97%. There are also other iterations achieved classification accuracies very close to the optimal result, for example the 8th iteration reflects 91.61%, while the 6th, 16th and 22th achieved 91.24%.
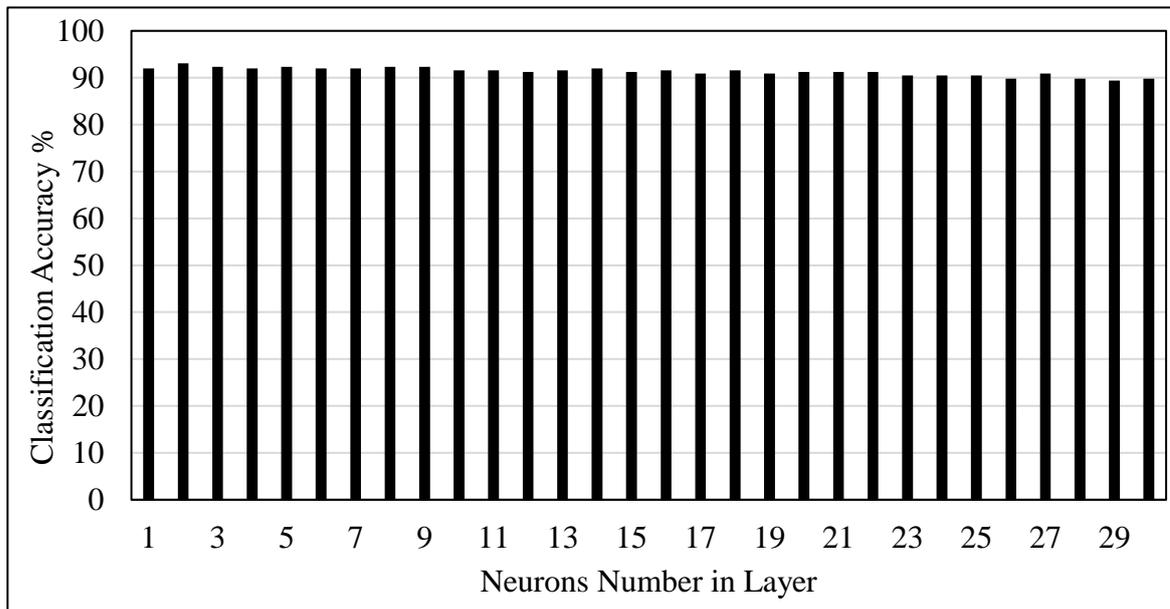
Figure 5.35 Classification accuracy results for the proposed system based on one-layer of NNs corresponding to the number of neurons in layer based on enabling the diversity in features space. This test is to classify between glioma low grades (I, II) and high grades (III, IV) using BRATS2015 dataset
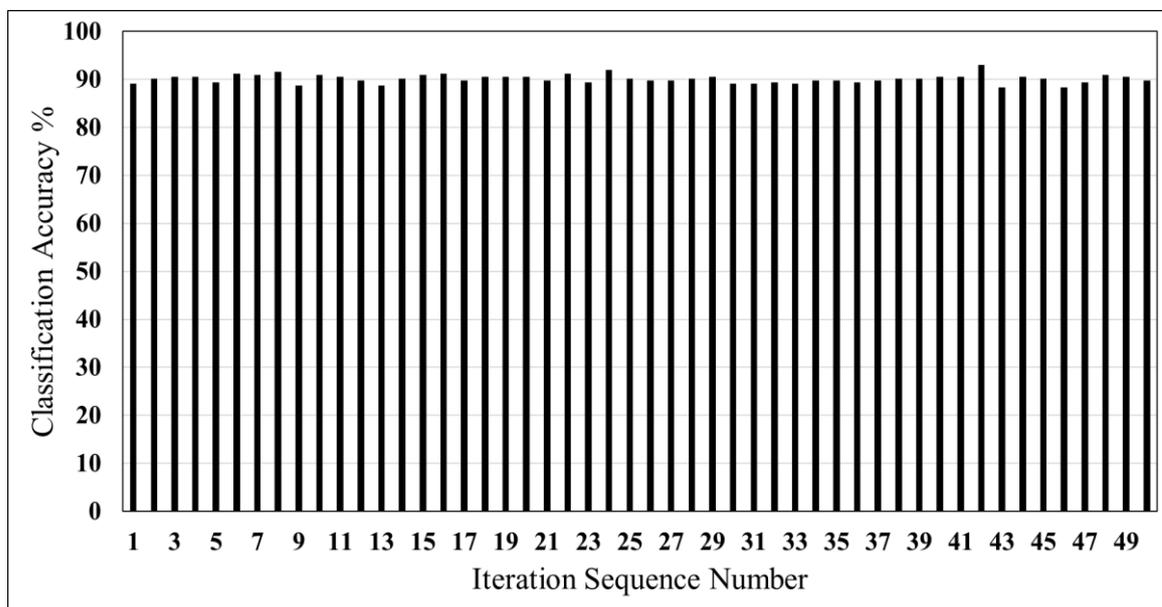
Figure 5.36 Classification accuracy results for the proposed system based on one-layer NNs according to the iteration sequence number for two neurons in the layer, incorporating the diversity in feature space. This test is to classify between glioma low grades (I, II) and high grades (III, IV) using the BRATS2015 dataset.

The two-layers NNs of the proposed MTMCS have also been investigated to seek for potential improvement in the classification accuracy. The results exhibit that the best classification performance achieved for the proposed MTMCS in terms of the classification accuracy, sensitivity, specificity (or sensitivity of low glioma grades), precision, and F-measure are 93.43%, 98.18%, 74.07 %, 93.9 %, and 96% respectively (Table 5.17).

Table 5.17 The results of the performance evaluation of the proposed MTMCS based on two-layer NNs and incorporated the diversity in features space using the BRATS2015 dataset. This test to classify between low grades glioma (I, II) and high grades glioma (III, IV). Class0 and Class1 refer to low and high grades respectively.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
Proposed MTMCS	Class0	40	14	74.07	90.91	81.63	93.43
	Class1	4	216	98.18	93.91	96.00	

It is observed that this achievement in the results is yielded based on the use of 19, and 5 neurons in the first and second layers respectively at the 24th iteration (Figure 5.37).

This experiment was extended to investigate the three-layer NNs for the proposed MTMCS incorporating diversity in the features space. Giving the confusion matrix for the optimal result after applying the proposed MTMCS (Table 5.18), the classification performance in terms of classification accuracy, sensitivity, specificity (or sensitivity of low grades), precision, and F-measure are 93.80%, 99.09%, 72.22%, 93.6%, and 96.24% respectively. These results are achieved by using 6, 18, 6 neurons in the first, second, and third layers respectively at the 10th iteration. It should be noted that it is extremely complicated to visualise a graph with five dimensions to present the obtained results in this experiment from three-layer and different iterations. Therefore, some of the results obtained by conducting the proposed MTMCS based on three-layer NNs and the diversity in the features space are depicted in (Appendix A).
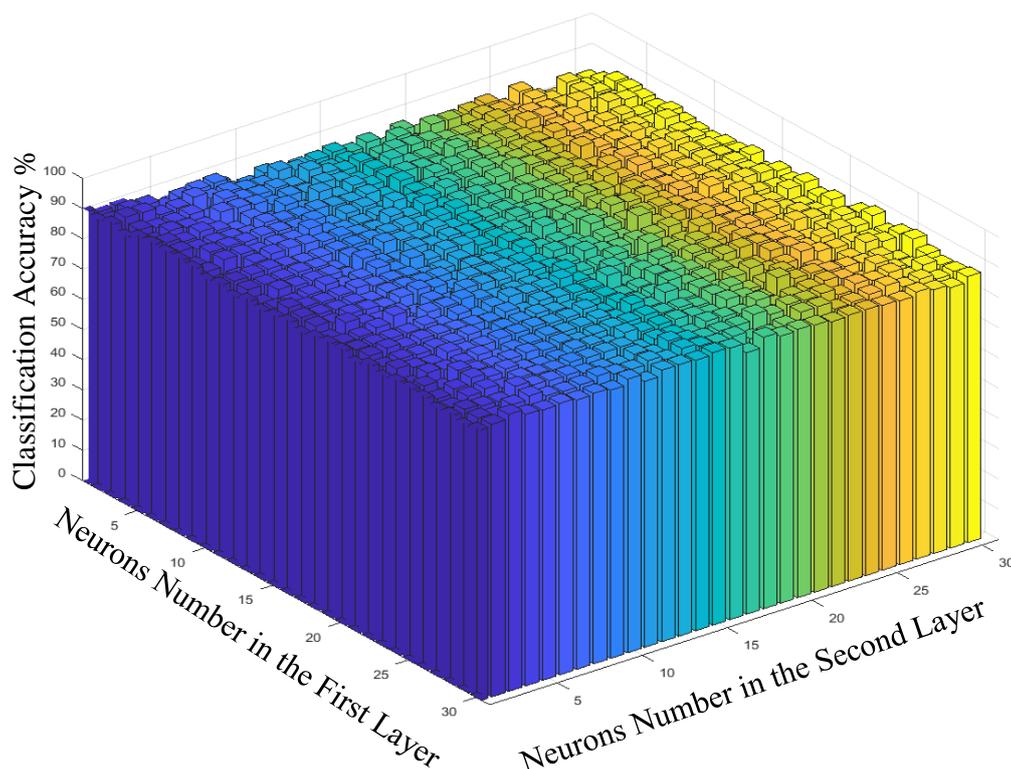
Figure 5.37 Classification accuracy with respect to number of neurons per first and second layers in the left and right side respectively. The proposed MTMCS is implemented based on two-layers NN and incorporated the diversity technique using the BRATS2015 dataset.

Table 5.18 Results of the performance evaluation of the proposed MTMCS based on three-layer NNs with enabling the diversity in features space using the BRATS2015 dataset. This test is to classify between low grades glioma (I, II) and high grades glioma (III, IV). Class0 and Class1 refer to low and high grades respectively.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
Proposed MTMCS	Class0	39	15	72.22	95.12	82.10	93.80
	Class1	2	218	99.09	93.56	96.24	

5.8.5.1   Redundancy Analysis and Select of Best Classifiers

The SC algorithm is applied and evaluated in this experiment for potential achievement in the classification performance of MTMCS based on the full set of features including the textural features associated with GLCM and the proposed features extracted from the tumour descriptors incorporated the activation of diversity in features space. The classifiers in this experiment are applied to five subsets of features where they are defined as follows (Table 5.20); those that are implemented based on the full set of features including FTD, 2DGLCM,

3rd GLCM are indicated by the description: classifier name_ALL. While the rest as follows; Classifier_2D refers to the classifiers implementation based on 2DGLCM. Classifier name_3rd refers to implementing classifiers based on the third part of GLCM. Classifier name_ALL3D refers to the ones conducted based on the features associated with 3DGLCM. Classifier name_TD refers to implementing the classifiers based on the proposed features associated with tumour descriptors. The algorithm starts with sorting the classifiers into descending order (Table 5.20). After the SC algorithm is implemented, the results indicate that the classification accuracy has been improved from 93.07% achieved by the initial Run to 93.43% obtained by Run11, where the dimensions of classifiers are reduced from 55 classifiers to 53 classifiers that are highlighted in yellow and two classifiers are eliminated, these are KNNCOS_3rd and DT_2D classifiers. It is noted that the classification accuracy is increased to its maximum value at 93.80% by Run32 when SVMQA_3rd classifier is eliminated in addition to the other two classifiers (Table 5.20). Accordingly, the best reduction in the classifiers dimensions is gained by Run32 where classifiers are reduced from 55 classifiers to 52 classifiers while maintaining the best classification accuracy at 93.80%. It is also observed that most of the runs conducted by the proposed MTMCS based on one-layer NNs incorporating the SC method reflected comparable accuracies close to the optimal result at 93.80%, for example, the last two runs defined by Run 45 and 55 are showing 93.43% accuracy (Table 5.20). Classification performance in term of sensitivity, specificity, precision and F-measure is illustrated in (Table 5.19), which was achieved by Run32.

Table 5.19 Results of the performance evaluation of the proposed MTMCS based on one-layer NNs and incorporated the diversity in features space and SCA using the BRATS2015 dataset. This test is to classify between low grades glioma (I, II) and high grades glioma (III, IV). Class0 and Class1 refer to low and high grades respectively.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
Proposed MTMCS	Class0	40	14	74.07	93.02	82.47	93.80
	Class1	3	217	98.64	93.94	96.23	

Table 5.20 Selection process of best set of classifiers with the corresponding classification accuracy *ACCnew* for each selected set of classifiers. Where 1 and 0 refer to the keep and removing actions of classifier respectively, Run represents running the process for each selected case. This test is classifying low grades (II, III) versus high grades (III, IV) using BRATS2015 dataset based on the full set of features (3DGLCM and FTD) and diversity in the features space

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run11	Run12	Run13	Run14	Run15	Run16	Run17	Run18	Run19	Run20	Run21	Run22	Run23	Run24	Run25	Run26	Run27	Run28	Run29	Run30	Run31	Run32	Run33	Run34	Run35	Run36	Run37	Run38	Run39	Run40	Run41	Run42	Run43	Run44	Run45	Run46	Run47	Run48	Run49	Run50	Run51	Run52	Run53	Run54	Run55
SVMQA_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
SVML_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
SVMG_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
SVML_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
KNNV_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
SVMCUB_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
KNNF_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
LDA_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
DT_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
SVMG_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
SVML_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
KNNV_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
KNNCUB_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNM_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMQA_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DT_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMQA_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNV_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVML_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMCUB_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNF_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNF_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNF_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMQA_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KNNCOS_ALL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMMG_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNF_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNV_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNV_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVML_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DT_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNM_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMCUB_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNM_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCUB_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMG_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCUB_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMG_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNM_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMCUB_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCUB_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMQA_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCOS_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNM_3rd	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCOS_3rd	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LDA_2D	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCUB_ALL3D	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMCUB_2D	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCOS_ALL3D	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCOS_TD	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DT_2D	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DT_ALL3D	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LDA_3rd	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LDA_ALL3D	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LDA_ALL	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ACCnew	93.07	92.70	92.70	92.34	92.70	93.07	92.70	92.70	92.70	92.70	91.97	93.43	92.70	93.07	92.70	92.70	92.70	92.70	92.70	93.07	92.34	92.70	93.07	92.70	92.70	93.07	92.70	92.34	92.70	92.70	93.07	92.70	93.80	92.70	92.70	92.70	93.07	93.07	92.70	92.34	92.70	93.07	93.07	93.07	92.34	92.70	93.07	93.07	92.70	93.07	92.70	92.34	92.70	92.70	93.43	93.43

The time spent for developing and optimising the proposed MTMCS based on one, two and three-layer NNs using the BRATS2015 dataset (Table 5.21) is measured and analysed using a personal computer with Core i7, RAM 16 M, having enabled the parallel computing, and Matlab software Ver. R2018. The time that was measured, indicated that additional layer of NNs leads to raise the implementation time further. It is also noted that time for developing the proposed system for this dataset is higher compared to the other datasets where this dataset has larger number of samples and this leads to increase the time required for the training phase, which consequently increases the implementation time of the proposed system. Although that time is significantly reduced by enabling parallel processing, it still takes considerable time for the implementation of the proposed MTMCS. However, this time is only required for the development of the system, and once the system model is developed and optimised, the required time for decision-making process or testing a sample will not take more than a few seconds. The time that is calculated for training and testing includes the training of N-1 where N is the total number of samples and time for testing one sample, this process is repeated for all samples through the LOO cross-validation technique and therefore the average ± standard deviation is measured (Mean ± SD) (Table 5.21). Number of classifiers being 55 refers to experiment 4 when diversity is activated, while the number of classifiers 11 indicates all other experiments.

Table 5.21 The development time spent by the proposed MTMCS including the total time, practical time, training N-1 and testing one sample. N is the total number of samples where the time is calculated according different number of layers and classifiers. This time is measured in seconds.

No. of Layer of NNs	No. of classifiers	Total Time	Practical time using parallel processing	Training (N-1) and testing one sample Mean ± SD
One-layer	11	61925.31	9243.96	0.150 ± 0.02
Two-layer	11	2338032	328009.97	0.189 ± 0.02
One-layer	55	65375.52	9387.87	0.159 ± 0.03
Two-layer	55	2409602	343778.08	0.195 ± 0.02
Three-layer	55	24695916	3522955.17	0.211 ± 0.03

5.9 Evaluating the Proposed MTMCS Using the BRATS2018 Dataset

To add further evaluation for the proposed MTMCS, BRATS2018 is used to train and test the automated classification system. Three experiments are implemented using this dataset; the first one is implemented based on the use of the texture feature associated with GLCM (Experiment 1). The second experiment is implemented based on the full set of features including the textural features and the new features developed to measure the tumour descriptor (Experiment 2). The third one (Experiment 3) is implemented based on enabling the diversity in the feature space using the same methodology develop with BRATS2015 which is explained in detail in section 5.8.5.

5.9.1 Experiment 1: Implementing the Proposed MTMCS Using the Textural Features

In this experiment, the input features are the textural statistics based on the full set of features associated with GLCM to classify malignant grades of glioma using the BRATS2018 dataset. In the implementation of the first stage of the MTMCS, the ODM is constructed based on the training and testing each classifier individually. The resultant ODM will have the dimensions - eleven classifiers $\times$ 285 samples.

Considering the maximum results that are achieved using the proposed MTMCS (Table 5.22), there are 55 samples out of 75 of low-grades samples are correctly classified as low-grade glioma, while 203 out of 210 of high-grades samples are correctly identified as high-grade glioma. The results indicate that the classification performance for the proposed system in terms of classification accuracy, sensitivity, specificity, precision, F-measure are 90.53%, 96.67%, 73.33%, 91% 93.76% respectively.

Table 5.22 Evaluation results of the proposed system for the discrimination between low glioma grades (I, and II) and the high grades (III, and IV) using the BRATS2018 dataset. Where Class1 and Class0 refer to high, and low grade respectively.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
Proposed MTMCS	Class0	55	20	73.33	88.71	80.29	90.53
	Class1	7	203	96.67	91.03	93.76	

It is noticed that these results were obtained when implementing the proposed MTMCS based on one-layer NNs using 10 neurons in the layer with the 22nd iteration (Figure 5.38 and Figure 5.39). There are also many other results which show slightly lower classification accuracy at 90.16% with the use of different numbers of neurons, for example when using 1, 8 and 14 neurons. However, the better choice would be 10 neurons as it shows better accuracy, and therefore the design based on 10 neurons was selected (Figure 5.38).
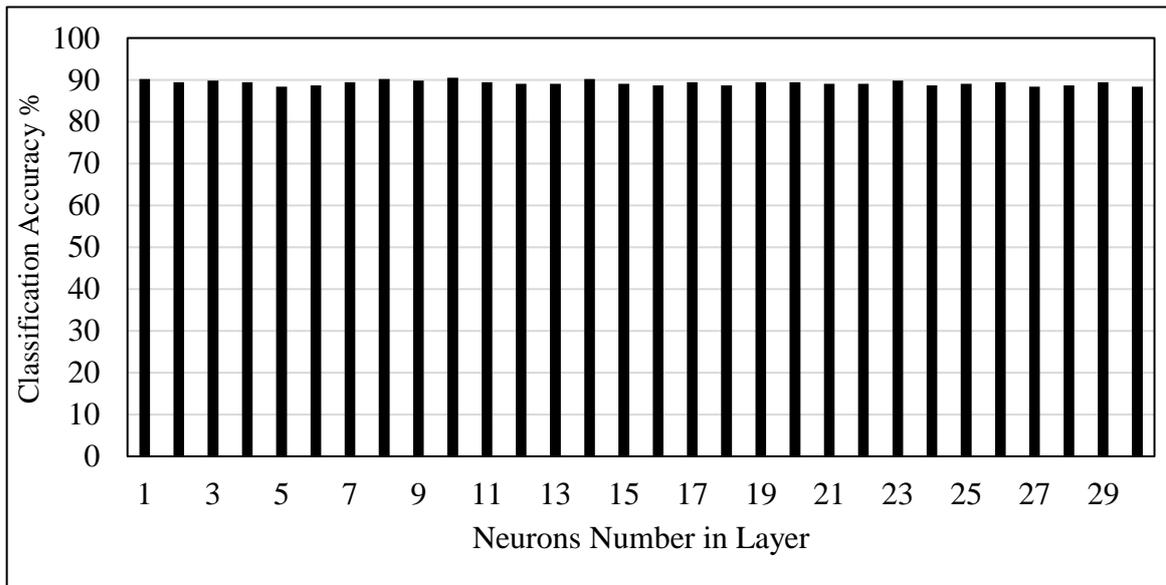
Figure 5.38 Classification accuracy results for applying the proposed system based on one-layer NNs corresponding to the number of neurons in the layer to discriminate low grades (I, and II) against high glioma grades (III, and IV) using BRATS2018 dataset.
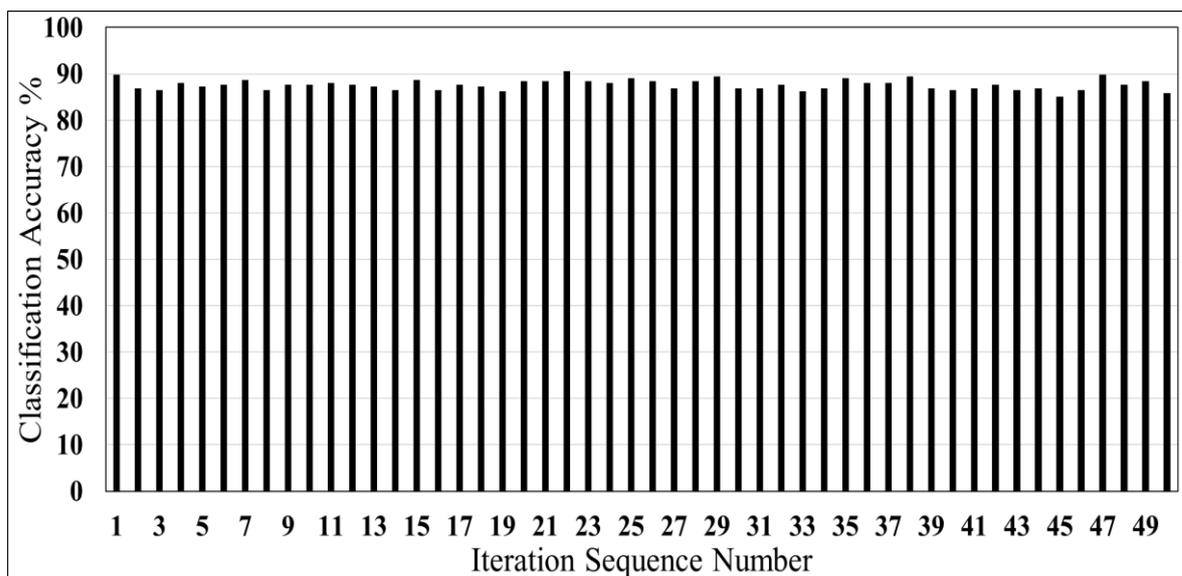
Figure 5.39 Classification accuracy results for applying the proposed system corresponding to the iteration sequence number using the ten neurons in the one-layer NNs to discriminate low grades (I, and II) against high glioma grades (III, and IV) using the BRATS2018 dataset.

5.9.1.1 Redundancy Analysis and Selection of Classifiers

The SC algorithm is implemented to optimise the classification performance by determining the best set of classifiers as well as eliminating redundant classifiers. The results obtained from applying this algorithm incorporated with the proposed MTMCS based on one-layer NNs showed that the best classification accuracy achieved at 91.58% by Run8 (Table 5.23). The highlighted classifiers are the optimal selected set of classifiers (Table 5.23), and the number of classifiers is reduced from 11 to 6 classifiers. Also, other runs showed a slightly lower classification accuracy at 91.23% with small number of classifiers such as Run6, and Run7. Following lower accuracy was achieved at 90.18% by Run10 where the number of classifiers is reduced from 11 to 5 classifiers. The results illustrate that the best classification performance for the proposed system based on one-layer NNs incorporated with the SC algorithm in terms of sensitivity, specificity, precision, F-measure are 97.62%, 74.67%, 91.52%, 94.47% respectively (Table 5.24).

Table 5.23 Selection process of best set of classifiers with the corresponding classification accuracy *ACCnew* for each selected set of classifiers. Where 1 and 0 refer to the keep and removing actions of classifier respectively, Run represents running the process for each selected case. This test is classifying low grades (II, III) versus high grades (III, IV) using the BRATS2018 dataset based on the full set of features (3DGLCM).

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	**Run8**	Run9	Run10	Run11
SVML	1	1	1	1	1	1	1	1	1	1	1	0
KNNCUB	1	1	1	1	1	1	1	1	1	1	0	1
KNNMED	1	1	1	1	1	1	1	1	1	0	1	1
KNNW	1	1	1	1	1	1	1	1	0	0	0	0
SVMQ	1	1	1	1	1	1	1	0	0	0	0	0
SVMG	1	1	1	1	1	1	0	0	0	0	0	0
SVMCUB	1	1	1	1	1	0	1	1	1	1	1	1
KNNCOS	1	1	1	1	0	1	1	1	1	1	1	1
DT	1	1	1	0	0	0	0	0	0	0	0	0
LDA	1	1	0	1	1	1	1	1	1	1	1	1
KNNF	1	0	0	0	0	0	0	0	0	0	0	0
ACCnew %	90.53	90.53	90.18	90.88	89.47	89.44	91.23	91.23	91.58	89.82	90.18	89.47

Table 5.24 Evaluation results of the proposed system based on one-layer NNs incorporated with the SC algorithm for the discrimination between low glioma grades (I, and II) and the high grades (III, and IV) using the BRATS2018 dataset where Class1 and Class0 refer to high, and low grade respectively.

| Classifier | Actual class | Confusion matrices Predicted class | | Sensitivity % | Precision % | F-measure % | Accuracy % |
		Class0	Class1				
Proposed MTMCS	Class0	56	19	74.67	91.80	82.35	91.58
	Class1	5	205	97.62	91.52	94.47	

5.9.1.2 Comparison with Other Methods

The overall classification performance in terms of classification accuracy is conducted (Figure 5.40). The results showed that the proposed MTMCS based on one-layer NNs has achieved an accuracy of 91.58% outperforming all other single and ensemble classification models. The next highest classification accuracy compared to all other the traditional approaches was achieved at 86.31% by the SVML classifier. While the lowest classification accuracy is obtained by KNNF classifier at 78.24%. With respect to the ensemble systems, the averaged classification accuracy is around 82%, where the majority voting was the one that shows the best accuracy at 84.91% compared to the other ensemble approaches (Figure 5.40).
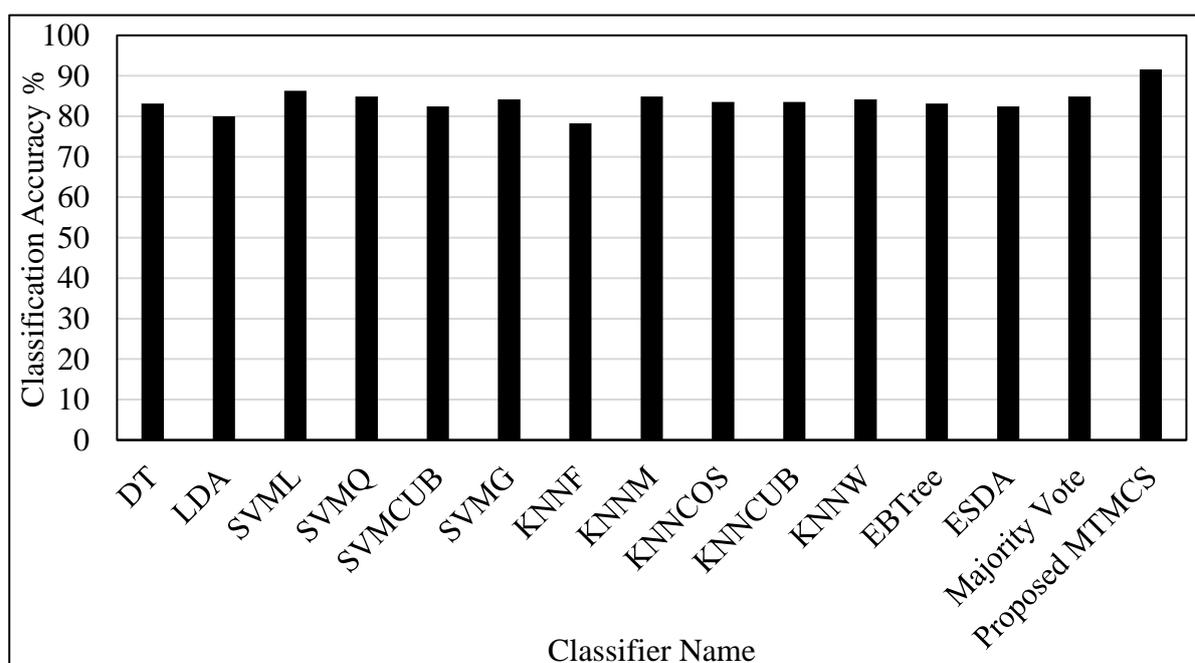
Figure 5.40 Comparative results in terms of the classification accuracy of the proposed MTMCS based on one-layer NNs against all other single and ensemble classification models using the BRATS2018.

5.9.2 Experiment 2: Implementing the Proposed System Using the Textual Feature and the Proposed Feature Associated with Tumour Descriptors

The automated classification system is further developed using all features extracted in this work including the full set of features associated with the GLCM and the new features FTD. The rationale of this experiment is to examine the ability of the classification system based on the proposed MTMCS using the textural features and the new features extracted from the tumour descriptors to classify the malignant grades of glioma.

After implementing the proposed MTMCS based on one-layer NNs, the results illustrate that the obtained classification performance in terms of classification accuracy, sensitivity, specificity, precision and F-measure are 95.09%, 98.10%, 86.67%, 95.37%, 96.71% respectively (Table 5.25). These best results are achieved by implementing the proposed MTMCS with the use of 13 neurons and 40th iteration (Figure 5.41 and Figure 5.42).

Table 5.25 Evaluation of the classification performance for the proposed system based on the one-layer NNs. This is for the discrimination between low grades glioma (I, and II) and the high grades (III, and IV) using the BRATS2018 dataset.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity%	Precision %	F-measure %	Accuracy%
		Class0	Class1				
Proposed MTMCS	Class0	65	10	86.67	94.20	90.27	95.09
	Class1	4	206	98.10	95.37	96.71	
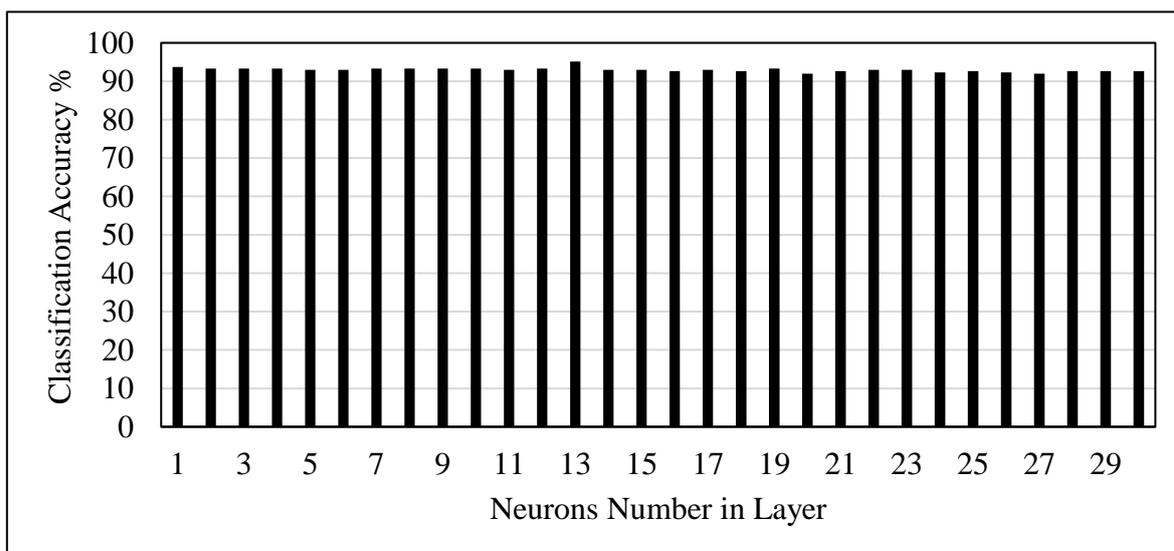
Figure 5.41 Classification accuracy results for demonstrating the proposed MTMCS based on one-layer NNs corresponding to the number of neurons in the layer based on the full set of features 3DGLCM and FTD derived from BRATS2018.
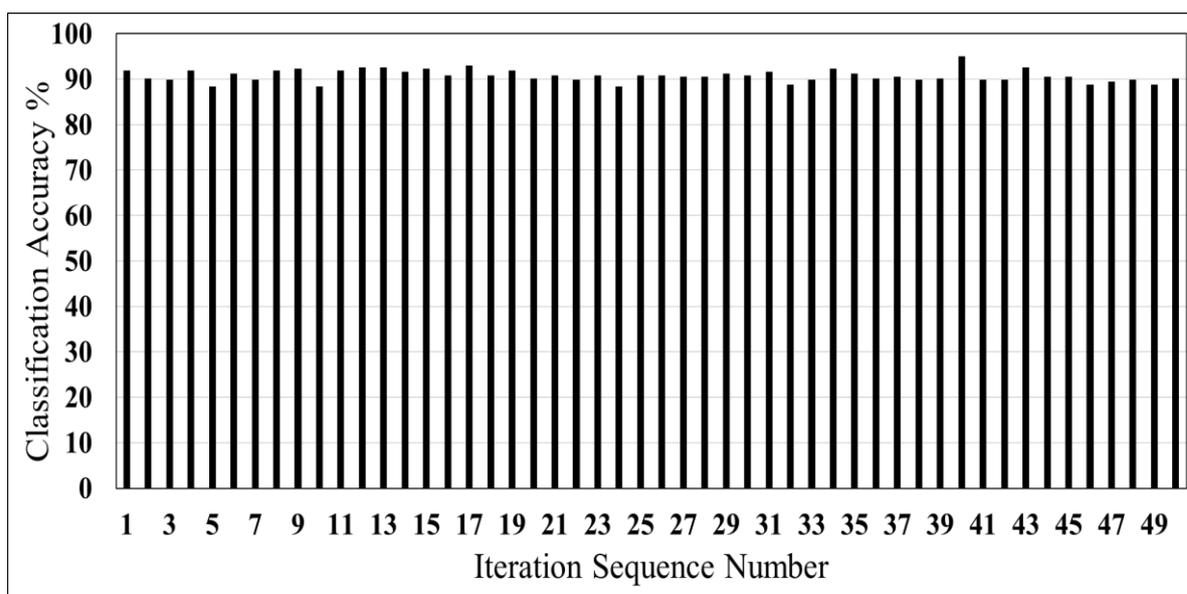
Figure 5.42 Classification accuracy results corresponding to the iteration sequence number with the thirteen neurons based on the full set of features 3DGLCM and FTD using BRATS2018.

5.9.2.1 Redundancy Analysis and Selection of Classifiers

After applying the SC algorithm to this experiment, the results indicated that the initial run achieved the maximum classification accuracy at 95.09% outperforming all other runs. Different runs have also been noticed to show slightly lower classification accuracy averaged around 94% with lower number of classifiers that is reduced from 11 to 10 classifiers such as Run1, Run2 and Run3. However, the best classification accuracy is achieved by the initial run based on the eleven classifiers (Table 5.26).

5.9.2.2 Comparison with Other Methods

It is observed from the comparison of the proposed MTMCS versus the other traditional single and ensemble classifiers that the proposed system based on the combination of the textural feature and the proposed features derived from the tumour descriptors has achieved a superior classification accuracy at 95.10 % compared to all other classification approaches (Figure 5.43). Among the single classifiers, 92.98% of classification accuracy is achieved by SVML. Following this result, the SVMQ classifiers obtained 91.22% (Figure 5.43). It is also noticed that the combination of the textural features with the new FTD integrated with the MTMCS enables the simple majority vote to achieve a significant classification accuracy at 90.17%. While all other classifiers result is between the lowest accuracy at 80.35% achieved by LDA to 89.82% achieved by both EBTree and SVMCUB classifiers (Figure 5.43).

Table 5.26 Selection process of best set of classifiers with the corresponding classification accuracy *ACCnew* for each selected set of classifiers. Where 1 and 0 refer to the keep and removing actions of classifier respectively, Run represents running the process for each selected case. This test is classifying low grades (II, III) versus high grades (III, IV) using the BRATS2018 dataset based on the full set of features (3DGLCM and FTD).

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run11
SVML	1	1	1	1	1	1	1	1	1	1	1	0
SVMQ	1	1	1	1	1	1	1	1	1	1	0	1
DT	1	1	1	1	1	1	1	1	1	0	1	1
SVMCUB	1	1	1	1	1	1	1	1	0	1	1	1
SVMG	1	1	1	1	1	1	1	0	1	1	1	1
KNNMED	1	1	1	1	1	1	0	1	1	1	1	1
KNNCOS	1	1	1	1	1	0	1	1	1	1	1	1
KNNW	1	1	1	1	0	1	1	1	1	1	1	1
KNNCUB	1	1	1	0	1	1	1	1	1	1	1	1
KNNF	1	1	0	1	1	1	1	1	1	1	1	1
LDA	1	0	1	1	1	1	1	1	1	1	1	1
ACCnew %	**95.09**	94.04	94.04	94.04	94.04	93.68	93.68	94.04	94.74	94.39	94.04	92.98
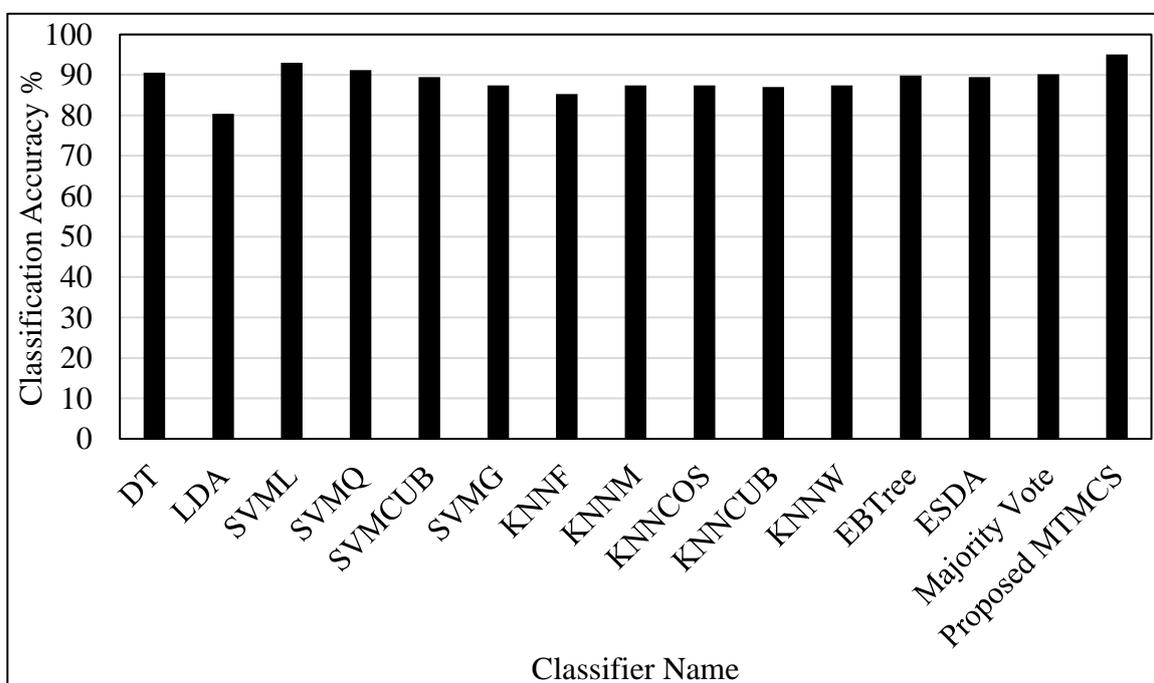
Figure 5.43 Comparative evaluation in term of classification accuracy for the proposed MTMCS based on one-layer NNs versus the other single and ensemble classifiers to classify low against high grade glioma using BRATS2018 based on the full set of features of 3DGLCM and the proposed tumour features.

5.9.3 Experiment 3: Develop the Proposed System Based on the Diversity in Feature Space

The full set of features including FTD and GLCM are used with enabling the diversity in feature space by splitting the features set into different subsets of features and through the same methodology developed in section 5.8.5. Then the classification models are trained and tested using these subsets of features whereby different output decisions are produced and utilised to build the ODM. The ODM is then employed in the second stage of MTMCS. The dimensions of the ODM are 11 classifiers $\times$ 285 samples $\times$ 5 subsets of features, resulting in 15675 tests that are implemented in the first stage of the MTMCS. For the implementation of the second stage of the proposed MTMCS based on one-layer NNs, 285 tests $\times$ 30 neurons $\times$ 50 iterations are investigated, as a result, in total, 427500 tests are implemented and evaluated.

The results obtained from testing the proposed MTMCS based on one-layer NNs showed that the best classification performance achieved in terms of the classification accuracy, sensitivity, specificity, precision, and F-measure are 93.33%, 97.62%, 81.33%, 93.6%, and 95.57% respectively (Table 5.27).

It is observed that the best classification accuracy of 93.33% is obtained by the implementation of MTMCS based on one-layer NN through the use of six neurons, and based on 32nd iteration (Figure 5.44 and Figure 5.45). The same classification accuracy at 93.33% is achieved by using 8 neurons. It is also noted that there are various numbers of neurons and iterations reflecting slightly comparable classification accuracy at 92.93% by the use of 9 and 15 neurons. While other cases achieved lower classification accuracy at 92.63% such as 1, 2, 3 and 4 neurons (Figure 5.44).

Table 5.27 The results of the performance evaluation of the proposed MTMCS based on one-layer NNs incorporated the diversity in features space using the BRATS2018 dataset. This test is to classify between low grades glioma (I, II) and high grades glioma (III, IV). Class0 and Class1 refer to low and high grades respectively.

| Classifier | Actual class | Confusion matrices Predicted class | | Sensitivity% | Precision % | F-measure % | Accuracy% |
		Class0	Class1				
Proposed MTMCS	Class0	61	14	81.33	92.40	86.52	93.33
	Class1	5	205	97.62	93.60	95.57	
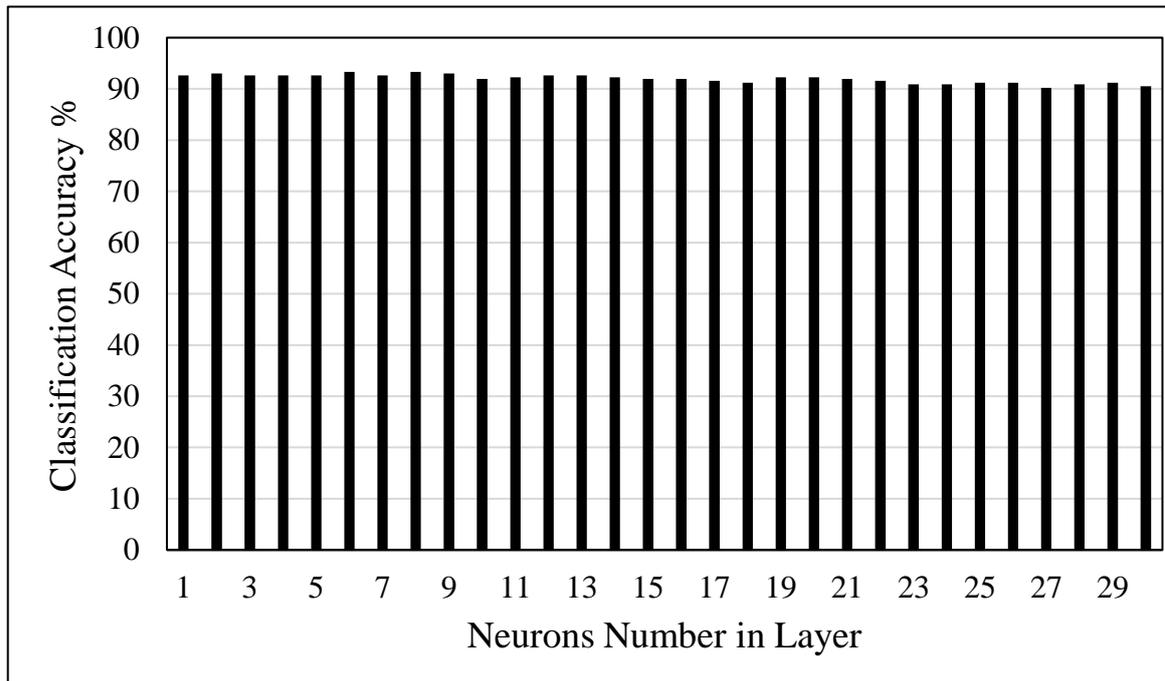
Figure 5.44 Classification accuracy results for the proposed system based on one-layer of NNs corresponding to the number of neurons in layer based on enabling the diversity in features space. This test is to classify between glioma low grades (I, II) and high grades (III, IV) using BRATS2018 dataset.
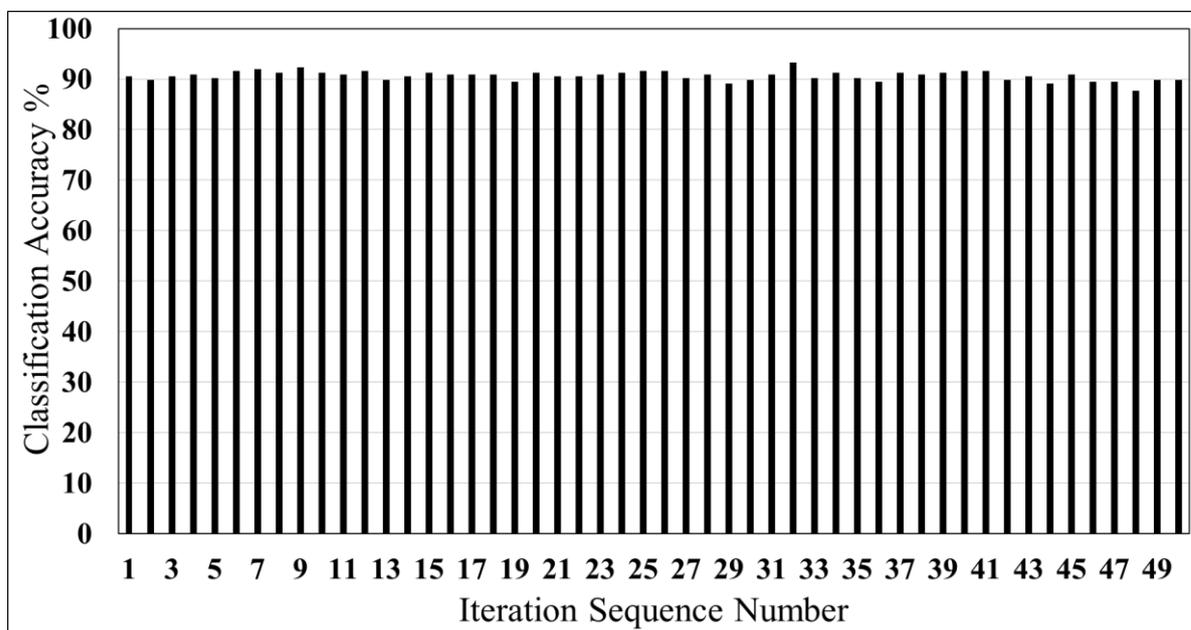
Figure 5.45 Classification accuracy results for the proposed system based on one-layer NNs according to the iteration sequence number for six neurons in the layer, incorporating the diversity in feature space. This test is to classify between glioma low grades (I, II) and high grades (III, IV) using the BRATS2018 dataset.

After implementing the proposed MTMCS based on two-layer NNs, the results indicate that the classification accuracy is improved to achieve its highest classification performance in terms of the classification accuracy, sensitivity, specificity, precision, and F-measure are 94.39%, 98.10%, 84%, 94.50%, and 96.26% respectively (Table 5.28). This achievement in the results is obtained based on the use of 12, and 6 neurons in the first and second layers respectively at the 7th iteration (Figure 5.46)

Table 5.28 The performance evaluation of the proposed MTMCS based on two-layer NNs and incorporated the diversity in features space using the BRATS2018 dataset. This test to classify between low grades glioma (I, II) and high grades glioma (III, IV).

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
Proposed MTMCS	Class0	63	12	84.00	94.03	88.73	94.39
	Class1	4	206	98.10	94.50	96.26	
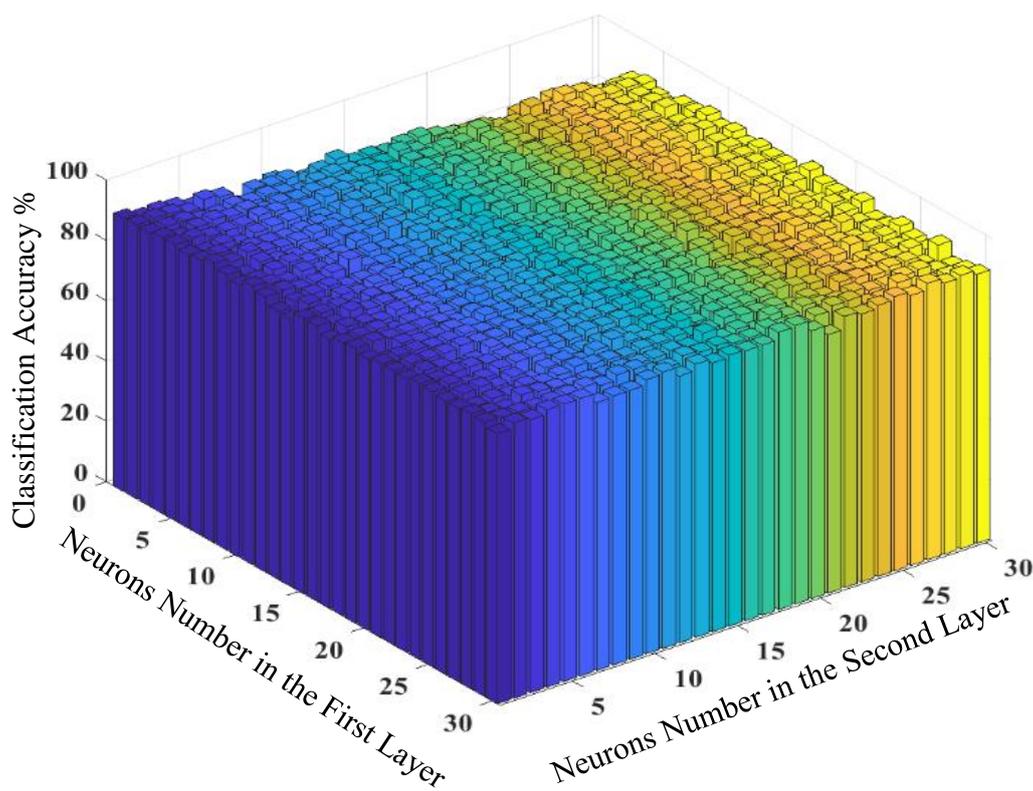
Figure 5.46 Classification accuracy results for the proposed MTMCS using two-layer NNs corresponding to number of nodes in the first and second layers in the left and right axis respectively. This test is to classify between lower grade (II, III) and high grades (III, IV) using the BRATS2018. The optimal selected design was found at the 7th iteration and 12 neurons in the first layer, and 6 at the second layer.

5.9.3.1   Redundancy Analysis and Selection of Classifier

The same algorithm of selecting the best set of classifiers (SCA) incorporated with the proposed MTCMS with enabling the diversity in features space is applied to BRATS2018 dataset. The rationale of this experiment is to add further evaluation to the developed system to investigate a better enhancement in the classification performance of the automated system in terms of classification accuracy and model design. The results indicated that the classification accuracy was enhanced from 93.33% achieved by the initial Run to 94.74% obtained by Run5 (Table 5.30). While the dimensions of classifiers are reduced from 55 classifiers to 52 classifiers that are highlighted in yellow and three classifiers are eliminated, these are LDA_3rd and DT_3rd and KNNF_2D classifiers (Table 5.30). It is observed that there are other cases have achieved slightly lower classification accuracy at 94.04% by RUN13, RUN20, RUN32 and RUN39 with better elimination of classifiers where the number of classifiers is reduced from 55 classifiers to 51 classifiers. It is also noted that most of the runs conducted by the proposed MTMCS based on one-layer NNs incorporating the SC method obtained slightly comparable accuracies around 93.33 to 94% (Table 5.30). Classification performance in term of sensitivity, specificity, precision and F-measure is illustrated in Table 5.29, which was achieved by Run5. It is noted that the maximum classification accuracy of 94.74% is achieved by the implementation of MTMCS based on one-layer NN through the use of twelve neurons and through using the 35th iteration (Figure 5.47 and Figure 5.48).

Table 5.29 Results of the performance evaluation of the proposed MTMCS based on one-layer NNs and enabling the diversity in features space and SCA using the BRATS2018 dataset. This test is to classify between low grades glioma (I, II) and high grades glioma (III, IV). Class0 and Class1 refer to low and high grades respectively.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
Proposed MTMCS	Class0	62	13	82.67	96.88	89.20	94.74
	Class1	2	208	99.05	94.12	96.51	

Table 5.30 Selection process of best set of classifiers with the corresponding classification accuracy *ACCnew* for each selected set of classifiers. This test is classifying low grades (II, III) versus high grades (III, IV) using BRATS2018 dataset based on the full set of features (3DGLCM and FTD) and diversity in the features space.

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run11	Run12	Run13	Run14	Run15	Run16	Run17	Run18	Run19	Run20	Run21	Run22	Run23	Run24	Run25	Run26	Run27	Run28	Run29	Run30	Run31	Run32	Run33	Run34	Run35	Run36	Run37	Run38	Run39	Run40	Run41	Run42	Run43	Run44	Run45	Run46	Run47	Run48	Run49	Run50	Run51	Run52	Run53	Run54	Run55
SVML_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
SVML_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
SVMQA_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
KNNCUB_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
SVMQA_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
KNNM_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
DT_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
SVMG_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
DT_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
KNNCOS_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
SVMCUB_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
LDA_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
SVMCUB_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNW_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMG_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNM_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCOS_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNW_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVML_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCUB_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNF_TD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVML_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVML_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNF_ALL3DTD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMQA_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNM_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCUB_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMG_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNW_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNM_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNW_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMG_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMQA_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNW_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCUB_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCOS_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMCUB_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMQA_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCOS_3rd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DT_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCOS_2D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMCUB_ALL3D	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMG_3rd	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SVMCUB_3rd	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNCUB_3rd	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNM_3rd	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DT_2D	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LDA_2D	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LDA_ALL3DTD	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LDA_ALL3D	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LDA_3rd	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KNNF_ALL3D	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNF_3rd	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KNNF_2D	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DT_3rd	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ACCnew	93.33	93.68	93.68	93.33	93.33	94.74	93.68	93.33	93.33	93.68	93.33	93.33	93.68	94.04	93.33	93.33	93.33	93.68	93.68	93.68	94.04	93.33	93.33	93.68	92.98	92.98	93.68	93.33	92.98	94.39	93.33	94.04	93.68	93.68	93.68	93.68	92.98	93.33	94.04	93.68	93.68	93.68	93.68	93.33	93.33	93.68	93.33	92.98	93.68	93.68	92.98	93.33	93.68	92.98	92.63

Figure 5.47 Classification accuracy results for the proposed system based on one-layer of NNs corresponding to the number of neurons in layer based on enabling the diversity in features space. This test is to classify between glioma low grades (I, II) and high grades (III, IV) using BRATS2018 dataset for the RUN5 of SC algorithm.

Figure 5.48 Classification accuracy results for the proposed system based on one-layer NNs according to the iteration sequence number for twelve neurons in the layer, incorporating the diversity in feature space. This test is to classify between glioma low grades (I, II) and high grades (III, IV) using the BRATS2018 dataset for the RUN5 of the SC algorithm.

5.10 Results Summary and Discussion

Overall results, from the implementation of the proposed MTMCS using the four datasets, are summarised in Table 5.31 and Table 5.32 where the results and findings of this chapter are discussed. It has been shown that the proposed methodology based on the integration of multiple classifiers and DNN offers an improved classification accuracy for brain glioma grades. A wide range of experiments and designs are investigated for the proposed MTMCS scheme using four benchmark datasets. The classification performance of the proposed system is evaluated and compared against different popular single classifiers and ensemble systems.

In general, the results illustrated that the proposed MTMCS has achieved optimal classification results at 100% accuracy when tested with two datasets; these are the BRATS2013 and Cancer dataset, while it yields a superior classification accuracy of 93.80% and 95.09% when tested with the BRATS2015 and BRATS2018 respectively.

Furthermore, the proposed MTMCS based on the textural features measured from GLCM has achieved a full accuracy rate at 100% using the BRATS2013 dataset, and the same full accuracy is achieved with the Cancer dataset when the SC algorithm is utilised. The proposed design obtains a classification accuracy of 91.24% with the BRATS2015 dataset when the SC method is used. While the highest achievement was 95.09% of classification accuracy using BRATS2018 dataset.

The results obtained from applying the proposed MTMCS based on integrating the full set of textural features measured for GLCM and the proposed features extracted from tumour descriptors indicate that the classification accuracy is improved to 92.7% using BRATS2015 and to 95.09% using BRATS2018 dataset. For BRATS2015, the classification accuracy is further improved to 93.07% by incorporating the SC method. In attempting to improve the classification accuracy further, the proposed MTMCS is updated by activating the diversity in the features space through dividing the feature space into five subsets of features and utilising the same eleven classifiers in the training and testing of each subset of features. It is observed that the results of the implementation of one-layer NNs have not shown a significant improvement where the classification accuracy remained at the same level at 93.07%. However, there is a slight improvement achieved in the specificity that is improved from 68.52% to 70.37%. On the other hand, when further hidden layers of NNs are investigated.

Table 5.31 Results achieved by implementing the proposed MTMCS using the BRATS2013, BRATS2015 and BRATS2018 datasets. Sensitivity refers to the high grades. Highlighted cell with (*) means there is no rationale to progress the experiment further.

N	Method design	BRATS2013					BRATS2015					BRATS2018				
		ACC %	Sensitivity %	Specificity %	Precision %	No. of Classifiers	ACC %	Sensitivity %	Specificity %	Precision %	No. of classifiers	ACC %	Sensitivity %	Specificity %	Precision %	No. of classifiers
1	Texture-GLCM	100	100	100	100	11	**90.51**	96.36	66.67	92.17	11	**90.53**	96.67	73.33	91.03	11
2	Texture-GLCM+SC (one layer)	100	100	100	100	5	**91.24**	97.27	66.67	92.20	7	**91.58**	97.62	74.67	91.52	6
3	Texture-GLCM + FTD (one-layer)	*	*	*	*	*	**92.70**	98.10	70.30	93.10	11	**95.09**	98.10	86.67	95.37	11
4	Texture-GLCM + FTD +SCA (one layer)	*	*	*	*	*	**93.07**	99.09	68.52	92.80	6	**95.09**	98.10	86.67	95.37	11
5	Enabling diversity in features space (Texture-GLCM+FTD) one-layer	*	*	*	*	*	**93.07**	98.64	70.37	93.10	55	**93.33**	97.62	81.33	93.60	55
6	Enabling diversity in features space (Texture-GLCM+FTD) two layer	*	*	*	*	*	**93.43**	98.18	74.07	93.91	55	**94.39**	98.10	84	94.50	55
7	Enabling diversity in features space (Texture-GLCM+FTD) one-layer +SCA	*	*	*	*	*	**93.80**	98.6	74.0	93.9	52	**94.74**	99.05	82.67	94.12	52

Table 5.32 Results achieved by implementing the proposed MTMCS using the Cancer dataset. Positive in confusion matrix and sensitivity refer to the high grades.

Cancer dataset, low grade (GII) against high grades (GIII, GIV)					
Method design	Classification Accuracy	Sensitivity	Specificity	Precision	No. of classifiers
Without feature reduction	96.67	100	90	95.23	11
With Feature selection (one-layer)	96.67	100	90	95.23	11
After SCA (one-layer)	100	100	100	100	7

It is observed that the classification accuracy is enhanced to 93.4% by the two-layer NNs and more improvement is gained by applying the three-layer NNs where the classification accuracy is increased to 93.80%. For BRATS2018 dataset, after implementing the one-layer MTMCS and enabling the diversity in feature space, the results showed lower classification accuracy at 93.33% and this indicates that redundant classifiers are conflicting with others in the pool of 55 classifiers, which impact negatively on the classification accuracy. Therefore, the experiment is further extended to investigate the proposed MTMCS based on two-layer of NNs and the full set of features that includes texture and FTD incorporated with enabling the diversity and thus the result indicated better enhancement and it is improved from 93.33% to 94.39%. The integration between the SC algorithm and the developed MTMCS is also implemented and the classification accuracy is evaluated. Hence the results illustrated notable enhancement in the classification accuracy to achieve 94.74% while the number of classifiers needed to obtain this result is reduced from 55 to 52 classifiers. This is a significant improvement in the system performance in terms of classification accuracy and classifies dimensions.

It was observed that the development of the proposed system based on adding more hidden layers would lead to the enhanced classification accuracy of glioma grades. However, adding more than one hidden layer of NNs increase the complexity of the proposed system design. Nevertheless, using the proposed methodology of the MCS based on the DINN incorporated the selection of the best set of classifiers (SCA) enable the proposed system to achieve the optimal classification rate with fewer numbers of layers of NNs.

The proposed MTMCS based on three-layer NNs has achieved a significant improvement in the classification accuracy. However, it requires high computation time mainly in the optimisation of the proposed MTMCS as well as having further complexity compared to other layers of NNs. Therefore, in attempting to tackle this issue and reduce the complexity in the proposed design, the SC algorithm is applied for further investigation, seeking for the best classification performance through conducting the proposed MTMCS based on only one-layer NNs.

The SC algorithm is advantageous as it enables the proposed MTMCS to achieve its maximum classification performance without the need to investigate further hidden layers of NNs where the same higher classification accuracy at 93.80% is obtained by incorporating SC algorithm with the proposed methodology based on only one-layer NNs. This achievement is gained with a lower number of classifier members that is only 52 classifiers at the second stage of the proposed MTMCS.

For the Cancer dataset, the full discrimination rate is obtained between grade IV and the lower grades (II, III), and between grade II versus the higher grades glioma (III, IV). Similarly, for BRATS2013, the full discrimination rate is obtained between low grades (I, II) and high grades glioma (III, IV), where all these achievements are conducted by the proposed MTMCS based on one-layer NNs and incorporating the SC algorithm. The proposed DINN is proven to overcome the problem of random selection of neurons and iterations in an attempt to achieve the best design of DNN. This is performed by a systematic examination for different intensive designs of NNs based on various numbers of neurons and iterations developed and implemented for the proposed system. To further elaborate, there are different points of divergence that come from testing different validation sets and initial weights, examined within various iterations of DNN.

Consequently, different network weights are produced in each iteration, where each iteration provides different results. The purpose of determining a considerable number of different designs of DNN with various subsets of validation is to find the optimal convergence to the global minimum value, which represents the best possible design of DNN that would reveal the highest classification accuracy. However, this will require a longer computation time in the learning phase.

Some datasets used in this experimental work could be considered relatively small, and this may have degraded the accuracy of the NNs. However, using the proposed MTMCS has

been proven to overcome this limitation through the integration of different classifiers that can handle datasets of small size and thus show remarkable results. Indeed, at present, it is a significant challenge to acquire a large image dataset of glioma grades with the approval of the histopathology test. This approval is vital to have a solid confirmation of the malignancy grade of glioma, thereby providing real validity of the evaluation and validation process of any classification or grading system. Consequently, the proposed classification system is not affected by this issue, and thus it achieved a promising result and a significant improvement in the classification accuracy of glioma grades. Besides, the proposed system can be considered critically as an alternative method to the traditional approach of deep learning that usually requires an enormous dataset and an intensive computation time.

A significant improvement is obtained in the results of the classification of MR images of glioma grades using the proposed MTMCS. The results show that the proposed MTMCS has achieved significant results compared to other single and ensemble classification method without the need to involve any selection process in the first stage of the proposed system. However, conducting the selection task in the second stage was observed to highly impact on the classification performance in terms of classification accuracy and gaining a reduction in the dimensions of the classifier member. This significantly improves the classification performance through the selection of best set of classifiers (SCA) that are able to contribute effectively to the classification system of glioma grades. It was found that the proposed system has the advantage of outperforming both the single classification method and the ensemble systems. However, it requires a high computation time, advanced computing hardware, and parallel computing making it possible to reduce the execution time.

5.11 Conclusion

Two stages of a novel multiple classification systems for automated glioma grading have been proposed in this chapter. The first stage was designed based on utilising different popular classification models developed for the grading of malignancy of brain tumours in MR images. In this stage, eleven classification models were trained and tested individually. Each of the classification models was trained based on different features including the co-occurrence of textural features and the proposed features associated with tumour descriptors, where these features were extracted from the conventional MRI modalities. The second stage adopted the meta-strategy of multiple classifier systems, which is developed based on the integration of different classifiers based on DNN. A systematic application of DNN is

conducted through the proposed DINN that enable DNN to achieve its highest classification accuracy. Consequently, this leads to an improved classification performance of the proposed system in the classification of glioma grades.

To assess the ability of the proposed MTMCS in the discrimination of different WHO glioma grades and achieve improved classification accuracy, the proposed MTMCS was implemented in different experiments with the use of four benchmark datasets.

The performance of different classification approaches has been evaluated in terms of classification accuracy, sensitivity, specificity, precision and F-measure. A comparison between the proposed MTMCS and different methods such as single classifier and ensemble systems have been examined in terms of these performance measures. The Leave-one-out cross-validation technique has been adopted in all stages to add more generalisation for the classification system and to validate the system performance. The results have illustrated that the proposed MTMCS has achieved a superior classification accuracy at 93.80% for BRATS2015 and 95.09% for BRATS2018 as compared to the existing methods for the classification between low-grade glioma (I, II) and high-grade glioma (III, IV). While, for the other datasets, the experimental results show that the proposed system has obtained the optimal discrimination rate at 100%. To elaborate, for the BRATS2013 dataset, the full classification rate is achieved to classify low grades (I, II) from high grades glioma (III, IV), and for the Cancer dataset, the full recognition rate is obtained to classify low-grade II versus the high grades (III, IV), as well as between grade IV and the lower grades glioma (II, III).

The experimental outputs have revealed that the integration of different single classifiers using deep neural networks is an efficient approach to stack a multi-classification model for the classification of different glioma grades. The proposed methodology presented by the DINN has further improved the final classification accuracy, thus enabling the MCS to achieve better classification accuracy for glioma grades.

CHAPTER 6 : Multi-Class Classification for Glioma Grades

Overview

This chapter presents the proposed system, its evaluation and results analysis to achieve a multi-class classification of WHO malignant grades of glioma. This starts by demonstrating the proposed hybrid feature selection algorithm (HFSA) incorporating different single and ensemble systems. This chapter also investigates the ability of the automated classification system incorporated with the proposed HFSA to improve discrimination accuracy for multi-class classification of glioma grades. Then, in attempting to achieve better classification accuracy, the proposed meta-trainable multiple classification systems (MTMCS) integrated with a hierarchical strategy is demonstrated to achieve the multi-class classification of glioma grades. To further elaborate, the proposed system takes the merit of the hierarchical classification scheme and developing the classification system by establishing the proposed MTMCS in each node of the hierarchical design. This system is built to enhance the automated multi-class classification for glioma grades towards precise and improved classification results. In this chapter, the need to develop a multiple classifier systems for multi-class classification for different WHO grades of glioma was addressed. MR images of public dataset with different WHO glioma grades were used to evaluate the proposed system. The input features to every single classifier in the first stage of MTMCS were the texture information extracted from T2-MR slices. In the second stage, the output decision vectors produced from each classifier were fed to the DNN. Leave-one-out (LOO) cross-validation scheme was used in all stages to validate the classification process and to add more generalisation to the proposed classification system. The classification performance of the proposed system was compared against other current and common approaches such as the single classification system, MCS based on the majority vote and other widely used ensemble algorithms that include Ensemble Bagged Tree (EBTree) and Ensemble Subspace Discriminate Analysis (ESDA). The comparison in classification performance between the proposed system and the other existing methods was conducted to evaluate the proposed system in the classification of WHO glioma grades. This evaluation is measured in terms of the confusion matrix and classification accuracy.

6.1 Introduction

Recognising the correct glioma grade is a significant challenge because the different grade of glioma can have high heterogeneity and mixed pathological characteristics of tumour. Hence, in this work, the proposed MTMCS based on different machine learning algorithms incorporated with MRI features of a brain tumour are employed to determine the correct glioma grade (Theeler and Groves, 2011, Siker et al., 2006). The development of a machine learning algorithm to undertake multi-class classification is challenging because many machine learning methods are designed basically to handle a two-class problem. Therefore, most machine learning methods perform well if applied to a two-class classification problem compared with its performance if applied to multi-class classification. The hierarchical scheme is an efficient method and recommended widely by the literature to overcome the challenge of multi-class classification problem (Rajan and Ghosh, 2004, Chen et al., 2004). Therefore, this approach was adopted in this study to deal with the multi-class classification problem. However, it was shown in the literature, that existing works that aim to achieve multi-class classification have developed the hierarchical strategy based only on a single classification approach while giving less attention to take the merit of MCS (Kumar et al., 2002). Due to the existing method including the single classifier system has less ability to handle multi-class classification, in addition, it is highly dependent on the input data distribution which may lead to low classification accuracy. Consequently, to achieve an improved classification accuracy of glioma grades, a new scheme was proposed through the integration between the proposed MTMCS and hierarchical scheme. This new system is performed by the development of the proposed MTMCS in each node of the tree structure of the hierarchical scheme to build an effective classification system for WHO glioma grades. LOO cross-validation technique is used in all stages of both approaches to achieve valid results and add more generalisation to the classification system. This work was proposed to achieve an automated classification of WHO glioma grades with a more accurate and precise outcome than other current and common approaches such as the single classification system and traditional ensemble approach. Therefore, the proposed system is evaluated by comparing the grading accuracy of glioma with other recent common approaches including single classifiers and ensemble classification models.

6.2 Materials and MRI Input Features

The dataset used in this chapter is acquired from the cancer archive collection. Thirty patients of MRI T2- weighted are used. This dataset is used in this study because it has different grades validated with the confirmation of histopathology diagnosis. Each group of glioma grades is diagnosed with one of the grades (II, III, and IV), and each one has 10 patients. Each patient has a different number of images ranging from 20 to 120 images, with varying post imaging setting such as different gap space and slice thickness which ranges from 2 to 7.5 mm (Clark et al., 2013). The MRI-T2 slices were utilised as described in Chapter 3, including cropping and intensities normalisation. Then, the texture information based on eighteen statistics were extracted from the 3DGLCM. The work started with the preparation of a multi-class classification of glioma grades, in which each grade of glioma was assigned with a unique index label, indicating different classes of glioma, and each patient has been indexed with one of the three grades in the training phase.

6.3 Evaluation of the Proposed Hybrid Feature Selection Method for Multi-Class Classification of Glioma Grades

Three experiments were conducted; in the first experiment, the full set of the textural features associated with the 3DGLCM was utilised as an input into the single classifier system. The next experiment includes the implementation of the ANOVA technique whereby the selected set of features was used as an input into the single classifier system. In the last experiment, the proposed HFSA was implemented in which the selected set of features was utilised in the training and testing of the single classifier system. The purpose of these experiments is to evaluate the proposed HFSA in the multi-class classification of glioma grades and to investigate possible enhancement in the classification accuracy of the developed system.

At the classification stage, the same thirteen classifiers which were used in this work, were applied and evaluated, eleven of which were based on the single classifier system namely decision tree (DT), linear discriminate analysis (LDA), support vector machine (SVM) with four kernels (SVML, SVMQ, SVMCUB and SVMG), and K-nearest neighbour (KNN) with five different designs (KNNF, KNNM, KNNCOS, KNNCUB and KNNW). While two classifiers were based on an ensemble system namely EBTree and ESDA. These classifiers are the most widely used in the classification of different applications and can efficiently achieve the multi-class classification task (section 2.14) and (section 3.6).

The input features to the classification system are the texture features driven from MRI-T2 weighted, which were utilised for training the eleven individual classifiers, and then the testing phase was determined using a LOO cross-validation technique, in which each sample in the dataset set was tested one by one and then the classification performance was evaluated. The evaluation of the classification performance was conducted using a confusion matrix that contains the classification results of the three grades of glioma.

In the first experiment, the full set of the 234 features were utilised in the training and testing of the thirteen classifiers that include single and ensemble systems. The results indicate that different classifiers reflect various classification accuracies. The best classification accuracy is obtained by DT classifier at 70%, outperforming all other classifiers. The results obtained by a single classification model shows that the next best classification accuracy is achieved by KNNF classifier at 50%. This is followed by accuracies of 40% achieved by both SVMCUB and KNNM classifiers. The results obtained based on the ensemble system illustrate that EBTree classifier achieved better results at 40% compared to the ESDA classification method (Table 6.1).

Table 6.1 Comparative results for applying different classification methods using the full set of texture features. Where Class0, Class1, and Class2 indicate to Grade II, Grade III, and Grade IV respectively.

Classifier	Actual class	Confusion matrices			Accuracy %
		Predicted class			
		Class0	Class1	Class2	
DT	Class0	7	2	1	**70.00**
	Class1	2	7	1	
	Class2	2	1	7	
LDA	Class0	2	8	0	30.00
	Class1	6	1	3	
	Class2	1	3	6	
SVML	Class0	5	5	0	40.00
	Class1	6	0	4	
	Class2	0	3	7	
SVMQ	Class0	3	7	0	36.67
	Class1	5	1	4	
	Class2	1	2	7	
SVMCUB	Class0	4	4	2	40.00
	Class1	5	1	4	
	Class2	1	2	7	

Table 6.1 Continued

Classifier	Actual class	Confusion matrices			Accuracy %
		Predicted class			
		Class0	Class1	Class2	
SVMG	Class0	2	8	0	30.00
	Class1	6	0	4	
	Class2	1	2	7	
KNNF	Class0	4	5	1	50.00
	Class1	5	3	2	
	Class2	0	2	8	
KNNM	Class0	4	6	0	40.00
	Class1	7	0	3	
	Class2	0	2	8	
KNNCOS	Class0	5	4	1	46.67
	Class1	6	0	4	
	Class2	0	1	9	
KNNCUB	Class0	5	5	0	40.00
	Class1	6	0	4	
	Class2	0	3	7	
KNNW	Class0	1	9	0	30.00
	Class1	6	0	4	
	Class2	0	2	8	
EBTree	Class0	4	6	0	40.00
	Class1	7	1	2	
	Class2	0	3	7	
ESDA	Class0	2	5	3	33.33
	Class1	5	2	3	
	Class2	2	2	6	

After applying the ANOVA technique, the feature dimensions are reduced from 234 features to 122 features. The selected set of features by this method are utilised in the training and testing of all classifiers used in this work; the results show that the best classification accuracy compared to all other classifiers is achieved by the DT classifier at 73.33%. Lower accuracies at 60% are obtained by the KNNM and KNNCUB classifiers, followed by the KNNCOS classifier at 56.67%. All other classifier reflects results of accuracies less than 50%. While with respect to the ensemble systems, the results indicate that both EBTree and ESDA achieved lower accuracies at 43.33% and 33.33% respectively (Table 6.2).

Table 6.2 Comparative results for testing the selected set of features chosen by the ANOVA method incorporated with different classification methods.

Classifier	Actual class	Confusion matrices			Accuracy%
		Predicted class			
		Class0	Class1	Class2	
DT	Class0	7	1	2	**73.33**
	Class1	1	8	1	
	Class2	2	1	7	
LDA	Class0	5	3	2	30.00
	Class1	5	1	4	
	Class2	1	6	3	
SVML	Class0	5	5	0	40.00
	Class1	6	0	4	
	Class2	1	2	7	
SVMQ	Class0	4	6	0	36.67
	Class1	6	0	4	
	Class2	1	2	7	
SVMCUB	Class0	4	5	1	33.33
	Class1	5	1	4	
	Class2	1	4	5	
SVMG	Class0	5	5	0	46.67
	Class1	6	0	4	
	Class2	0	1	9	
KNNF	Class0	4	6	0	36.67
	Class1	5	1	4	
	Class2	0	4	6	
KNNM	Class0	10	0	0	60.00
	Class1	7	0	3	
	Class2	1	1	8	
KNNCOS	Class0	7	3	0	56.67
	Class1	6	0	4	
	Class2	0	0	10	
KNNCUB	Class0	10	0	0	60.00
	Class1	6	0	4	
	Class2	1	1	8	
KNNW	Class0	4	6	0	43.33
	Class1	7	0	3	
	Class2	0	1	9	

Table 6.2 Continued

Classifier	Actual class	Confusion matrices			Accuracy%
		Predicted class			
		Class0	Class1	Class2	
EBTree	Class0	5	5	0	43.33
	Class1	6	1	3	
	Class2	0	3	7	
ESDA	Class0	3	5	2	33.33
	Class1	5	1	4	
	Class2	3	1	6	

The next experiment is the implementation of the proposed HFSA, whereby the selected set of features is fed into the classification system and used in the training and testing of each classifier individually. For the initialisation of the proposed HFSA, the feature (*Inverse difference normalized predictor*, *4545*) is selected as a reference predictor and after the implementation of the proposed algorithm, the chosen features namely (*Inverse difference normalized predictor*, *4545ᵒ*), (*homogeneity, 0ᵒ*), (*homogeneity, 90ᵒ*), (*Dissimilarity*, *45ᵒ*). The results yielded from the use of these features with different classification methods show that the maximum classification accuracy compared to all other classifiers is obtained by DT classifier at 76.67%. The next best result with slightly lower accuracy at 73.33% is achieved by the EBTree classifier. With respect to single classification models, these results are followed by the accuracy of 63.33% obtained by SVMQ classifier and 60% achieved by both the KNNCOS and KNNCUB classifiers. While all other classifiers reflect smaller accuracies for example both LDA and ESDA classifiers achieved an accuracy of 36.67% (Table 6.3).

Table 6.3 Comparative results for examining the selected set of features chosen by the proposed HFSA incorporated with different single and ensemble classification approaches.

Classifier	Actual class	Confusion matrices			Accuracy%
		Predicted class			
		Class0	Class1	Class2	
DT	Class0	8	1	1	**76.67**
	Class1	1	8	1	
	Class2	2	1	7	
LDA	Class0	4	5	1	36.67
	Class1	7	0	3	
	Class2	1	2	7	

Table 6.3 Continued

Classifier	Actual class	Confusion matrices			Accuracy %
		Predicted class			
		Class0	Class1	Class2	
SVML	Class0	4	5	1	40.00
	Class1	7	0	3	
	Class2	0	2	8	
SVMQ	Class0	7	2	1	63.33
	Class1	4	4	2	
	Class2	0	2	8	
SVMCUB	Class0	6	3	1	53.33
	Class1	4	3	3	
	Class2	0	3	7	
SVMG	Class0	5	4	1	46.67
	Class1	6	2	2	
	Class2	0	3	7	
KNNF	Class0	5	4	1	46.67
	Class1	3	3	4	
	Class2	0	4	6	
KNNM	Class0	8	1	1	56.67
	Class1	7	1	2	
	Class2	1	1	8	
KNNCOS	Class0	9	0	1	60.00
	Class1	6	0	4	
	Class2	1	0	9	
KNNCUB	Class0	8	1	1	60.00
	Class1	7	1	2	
	Class2	0	1	9	
KNNW	Class0	7	2	1	56.67
	Class1	5	2	3	
	Class2	0	2	8	
EBTree	Class0	8	1	1	73.33
	Class1	1	6	3	
	Class2	2	0	8	
ESDA	Class0	4	5	1	36.67
	Class1	8	0	2	
	Class2	0	3	7	

The overall comparison of classification performance for applying the proposed HFSA against the ANOVA technique and the set of features was conducted (Figure 6.1). The results illustrate that the classification accuracy through the use of selected set of features by ANOVA is increased compared to using the full set of features from 70% to 73.33% by DT, from 40% to 60% by both KNNCUB and KNNM, from 30% to 46.67% by SVMG, from 30% to 43.33% by KNNW. While the classification accuracies based on an ensemble system is improved from 40% to 43.33% by EBTree, the accuracy of ESDA remained at the same level at 33.33% (Figure 6.1).

When the selected set of features by the proposed HFSA was used in the multi-class classification of glioma grades, the results show that the classification accuracy is improved compared to both ANOVA methods and the use of the full set of features, where the best classification improvement is achieved by the DT classifier at 76.67%. Followed by the EBTree classifier where the classification accuracy is enhanced from 40% to 73.33%. The next best improvement is obtained by the SVMQ classifier where the accuracy is increased from 36.67 to 63.33%. Similarly, the accuracy of the SVMCUB classifier is enhanced from 40% to 53.33% (Figure 6.1).
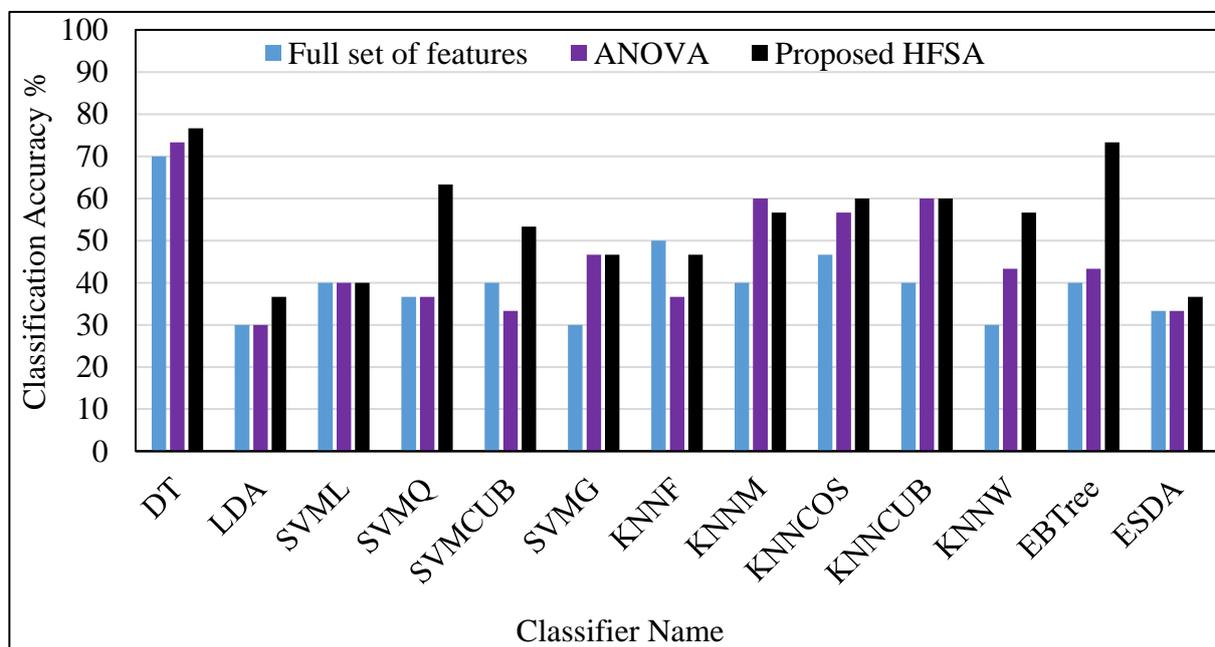
Figure 6.1 Comparative results for examinations of three cases.
The first case is the testing of full set of features, the second case is testing of the selected set of features by the ANOVA technique, and the third case is the testing of the selected set of features by the proposed HFSA. This test is to evaluate the behaviour of the proposed HFSA against other cases using different classification models for multi-class classification of glioma grades.

Although both the proposed HFSA and ANOVA reveal an enhancement in the classification accuracy for the multi-class classification of glioma grades, the proposed HFSA achieved better enhancement with many classifiers. Also, the proposed HFSA has achieved a better reduction in the feature dimensions, where the feature space is reduced by ANOVA from 234 to 122 features. The proposed HFSA has reduced the features space to only four features, which is considered a significant reduction in the features dimensions while gaining better classification accuracy.

The results show that the proposed HFSA achieved a notable improvement in the classification performance from 70% to 76.33% by single classifiers systems including DT and EBTree classifiers. It is noted that different classifiers reflect various behaviours for a different subset of features and mainly single classifier, showing relatively lower accuracy when developed for multi-class classification of glioma grades. While the ensemble system shows low classification accuracy due to its dependency on the majority vote that is limited to sensing only linear relationships. Nevertheless, in attempting to achieve further improvement and better classification accuracy in the multi-class classification of glioma grades, the proposed MTMCS is implemented and tested in the multi-classification of glioma grades. Furthermore, the hierarchical strategy is applied because of its ability to overcome the multi-class classification problem. Therefore this strategy is integrated with the proposed MTMCS for multi-class classification of WHO glioma grades, more details are included in the next section.

6.4 Hierarchical Meta-Trainable Multiple Classifier System

In this work, a hierarchical approach is adopted for the multi-class classification of glioma grades due to its superior capability, successfully outperforming other methods, to tackle the multi-class classification problem. Where the classifier members are not trained on multi-classes but instead on binary classification, within the proposed MTMCS through the tree structure of the hierarchical design, which they can perform better. The hierarchical strategy is integrated with the proposed MTMCS and denoted the hierarchical meta-trainable multiple classifier systems (HMTMCS). The proposed HMTMCS is then tested in the multi-class classification of glioma grades to examine the ability of this approach in improving the classification performance of glioma grades into the WHO glioma grades.

The proposed design was implemented based on developing each node of the tree of the hierarchical scheme using the proposed MTMCS. The design structure of the proposed

HMTMCS (Figure 6.2) which represents the proposed diagram for multi-class classification of the three classes where the proposed MTMCS was developed in Node 1 and Node 2 of the hierarchical structure (HS) and then both nodes were utilised for the decision making process for glioma grading. The proposed ensemble system has two main stages; the first stage is based on developing of single classifiers (Ψ_1, ..., Ψ_n), where Ψ represent a classification model, and n is the total number of classifiers. Then, the second stage is that the ODM produced from the classifiers are provided to DNN. For the classification of three classes (1, 2, 3), the proposed design starts with the first level of separation of the tree structure which includes the separation between class label Ω_3, and label Ω_2 based on the outcome of Node1. Due to the results from the discrimination between grade IV and the lowest grades (II, III) obtained at the full rate of 100%, the first separation of the HS starts with the development of the proposed MTMCS to classify between grade IV against the other grades (II, III).

The label Ω_3 indicates class 3 (grade IV), label Ω_2 designate grades (II, and III) that will be classified to label Ω_4 (class 1 or grade II) and label Ω_5 (class 2 or grade III). In more details, one-versus-all classification based on using MTMCS is determined in Node1 of the HS. The proposed HMTMCS will make a decision on an unknown sample and classify it as class 3 if a positive decision results from Node 1. Similarly, the unknown sample will be classified as belonging to the label Ω_2 if the output decision of Node 1 is negative. In the second split, that is, in Node 2 of the HS, another classification model is developed using the proposed MTMCS to discriminate the label Ω_2 into two classes (1 and 2 or grade II and III respectively). Similarly, positive and negative decisions generated from Node 2 will decide whether the unknown sample belongs to class 2 or 1 respectively (Figure 6.2).

Accordingly, a binary classification of glioma grades is applied in each tree node of HS, and then each node should produce a final decision on testing for an unknown brain tumour. For example, let class 3 refer to a brain tumour with grade IV and class 1 and 2 indicate grade II and III respectively. If the output decision of Node 1 is positive, then unknown brain tumour will be classified as grade IV, while if the output of Node 1 is negative, and then based on the output of Node 2, if it is negative, the unknown brain tumour will be classified as grade II. Similarly, if the outputs of Node 1 and Node 2 are negative and positive respectively, then the brain tumour will be classified as grade III.
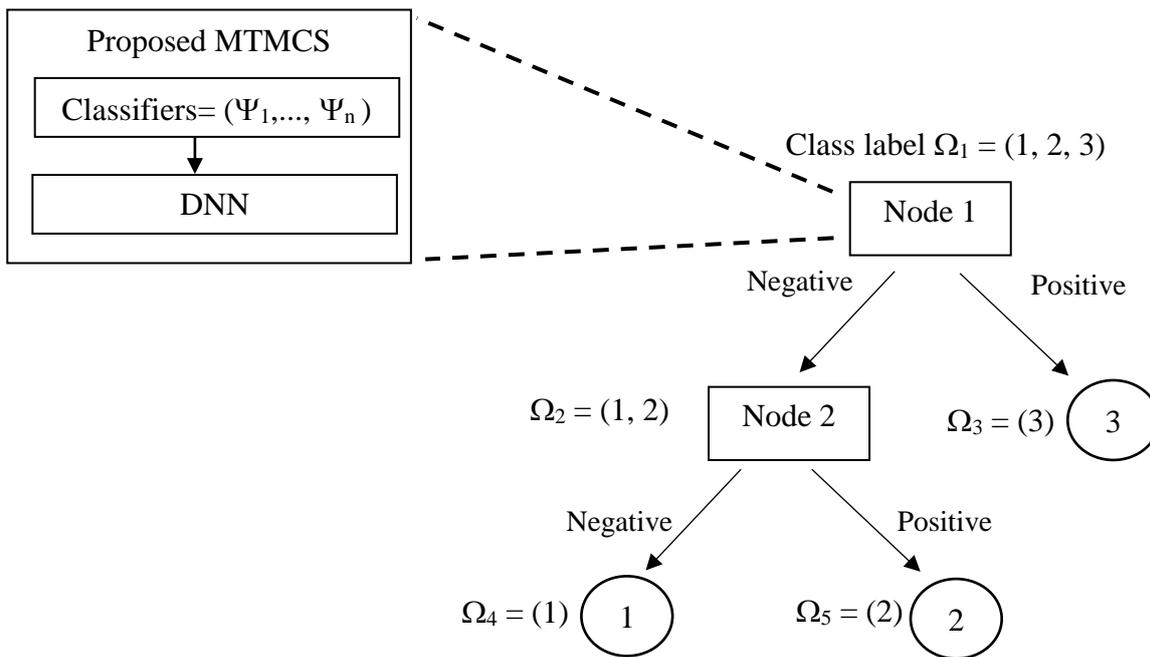
Figure 6.2 An illustration of the proposed hierarchical ensemble structure for classification of three classes (1, 2, 3), with two internal nodes, and three leaf nodes. The proposed MTMCS is developed in each internal node.

It should be noted that the setting parameters of classifiers ($\Psi 1$, ..., Ψn) undertaken in the proposed design are the same as the setting used with the single classifier system, which are illustrated in (section 3.6). Each node of the proposed HMTMCS is trained, tested, and evaluated independently. Two classification models of the proposed MTMCS are developed in the tree nodes of the HS, where both classification models are used to determine the multi-class classification of glioma grades. It is worth noting that through the development of the HMTMCS, it is necessary to change the class label to match the new subset of classes Ω, and then performing a new task of training and testing phases with the new class label. That is accomplished based on assigning different indices to different class labels. The SC algorithm is also investigated in the development of MTMCS within each node of the proposed design in which possible enhancement can be achieved.

LOO cross-validation technique is applied in all stages and all tree nodes of HS of the classification system to validate and add more generalisation to the proposed HMTMCS in the classification of glioma grades. Eventually, classification performance is evaluated and

compared in terms of the confusion matrix and classification accuracy between the proposed HMTMCS and other current approaches, including the single classifier and ensemble classification systems.

6.5 Results and Discussion

According to the HS that is utilised for the proposed MTMCS, which is developed at the first splitting of the HS in which the classification is conducted between glioma grades IV, and the other grades (II, III). Then at the next splitting of the HS, the system is further developed to classify the grades II against the grade III. These two developments of the proposed HMTMCS are conducted through two divisions of HS (Figure 6.2). Development of the proposed MTMCS at the first splitting of HS was already presented and thoroughly explained in (section 5.7.1) in which the results showed full discrimination rate between grade IV and the lower glioma grades (II, III). Therefore, in this chapter, the results of the development of the proposed MTMCS at only Node 2 will be explained and further discussed.

With regards to the system development at the second splitting of the HS or Node 2, the SC algorithm is applied with the proposed system in attempting to enhance the classification performance of the proposed system for multi-class classification of glioma grades. Considering the results obtained from performing this method, a full discrimination rate at 100% between grade II and grade III is achieved at Run1 (Table 6.4), in which the dimensions of classifiers are reduced by eliminating SVMG classifier. The highlighted classifiers in Table 6.4 are considered the significate classifiers that achieved the optimal classification accuracy at 100% by Run1; these classifiers are namely DT, LDA, KNNM, KNNCUB, KNNF, SVML, SVMCUB, KNNCOS, KNNW, and SVMQ. While other cases showed lower classification accuracy ranging from 90% to 95%, it is seen that the SC algorithm has significant impact on enhancing the classification performance of the proposed MTMCS in terms of accuracy and classifier dimensions in the discrimination between grade II and grade III. Consequently, the overall classification accuracy of the proposed system for the recognition of an unknown brain tumour has reached the full rate at 100% for all grades of glioma.

Considering the resultant confusion matrix, in which the results show that the proposed HMTMCS based on one-layer NNs is able to recognise all samples correctly and achieve

the full classification accuracy at 100% (Table 6.5) for all malignant samples using the proposed methodology and incorporated the SC algorithm.

Table 6.4 Selection process conducted based on the SCA. The first column in the right represents the sorted classifiers according to their corresponding classification accuracy at the first stage of the proposed MTMCS. Table cells that include 1 and 0 refer to keep and remove actions respectively, which are applied to classifiers in different runs for the system (Run1 to Run11). *ACCnew* represents the final classification accuracy of the proposed system through the selection process using the Cancer dataset.

Classifier Name	Run Initial	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run11
DT	1	1	1	1	1	1	1	1	1	1	1	0
LDA	1	1	1	1	1	1	1	1	1	1	0	1
KNNM	1	1	1	1	1	1	1	1	1	0	1	1
KNNCUB	1	1	1	1	1	1	1	1	0	1	1	1
KNNF	1	1	1	1	1	1	1	0	1	1	1	1
SVML	1	1	1	1	1	1	0	1	1	1	1	1
SVMCUB	1	1	1	1	1	0	1	1	1	1	1	1
KNNCOS	1	1	1	1	0	1	1	1	1	1	1	1
KNNW	1	1	1	0	1	1	1	1	1	1	1	1
SVMQ	1	1	0	1	1	1	1	1	1	1	1	1
SVMG	1	0	0	0	0	0	0	0	0	0	0	0
ACCnew %	90	100	90	90	90	90	95	95	95	90	95	95

Table 6.5 Resultant confusion matrix of applying the proposed HMTMCS based on one-layer of NNs in Node1 and Node2 of the hierarchical design for multi-class classification of glioma grades.

Actual Classes	Predicted Classes		
	GII	GIII	GIV
GII	10	0	0
GIII	0	10	0
GIV	0	0	10

These optimal results are yielded by implementing the proposed system based on using 8 neurons in the one-layer NNs (Figure 6.3 and Figure 6.4). Other neuron numbers have achieved good classification results; for example, when 28 neurons are used the results show 95% classification accuracy. Many other neuron numbers reflect lower classification accuracy at 90% such as for 5, 11, 15 and 16 neurons. With respect to the iterations of NNs conducted by the proposed methodology (Figure 6.4), various iterations reveal different classification accuracies where the majority of the iterations reflect accuracies ranging from 60 to 70%, while the accuracies of both 39th and 48th iteration are 80%. However, the optimal classification accuracy at 100% to discriminate grade II versus grade III is observed at the 43th iteration.
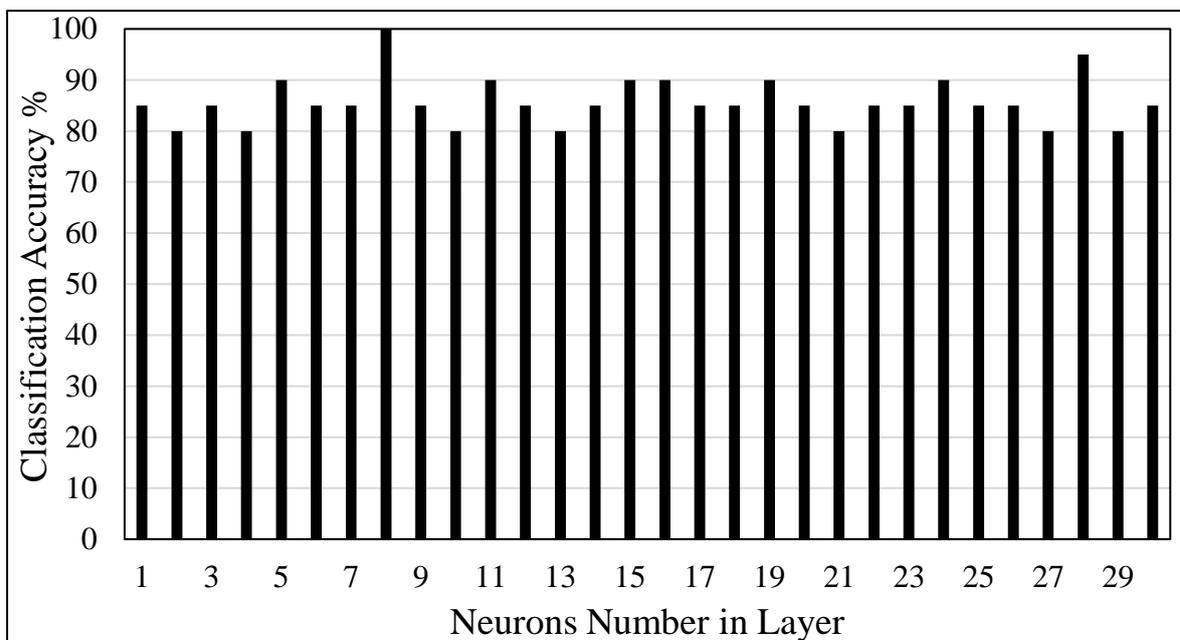
Figure 6.3 Classification accuracy results for applying the proposed HMTMCS based on one-layer NNs corresponding to the number of neurons in the layer to discriminate grade II against high glioma grade III using the Cancer dataset. These results are obtained from Run1 conducted by the SC algorithm.
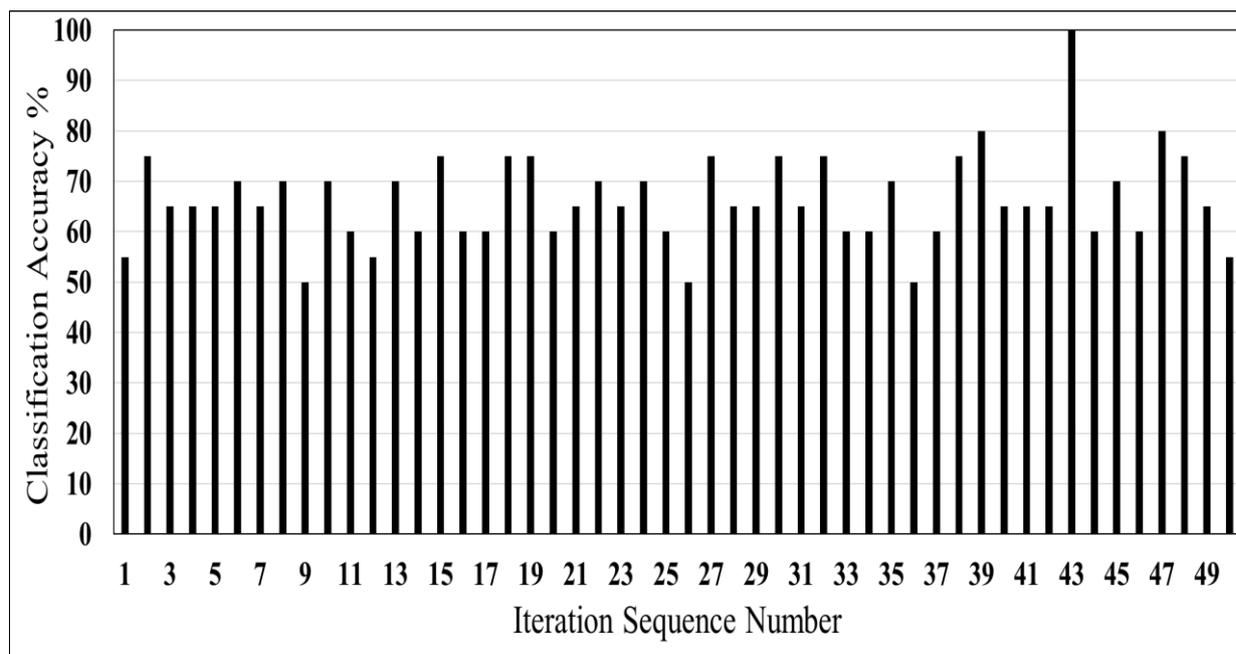
Figure 6.4 Classification accuracy results for performing the proposed HMTMCS corresponding to the iteration sequence number based on the 8 neurons in the one-layer NNs to discriminate grade II against high glioma grade III using the Cancer dataset. These results are obtained from Run1 conducted by the SC algorithm.

The overall comparative results in term of classification accuracy for the proposed system versus the single and ensemble classification approaches illustrate that the proposed system has achieved the optimal results at 100% outperforming all other approaches. The next best accuracy is achieved by DT classifier at 70% followed by the majority voting classifier at 56.66%, and with slightly lower accuracy at 50% shown by the KNN classifier while all other classifiers achieved lower accuracies with the range from 30% to 50% (Figure 6.5).

The proposed HMTMCS has achieved a full recognition rate between different WHO glioma grades. Unlike the single classifiers system for which the multi-class classification is a great challenge and can impact negatively on many classifiers, the proposed system can take advantage of the proposed MTMCS integrated with the HS, which led to significant classification results outperforming all other classification schemes. However, the proposed approach and according to the adopted HS has more classes involved, and more splitting by the hierarchical tree is required for additional classes. Consequently further system development and complexity is needed.
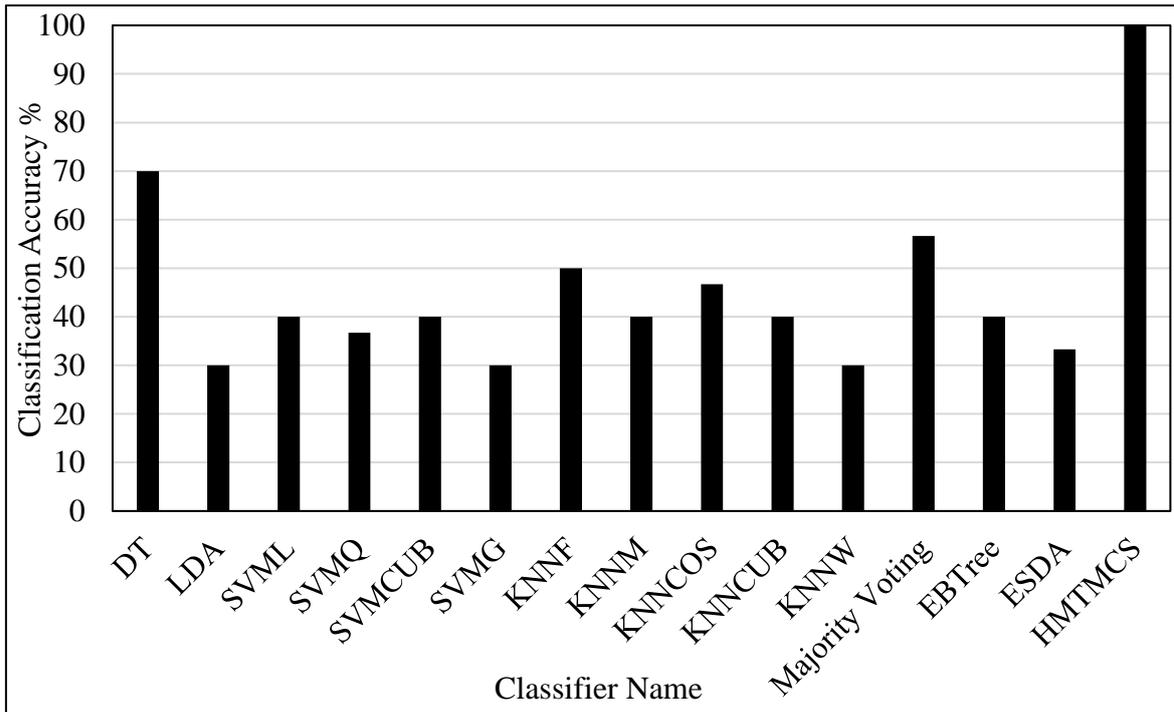
Figure 6.5 Comparative results in term of classification accuracy for the proposed HMTMCS against both the single classification model and the ensemble system for multi-class classification.

6.6 Conclusion

The proposed MTMCS was developed in the tree nodes of HS, and consequently a novel system HMTMCS was developed for the multi-class classification to classify three WHO grades of glioma (II, III, and IV). In the first splitting of the HS, the proposed MTMCS was developed to classify grade IV against the lower grade (II, III). The second splitting includes the development of the proposed MTMCS to classify grade II versus grade III. The classification performance of the proposed system was evaluated in terms of the confusion matrix and the classification accuracy. The LOO technique is applied in all stages of this work to validate the classification performance and to ensure the generalisation. Furthermore, the classification performance is compared in terms of this evaluation method with different popular single classifiers and ensemble classification systems. The results showed that the proposed system achieved the full accuracy rate at 100% in the classification of WHO glioma grades with the optimal recognition for each glioma grade.

CHAPTER 7 : Discussion, Evaluation, Conclusion and Future Trends

Overview

This chapter provides a summary of the research undertaken and reviews the aim, objectives, methodology, contributions and results obtained by the development of an automated classification system for glioma grades. The contributions of this work are explored in terms of feature extraction, selection and classification. The chapter also highlights the method used for performance evaluation, in addition to comparing the classification performance of the proposed system against other common approaches, through exploring the robustness and classification stability of the system's performance. Critical evaluation is conducted for the proposed system versus other recent methods in terms of features extraction, selection and classification. A critical comparison between the proposed system and other recent algorithms is discussed and presented. This chapter ends with the conclusion, findings, recommendations, possible opportunities, and suggested future works. The purpose of this chapter is to present a comprehensive overview and discussion of the research work conducted in this thesis with regards to different aspects highlighted and explored thoroughly which will give a clear understanding with summary of all the work done in this thesis.

7.1 Discussion

This section presents an overview of the research undertaken, starts with reviewing the aim and objectives of this thesis followed by an exploration of the research methodology, novelty and contributions. The review of novelty and contribution is highlighted in terms of features extraction, selection and classification. A summary of the overall results achieved by this experimental work is also discussed and reviewed.

7.1.1 Review Aim and Objectives

The aim of this thesis is to develop an automated classification system to discriminate different glioma grades and describe them in terms of WHO standard clinical grading schemes based on objective predictors extracted from MR images. Particularly, it explores

an effective solution to the glioma grading problem that would allow an automatic, objective and accurate classification for glioma grading.

This work was concerned with researching how automated methods could be defined to generate a fast, accurate and objective assessment for glioma grades from medical images.

To achieve the aim of this work, the following research objectives were established.

Obj.1 Review the literature to understand the problem domain and to identify the research requirement, opportunities and boundaries of the research undertaken, in addition to evaluating and investigating the robust methods and techniques, which support the development of an automated classification system for glioma grades towards specific goals of a non-invasive (without clinical surgery), automated and objective analysis.

Obj.2 Develop an effective method that would support the automated classification for glioma grades based on developing the feature extraction and selection stages.

Obj.3 Build a new classification approach within the development of the classification stage of the automated system for glioma grading which can achieve better discrimination for glioma grades.

Obj.4 Evaluate the new method experimentally for improving the classification accuracy by applying it and using common quantitative measures such as the confusion matrix. This seeks to determine that the new method aligns to the recent state-of-of the-art.

Initial work was undertaken to explore the research domain of glioma grading (**Obj.1**) through defining them in terms of the WHO grading scheme (**Chapter 2**), providing the possible approaches for the classification of glioma grades for developing an automated method to extract features from medical images (**Chapter 2**). In Chapter 2, the quantitative analysis of medical images using an automated method was explored, the refinement of this approach in terms of various aspects including features extraction, selection and classification with incorporation of machine learning algorithms were sought. At the end of the review chapter, the research boundaries, challenges, opportunities and requirements to achieve the aim of this study were finalised whereby the research directions were determined. A novel method was proposed for features selection and reduction to enhance the classification performance of the automated system for glioma grading in terms of classification accuracy and features reduction (**Obj.2**) (**Chapter 3**), where the classification

performance of the developed system based on the proposed method was evaluated using different popular quantitative evaluation metrics additional to further evaluation conducted using four benchmark datasets. The integration of this method with the machine learning algorithms has resulted in a significant improvement in the classification accuracy compared with both the original features and ANOVA technique where it achieved an optimal classification accuracy using the BRATS2013 at 100% while it achieved 93.3% with the use of the Cancer dataset, and 87.9% with use of the BRATS2015. This method has thus obtained a significant enhancement in terms of classification accuracy and features reduction. Conversely, it is observed that this method achieved lower classification accuracy using both Cancer and BRATS2015 datasets where this method relies on image texture feature that is highly subjected to the tumour homogeneity of MR images that may influence the level of classification accuracy. Nevertheless, in attempting to enhance the classification accuracy, new features were extracted from MR image of brain tumour independent to the variance in tumour homogeneity of MR image (**Obj.2**) (**Chapter 4**). The experimental results showed that the automated classification system developed based on the integration of these new features and the machine learning algorithm achieved comparable results, where it obtained 93.33% with the use of BRATS2013 dataset, and 90.51% with the use of BRATS2015 dataset. However, it is observed from both developed systems in Chapter 3 and Chapter 4 that the machine learning models exhibit various behaviour and different classification accuracies dependant on the use different datasets. This is due to the different sensitivities of single classification systems affected by various data distributions generated by different datasets. This created the motivation to develop further the classification system that is able to overcome this challenge and has more robustness in handling the variation in data distribution with a more stable classification accuracy. This was achieved by developing two-stage learning strategy through the integration of multi-classification models using DNN conducted by the proposed MTMCS (**Obj.3**) (**Chapter 5**). A part of the work undertaken in Chapter 5 has been published in International Conference on Automation and Computing (ICAC'2018) (AlZurfi et al., 2018).

The multi-class classification of glioma grades was also demonstrated to achieve accurate classification of different WHO grades of glioma (II, III, and IV) (**Chapter 6**). In this chapter, the proposed MTMCS was developed by taking advantage of using the hierarchical scheme to optimise the multi-class classification of glioma grades. The proposed HMTMCS has achieved the optimal classification accuracy at 100% for glioma grading into different

WHO grades (II, III, and IV). In this work the most efficient method is used including the SC algorithm. The results showed that the SC algorithm has enabled the proposed system to reach of its optimal accuracy as well as gaining a significant reduction in the dimensions of the classifier members by further reduce the number of classifiers required to achieve the optimal classification results of glioma grades.

All samples used in this work are pre-diagnosed in advance, and thus they are utilised as class truth label through the supervised learning process. Consequently, the evaluation metrics presented in this research were calculated based on comparing the outcome of the proposed classification system with the class truth label (**Obj.4**). The classification performance of the developed system is evaluated in terms of the classification accuracy, sensitivity, specificity, precision and F-measure, demonstrated from the confusion matrix that is also presented in this thesis. The evaluation also covers the application of the leave-one-out cross-validation technique, to validate the classification model. In this technique the model is tested using new samples that have not been seen by the trained model through splitting the dataset into k-folds (multi subsets). Each fold should be used twice, first is in the training phase and the second in testing phase with the replacement. This process should be repeated to cover all samples in the dataset. The outcome of the proposed system is compared with the recent common methods in terms of the evaluation metrics. The evaluation process is extended further by comparing the proposed classification system with the existing algorithms using four public datasets. The evaluation procedure mentioned above is conducted and presented in each chapter of the thesis to create a deep evaluation of the developed system based on the new method.

7.1.2 Review Methodology, Contributions and Implications to the Literature

The research design adopted in this work is a combination of quantitative and experimental research approaches to answer the research questions and achieve the aim and objectives of this thesis. Within this research design, all steps and processes of the proposed work are developed and evaluated quantitatively, and the empirical design is utilised to emerge the findings, analysis results and draw conclusions. The work starts with a review the research problem of this work, exploring existing solutions, boundaries and possible opportunities and accordingly a proposed solution was formulated. The performance of the proposed solution is validated and evaluated through implementing it and using several widely used evaluation techniques that objectively measure its performance using several different

datasets. This seeks to determine if the new method developed by the proposed solution has a competitive outcome and is in line with the standards by the current state-of-the-art.

The key contribution of this study is the automated classification system and the methodology by which this is undertaken for distinguishing various WHO glioma grades. This will offer aid to the clinicians to achieve a diagnostic decision and support them towards an accurate, objective and automated classification of glioma grades.

The main contributions and novelties of this study are summarised as follows:

1- New features extracted from MR images of the brain tumour are proposed, based on generating ratio predictors and an objective analysis extracted from the presence of different descriptors of a brain tumour, such as contrast enhancement, non-enhancement, necrosis, and edema, which offer a more reliable analysis of MRI attributes of a brain tumour. The existing work relies on the expert domain to analyse these tumour descriptors, which have many challenges such as the inter and intraobserver variabilities and does not consider the objective analysis of the relations with each other. Similarly, the influences of these features or their relations were not investigated with machine learning algorithms on the classification of glioma grades. The proposed method is beneficial as the discrimination ability of these tumour descriptors can be analysed and utilised to differentiate malignant gliomas. Consequently, more benefit from these tumour descriptors can be gained to improve the classification accuracy of glioma grades. This is the first study that proposes and examines the impact of the new features on glioma grading.

2- A new feature selection method is developed. The advantages of this method are eliminating redundant features, not only related to maintaining the same level of accuracy but also achieving further improvement in the classification accuracy for glioma grades. This method is based on taking advantage of a fusion between filter and wrapper methods. It is based on the correlation analysis incorporated with several classifiers to update and guide the selection process.

3- A comprehensive analysis of the three-dimensional textures feature based on 3DGLCM is developed to support an automated MRI classification system for glioma grades.

4- A novel method is proposed to support the integration of different machine learning algorithms that further improve the classification accuracy for glioma grades. This includes the development of the meta-trainable strategy based on Deep Neural

Networks (DNN). The existing works mainly concentrate on either using a single machine learning algorithm or using one stage of the multiple classifier systems. Therefore, better classification accuracy can be achieved by the development of a multiple classifier system (MCS) based on two stages of learning for glioma grading.

5- An effective method is proposed to optimise the output accuracy of DNN, which provides a systematic trainable design for the MCS that can be beneficial to improve the classification of glioma grades. The existing work applies the few trials and selected randomly in attempting to achieve the best design and optimal parameters of the NNs. While, in development of the DNN a systematic approach can play an important role in the optimisation of MCS and improve the classification accuracy for glioma grades.

6- A new method is proposed to reduce the number of classifiers required for the proposed system achieving a significant improvement in the classification accuracy. This method also eliminates many of the redundant classifiers and keeps only the most active classifiers, which enable the developed system for more efficient classification performance for glioma grades, in terms of classification accuracy, and dimensions reduction. Consequently, this leads to reduce the complexity of system design and add further system enhancement.

Further discussion on the contributions and novelties that have been achieved by this study is found in the following subsections.

7.1.2.1 Features Extraction

Many existing current works rely on the combination of using an advanced MRI technique in addition to the conventional MRI method in the assessment of glioma grades. However, the advanced technique costs more, require advanced instruments and a high level of qualification, which is not available in all MRI clinical centres. Alternatively, developing a classification system for glioma grades based on the conventional MRI technique would be of great benefit to those who do not have access to an advanced MRI technique. Accordingly, this thesis has been developed based on the conventional MRI methods particularly, T2- MR images for extracting the texture features, while the other tumour descriptors are originally provided by using T2, T1, T1c, and Flair MRI modalities.

It is widely known that the image texture feature is the most common feature to access the tumour heterogeneity, specifically the textural predictors extracted from the GLCM are the

most efficient features used for brain tumour classification. Therefore, this feature has been utilised and assessed in this thesis. It has been shown that these features could be subjected to the variance in tumour homogeneity (**Chapter 3**). Consequently, in attempting to enhance the classification performance and solve this problem, new features are proposed within a novel approach, which are independent to the variance in tumour homogeneity of medical image of a brain tumour. The idea of generating the novel features started from the standpoint that quantitative ratios among the presence of the tumour descriptors integrated with machine learning algorithm can lead to an effective classification of glioma grades. The developed classification design based on the novel features that were assessed thoroughly (**Chapter 4**). A critical comparison between the developed system based on these features and the textural-GLCM features are discussed (**Chapter 4**), where it is proven that the developed design based on the new features has achieved comparable results in the classification of glioma grades.

7.1.2.2 Feature Selection

One of the major challenges in the design of a classification system is the selection of the most crucial features that can contribute to enhance the classification accuracy and eliminates redundant features. To tackle this challenge, a novel method was proposed and utilised for the feature selection and features dimensions' reduction. This algorithm has been called HFSA, which was evaluated and incorporated with different machine learning algorithm in the classification of glioma grades. The performance evaluation successfully proves the usefulness of this algorithm in enhancing the classification performance in terms of classification accuracy and the reduction of features dimensions. It has been confirmed that this algorithm enables the single classification model to show better classification accuracy over the four MR images datasets. In addition, this algorithm contributes in reducing the features space to smaller dimensions that will decrease complexity in the system design (**Chapter 3**).

The proposed MTMCS has been proven to achieve the best accuracy without the necessity to involve features selection method in the first stage of the proposed MTMCS. The experimental results showed that the proposed system provides the best classification accuracy irrespective of whether the original set of features or the selected set of features are used (**Chapter 5**). The reason behind this is that the proposed system is robust to the variation in input features while it is relatively more dependent on the input of the second stage rather than the input of the first stage of the proposed system. The proposed MTMCS

has taken the merit of each classifier member, in addition to the complementary relationship between them. Consequently, this overcomes the limitation that could have occurred from using different set of features. However, if the selected set of features causes many drawbacks in the sensitivities or the specificities of the classifier members, then this would lead to slight degradation in the final classification accuracy. In contrast, it is possible to overcome this potential issue and optimise the classification accuracy by adding a further hidden layer of NNs within the proposed system in which the classification performance can be improved further to a better classification result.

Selection of best set of classifiers that has a significant contribution to the classification accuracy of the proposed system is a big challenge. The outcome of the machine learning models, which were integrated within the proposed MTMCS, can play the greatest role in drawing the shape of the classification behaviour of the proposed MTMCS. Therefore, the SC algorithm was utilised in the second stage of the proposed MTMCS, through taking the merit of different outcome of the classifiers members to select the best set of classifiers and enhance the classification accuracy of the proposed system. This algorithm was used with the proposed MTMCS to overcome the challenge of the classifiers selection and gain further reduction in the design of MCS. The results of this algorithm showed a significant ability to select the best set of classifiers and enabled the proposed MTMCS to demonstrate better classification accuracy. This algorithm is also advantageous in achieving better reduction of classifier dimensions, which can lead to gain further decrease in the complexity of the proposed system design (**Chapter 5**).

7.1.2.3 Classification

To achieve an effective classification model, which can show a stable classification performance in term of classification accuracy regardless of the potential variation in data distribution, various datasets or different features is a challenging task. In this research work, two directions of developing machine learning-based classification systems for glioma grades were explored. The first direction was concerned with developing a single classification model, and the second direction was based on developing a MCS. In both directions, the best and efficient strategies have been used in attempting to enhance the classification accuracy of glioma grades. It was found that the proposed MTMCS, which is based on MCS, showed more robustness and stable classification accuracy over all experiments and datasets undertaken in this work. The superiority and stability in classification performance of the proposed MTMCS are due to developing two stages of

learning based on the integration of different classification models where the advantage of each single classifier has been considered in a complementary strategy that can overcome the limitation that may arise from weak classifiers. However, integrating these classifiers to achieve an effective design of MCS is a challenge. Most existing methods rely on the majority vote, which is limited to sensing a nonlinear relationship among integrated classifiers. Therefore, DNN was utilised to integrate the classification models due to its ability to sense the intricate relationship among classifiers, leading to enable the proposed MTMCS efficiently to further overcome any drawback that may arise because of weak classifiers, which leads to improving the classification accuracy of glioma grades further (**Chapter 5**). Using DNN in the integration of multiple classification models is also advantageous because it is a systematic approach examining several weights multiplied by the output decisions of the classifier members through generating varied range of weights by the backpropagation strategy and the process of searching for the optimal design of NNs (**Chapter 5**).

The other challenging issue is how to optimise NNs and choose the most efficient design of NNs that enable it to achieve the best possible classification accuracy. Possible existing solution for this problem is based on determining random trials and selecting the best one among them (Khalid and Noureldien, 2014). Although this solution is efficient requiring few trials and may lead to better result, it has a lack of generalisation ability and it is not a systematic approach. To overcome this challenge, a novel method is performed by developing the DINN that enable the DNN to achieve its highest classification accuracy and choose the best number of neurons and iterations, leading to the best most efficient design of the proposed MTMCS. The proposed system was proven to obtain an efficient classification accuracy over all experiments and datasets and showed the most stability and robustness in classification performance regardless of the variance generated by using different datasets (**Chapter 5**).

7.1.3 Overall Results

The overall summary of results generated in this thesis for BRATS2013, BRATS2015 and BRATS2018 datasets are presented in Table 7.1, and for the cancer dataset are shown in Table 7.2, in which the work starts with developing an automated classification system based on the textural features measured from GLCM extracted from the T2-MR images of brain tumours. The textural features were utilised and evaluated by using several machine learning algorithms within a single classifier system.

Table 7.1 Overall summary and comparisons for the results obtained by the experiments developed based on BRATS2013, BRATS2015 and BRATS2018 datasets. Sensitivity and specificity are associated with high and low grades of glioma respectively. Cells marked with asterisk (*) indicates the cease of system development as there is no rationale to develop it further.

No.	Approach Types	Method design	BRATS2013				BRATS2015				BRATS2018			
			ACC	Sensitivity	Specificity	Precision	ACC	Sensitivity	Specificity	Precision	ACC	Sensitivity	Specificity	Precision
1	Single	Texture + single classifier	**93.33**	95.00	90.00	95.00	**85.77**	90.00	68.52	92.10	**86.32**	92.86	68.00	89.00
2	Single	Texture + single classifier + proposed HFSA	**100**	100	100	100	**87.9**	93.18	66.67	91.90	**88.07**	93.33	73.33	90.70
3	Single	Proposed FTD + single classifier	**83.33**	90	70	85.7	**89.05**	97.27	55.56	89.90	**89.12**	94.76	73.33	90.87
4	Single	Proposed FTD + single classifier + proposed HFSA	**93.33**	95	90	95	**90.51**	96.81	64.81	91.80	**93.33**	99.05	77.33	92.44

Table 7.1 Continued.

No.	Approach Types	Method design	BRATS2013				BRATS2015				BRATS2018			
			ACC %	Sensitivity %	Specificity %	Precision %	ACC %	Sensitivity %	Specificity %	Precision %	ACC %	Sensitivity %	Specificity %	Precision %
5	MCS	Texture + proposed MTMCS+SC	**100**	100	100	100	**91.24**	97.27	66.67	92.20	**91.58**	97.62	74.67	91.52
6	MCS	Texture+ proposed FTD + proposed MTMCS (one layer) +SCA	*	*	*	*	**93.07**	99.09	68.52	92.80	**95.09**	98.10	86.67	95.37
7	MCS	Texture + proposed FTD + proposed MTMCS (Two layer of NNs) with diversity	*	*	*	*	**93.43**	98.18	74.07	93.91	**94.39**	98.10	84.00	94.50
8	MCS	Texture + proposed FTD + proposed MTMCS (one layer of NNs) with diversity + SCA	*	*	*	*	**93.80**	98.64	74.07	93.9	**94.74**	99.05	82.67	94.12

Table 7.2 Overall summary of the results obtained by the experiments conducted based on the Cancer dataset. Sensitivity and specificity are associated with high and low grades of glioma respectively.

No.	Approach Types	Approach Design	Cancer dataset (low (II) against high (III, IV))			
			ACC %	Sensitivity %	Specificity %	Precision %
1	Single	Texture + single classifier	**86.67**	90.00	80.00	90.0
2	Single	Texture + single classifier + proposed HFSA	**93.33**	95.00	90.00	95.0
3	MCS	Texture + proposed MTMCS	**96.67**	100	90	95.23
4	MCS	Texture + proposed MTMCS +SC	**100**	100	100	100

In this part of the work a novel algorithm for features selection and reduction was proposed. The purpose of this algorithm was to enhance the classification performance by eliminating the redundant features. This was performed based on the integration of ANOVA technique and Pearson correlation in addition to involving different classification algorithms, which are utilised in a search strategy to find the optimal set of features. The results obtained from applying the developed system based on the original features without the proposed HFSA in terms of classification accuracy was 93.33% for BRTAS2013 dataset, 85.77% for BRATS2015 dataset and 86.32% for BRATS2018 dataset (Table 7.1, No.1). Afterwards, when the proposed HFSA was determined, the classification performances in terms of classification accuracy, sensitivity, specificity and precision were improved to optimal results at 100% using BRATS2013 dataset (Table 7.1, No.2), and to 87.9%, 93.18, 66.67% and 91.90% respectively by BRATS2015 dataset and to 88.07%, 93.33%, 73.33% and 90.70% respectively by BRATS2018 (Table 7.1, No.2). When the proposed HFSA was evaluated with Cancer dataset, the classification performance in term of classification accuracy also showed a significant enhancement from 86.67% to 93.33% (Table 7.2, No.1 and No.2). Consequently, the performance evaluation of the classification system after using the proposed HFSA has shown an improved classification accuracy over the four datasets. This method used the combination of a filter approach presented by ANOVA, the wrapper approach based on Pearson correlation and the outcome generated by testing several different subsets of features. To further elaborate, in this algorithm, ANOVA was used to remove the redundant features that have no statistical significance in the features space. After that the Pearson correlation among features was used to guide the search process for an optimal subset of features. This also considers the feedback in term of classification accuracy produced from different classifiers. This feedback also includes the outcome of different subset of features generated by correlation interaction among the feature set. This generates the ability to consider the interaction among features and updating the selection process based on the outcome from different machine learning models using the generated subset of features. Consequently, this supports the ability of the proposed algorithm to select an efficient subset of features, which enables the classification model to achieve the best classification performance in terms of classification accuracy and reduction in the feature dimensions.

It has been shown that texture features are dependent on tumour homogeneity, where it is analysed statistically using box-plot (**Chapter 3**). The tumour homogeneity of an image

plays an important role in determining the output behaviour of the classification performance. While the optimal classification accuracy at 100% is achieved with BRATS2013 dataset, the results of other datasets showed lower classification accuracies. The reason behind this is that there is large variance in the level of homogeneity of the tumour image, and thus the classification accuracy can be affected when there is a high level of homogeneity of an image, leading to lower classification accuracy, such as results achieved by both BRATS2015 and BRATS2018 dataset. In attempting to enhance the classification accuracy, new features were proposed, which were extracted from the tumour descriptors of brain tumour namely tumour necrosis, edema, enhancement and non-enhancement. The results exhibit comparable classification accuracy to the previous developed classification system based on textures (Table 7.1, No. 2 and 4). Unlike the texture features, these features are independent of tumour homogeneity of an image and therefore the developed system based on these features integrated with a machine learning algorithm showed competitive classification results as they achieved better results at 90.51% and 93.33% when assessed with the BRATS2015 and BRATS2018 dataset respectively (Table 7.1, No. 2 and 4). However, the result obtained from applying the classification system based on these features was 93.33 % for BRATS2013 dataset, and thus the developed system based on these features did not show optimal accuracy compared to the developed system based on the textural features. This is due to the nature of the presence of these tumour descriptors, for example, the presence of tumour enhancement which is dependent on the leakage in the blood-brain barrier caused by the tumour cell invasion (Li et al., 2015), which consequently led to a variance in the classification outcome depending on the behaviours of malignant tumours.

It is observed that both developed systems based on a single classification approach have a high variation when examined with different datasets. They are affected by different issue such as the difference in data distribution, homogeneity level of image sample, and selecting an appropriate classifier. These were the most challenging issues in finding the classification system with best classification accuracy. In addition, the developed classification model that achieved high accuracy is not the same overall datasets. For example, the results obtained by testing the developed system based on a single classification approach illustrated that the best classifier compared to all other classifiers was the KNNF classifier for BRATS2013. While for the Cancer dataset, the results showed that the classifier achieving the best classification performance was the DT classifier. Consequently, the developed classification

system based on a single classification approach has a low stability in classification performance. In attempting to overcome this challenge, a novel deep learning approach based on two stages of learning within multiple classifier systems was proposed. The results showed that the classification accuracy of the proposed MTMCS was improved compared to the use of a single classification system where the proposed MTMCS based on only the textural feature has achieved an optimal classification accuracy at 100% for BRATS2013, and 91.24% for BRATS2015 and 91.58% for BRATS2018 (Table 7.1, No. 5). Similarly, it achieved high classification rate at 96.67% (Table 7.2, No.3) for the Cancer dataset. While the best classification accuracy achieved for the evaluation of the proposed MTMCS with BRATS2015 dataset was 93.8%. This best result was obtained through applying the proposed MTMCS based on one-layer of NNs and using the texture features and the proposed feature associated with tumour descriptors, and through integrating with the SC algorithm and enabling the diversity in the feature space (Table 7.1, No. 8).

To further enhancement the classification performance of the proposed MTMCS, the SC algorithm was investigated to select the best set of classifiers that have the highest impact on the classification accuracy and produce a better reduction in the classifier dimensions. When the SC algorithm was utilised with the proposed MTMCS for BRATS2013 dataset, the results illustrated a significant enhancement in the classifier dimension where the classifier members were reduced from 11 to 5 classifiers (**Chapter 5**) as well as maintaining the optimal classification accuracy at 100% (Table 7.1, No. 5). While with BRATS2015 dataset, the classifiers space was reduced from 11 to 7 classifiers (**Chapter 5**) and the classification accuracy achieved was 91.24%, and for BRATS2018, the classifier dimensions were further eliminated from 11 to 6 classifiers with 91.58% of classification accuracy (Table 7.1, No. 5). This is an important achievement as the reduction in number of classifiers decreases the complexity of the system design with a significant classification accuracy was achieved.

When the SC algorithm was investigated with the proposed MTMCS using the Cancer dataset, the results showed that the classification performance in terms of classification accuracy, was improved from 96.67% to 100% (Table 7.2, No.3 and 4), and in terms of classifier dimensions, the result indicated that the classifier dimensions were reduced from 11 to 7 classifiers (**Chapter 5**). The SC method has shown a notable improvement when utilised with the proposed MTMCS using different datasets. This was achieved by applying the proposed MTMCS based on the combination of texture features and proposed FTD, and

with enabling the diversity in feature space. The results showed that the proposed approach achieved notable classification accuracy at 93.80% for BRATS2015 and 94.74% for BRATS2018 (Table 7.1, No. 8) with a less complex design based on only one layer of NNs. The advantage in developing the proposed MTMCS incorporating the SC method are as follows; (i) superior enhancement was achieved by using only one layer of NNs, (ii) a lower number of classifiers were obtained as the classifier dimensions were reduced from 55 to 52 classifiers (**Chapter 5**). The SC algorithm enables the proposed MTMCS to select the best set of classifiers and maintain the highest classification accuracy by considering the outcome from the proposed MTMCS using different subset of classifiers generated by the SC algorithm. Consequently, the classification performance was further enhanced in terms of classification accuracy and classifier dimensions.

It was also observed that a significant evolution was achieved in the classification accuracy when an additional hidden layers of NNs were added and utilised in the proposed MTMCS, where the classification accuracy was improved by using two-layer of NNs to 93.43% for BRATS2015 and to 94.39% for BRATS2018 dataset (Table 7.1, No. 7). However, it requires high computation time and extensive complexity in the system design. Alternatively, the integration of the SC algorithm with the proposed MTMCS based on one-layer and the combination of FTD and the texture features has also achieved a notable increase in classification accuracy with gaining lower complexity in system design to obtain 95.09% for BRATS2018 dataset (Table 7.1, No. 6).

It is observed that the combination of the new features extracted from the brain tumour descriptors and texture features derived from the BRATS2018 have a notable contribution to support different machine learning algorithms and particularly in achieving better sensitivities and specificities of BRATS2018 compared to BRATS2015 (APPENDIX C Table 7, Table 5.15). Accordingly, this leads to achieve better classification accuracy results. The reason behind this is that samples from BRATS2018 dataset have better recognised features that reflect the structure of the descriptors of brain tumour. In more details, it is noted that the statistical ration of contrast enhancement (tC_R) measured from samples of BRATS2018 is more significant at 9.0504×10^{-26} (Table 4.9) than the P-value measured from samples of the BRATS2015 which showed lower P-value at 3.8146×10^{-18} (Table 4.6). This difference in the significance level of contrast enhancement of the two datasets is because of the nature of the malignant tumour and its aggressiveness to infiltrate different region of the brain tumour. Notably, the different effect on the blood-brain barrier of the

brain tumour can occur as the more leakage in the blood-brain-barrier indicating the more contrast enhancement in the brain tumour (Xi et al., 2019, Geneidi et al., 2015).

The SC method has a great influence on enhancing the classification performance of proposed MTMCS. This is basically due to the ability of the SC algorithm to eliminate the classifier that has lowest impact on the outcome of the classification accuracy of the developed system and eliminate the classifier that causes conflict with others thereby enabling the proposed system to reach its highest classification performance. Also, considering the outcome of the developed system as feedback is a crucial approach to control the search strategy for the best subset of classifiers, which can enable the proposed system to achieve the best classification accuracy and design.

In the proposed MTMCS, an intricate design was conducted based on the development of two stages of learning of MCS incorporated the DINN. In this design, a number of neurons ranging from 1 to 30 in which each neuron was investigated with 1 to 50 iterations, producing 1500 experiments determined by the proposed MTMCS for one sample in the dataset. Experimentally, using large number of neurons or iterations for the proposed MTMCS requires extensive computation time on a standard Personal Computer. Consequently, the classification accuracy of the proposed MTMCS has been measured against a number of neurons ranging from 1 to 30 and number of iterations varied from 1 to 50, which are used in the development of DINN. The notable classification accuracy does not vary linearly with neither the number of neurons nor the number of iterations used in the development of the proposed MTMCS. It is observed that there is no significant pattern of variation between the classification accuracy and the number of neurons or number of iterations where the variation is highly irregular. In addition, the optimal classification accuracy achieved by the proposed MTMCS does not exceed the 50th iteration or 30 neurons in all experiments over all datasets. Therefore, these ranges are all that is considered in the design of the proposed MTMCS.

7.2   Evaluation

The proposed system was evaluated using different common statistical measures including the confusion matrix, sensitivity, specificity, precision, and F-measure. These metrics were used in the comparing the classification performance of the proposed system against different popular single classifier and ensemble classification methods. To validate the classification performance and investigate the ability of generalisation, the leave-one-out

cross-validation technique was utilised in all stages and for all classification methods applied in this work. In the cross-validation technique, the developed system is tested and evaluated by using samples not seen by the developed system through the training phase. This method is vital to achieve an accurate evaluation and to avoid an overfitting problem. Further evaluation was added by examining the classification performance of the proposed system using four different benchmark datasets where the behaviour of the proposed system was measured, assessed, and compared with other alternative approaches over all the four datasets.

7.2.1 Stability of Classification Accuracy

It is observed that selecting an appropriate classifier that can perform successfully in the classification of a dataset is a challenging task and is the most important issue to find a classifier system able to achieve optimal classification accuracy. Particularly in this work, a single classifier system achieved the optimal accuracy at 100% using BRATS2013, in which samples have good spatial resolution with an appropriate level of homogeneity, which highly impacts the amount of texture extracted from the medical images. However, using different datasets can reduce the classification performance. It is noted that when the BRATS2015 dataset was used, in which the level of homogeneity is higher compared to the samples from BRATS2013 dataset, the single classifier system achieved lower classification accuracy. It is also noted that the best accuracy is not always achieved by the same classifier for different datasets. It is observed from the results of different experiments that even though a single classifier system can achieve optimal classification accuracy, it has less stability in the prediction performance for different datasets. This is possibly a consequence of various sensitivities of classifiers to different data characteristics, distributions and sample sizes. These findings mentioned above agree with the results of the recent study carried by Zhang et al. (2017), in which they discussed the complexity of selecting an optimal classification approach for the complex cancer dataset, stating that it is a big challenge to improve the stability and generalisation ability and overcome the high variations in the classification outcome.

However, to overcome the above-mentioned challenges, a novel MTMCS has been proposed and developed for glioma grading based on the integration of a multi-classification model and by continuing to train the weight of output decision of each classifier by DNN, overcoming the drawback of a weak classifier in a complementary strategy. In the proposed system, classifiers with higher accuracy have a superior role to play and interference

information of the classifiers with minor accuracy is excluded. Therefore, the advantages of each classifier are entirely considered, and utilised, and improved prediction performance was achieved for all datasets, providing significant evidence that the proposed MTMCS is more stable in the classification performance and is more robust to the variation of input data characteristics.

7.2.2 System Scalability

One of the significant factors that are challenging many machine learning algorithms and affect the classification performance is the number of samples in the training phase, which are used to learn the classification model (Pan et al., 2015). For instance, it is a significant issue with the traditional classification system based on the deep learning approach (Mohsen et al., 2018, Hegde et al., 2019). However, to overcome this drawback, a new approach was adopted by developing the MTMCS. To elaborate, the system is developed based on utilising the advantage of the fusion of several classification models that are less affected by the problem of limited sample size in the training phase (Kuncheva, 2014). Furthermore, the complementary of these classifiers leads to optimise the output decisions of the classifiers through an optimising process and to continue learning these output decisions through the meta-trainable strategy, which achieves the best possible classification accuracy and more scalable performance. Accordingly, the proposed MTMCS has the better scalability to adapt the large and small number of samples leading to more robust performance than other recent approaches for glioma grading. The proposed MTMCS has also been experimentally evaluated using several datasets with different sizes and distributions where the results showed that the proposed system has achieved the more stable performance of enhanced classification accuracy as compared to the traditional single and ensemble classification models. However, due to the complexity of system design generated from developing two stage of learning, the more samples would lead to larger computation time in the training phase.

7.2.3 Critical Evaluation

In this section, the entire work developed in this thesis is discussed and critically evaluated in comparison with different alternative approaches that are recommended by the literature as efficient and successful approaches. The discussion covers the comparison in terms of three aspects including features extraction, selection, and classification.

7.2.3.1 Features Extraction

This work starts with exploring the problem domain and possible opportunities for classification of glioma grades. Consequently, several tumour descriptors were investigated where the concentration was given to tumour heterogeneity, necrosis, edema, enhancement and non-enhancement due to their popularity and that they can be extracted from the conventional MRI modalities, which are usually available in clinical MRI centre. However, there are other tumour descriptors which can be utilised for assessing the malignancy level of glioma grades such as tumour vascularity and cellularity, which require a high level of qualification, more costs, and an advanced MRI technique (Kono et al., 2001, Geneidi et al., 2015).

Among the tumour descriptors, tumour heterogeneity is the most widely used and, especially for studies, utilises quantitative analysis to characterise a tumour using the statistical features extracted from medical images. Tumour heterogeneity is also successfully investigated in the classification and segmentation of many types and grades of brain tumours (Roy et al., 2013, Tantisatirapong, 2015). Therefore, the heterogeneity descriptor of brain tumour was considered in this work through extracting several popular statistics measured from GLCM using MR T2-weighted images (Gómez et al., 2012, Larroza et al., 2016) (**Chapter 3**).

Tumour necrosis, edema, enhancement and non-enhancement are common tumour descriptors used in the visual diagnosis for evaluating the malignancy of glioma grades. They are also frequently provided as a standard segmented data in addition to the whole region of brain tumour by two benchmark datasets (BRATS2013, and BRATS2015). The identification of these tumour descriptors is an important and challenging task (Menze et al., 2015). Consequently, in this thesis the usefulness of these tumour descriptors and the proposed features thereof were investigated through novel quantitative criteria to assess the malignancy level of glioma grades. These features were utilised in the development of an automated classification system for glioma grades (**Chapter 4**). Several MRI modalities were suggested by the literature to assess the brain tumour characteristics. For example, T1c-weighted was used for texture extraction and employed in developing a classification system for glioma grades (Hsieh et al., 2017b). Despite this modality being the least invasive procedure with contrast enhancement, it is highly dependent on leakage in the blood-brain barrier, which can affect the tumour structure and consequently affect the amount of texture extracted from this modality. However, the proposed work in this thesis, which is concerned with tumour heterogeneity, extracted texture features based on T2-modality by a non-

invasive process and brain tumours are usually possible to recognise visually using this MR modality. Nevertheless, the proposed features associated with FTD were measured using the standard segmented dataset, which originally used T1c modality in delineation of the brain tumour sub-regions.

It was argued that the visual diagnosis based on conventional MRI technique is less sufficient compared to the advanced MRI method in the classification between low-grade glioma and high grades (Law et al., 2003). In that research work, recognised information extracted from advanced MRI has shown a better result than the traditional diagnosis based on conventional MRI (the diagnosis was performed by two qualified experts blinded to the actual label of glioma grades). However, the decision for glioma grading has been made based on a visual diagnosis that is a complex task and prone to the subjectivity. Several current studies suggested the use of advanced MRI techniques alone or combined with conventional MRI modalities for the classification of glioma grades (Ryu et al., 2014, Inano et al., 2014, Zhang et al., 2017, Citak-Er et al., 2018). However, the advanced MRI techniques are costly, require special experience, and advanced MRI equipment. Therefore, developing a classification system based on conventional MRI modalities is of great interest especially for those who do not have access to advanced MRI resources. Consequently, the proposed work in this thesis was designed completely based on the conventional MRI technique.

7.2.3.2 Features Selection

Features selection is a vital task and necessary in classification design to select the most significant features and eliminate redundant features. The features selection and feature dimensionality reduction decrease storage requirements, reduce complexity and computational cost, reduces training and implementation times and thus leads to improved classification accuracy (Pantelis, 2010, Tantisatirapong, 2015, Al-Waeli, 2017).

An examples of studies that have demonstrated features selection method to enhance the classification of brain tumours would be the research conducted by Ryu et al. (2014), Hsieh et al. (2017a), Hsieh et al. (2017b) in which they have used a filter approach to select the most important features. Although the impact of using this method on the classification accuracy was not mentioned in their works clearly where the behaviour of the classification system before and after applying the selection method was not reported, this approach is effective, fast and simple to implement. However, this approach is conducted independently to the outcome of the classification accuracy and the interactions among the features have

not been considered. This can significantly influence the behaviour of the classification system. Using the better and most effective approach, the wrapper method is demonstrated to select the most important subset of features in developing a classification system for grading of glioma (Citak-Er et al., 2018). While Zacharaki et al. (2009) have combined the use of the filter and wrapper approaches, the filter method was used first and then the selected subset of features was fed to the wrapper technique to get more enhancement in classification results. Even though the influence of each method on the classification accuracy was not mentioned in their work, this approach is working effectively and can overcome the limitation explored by the filter approach through considering the outcome of the classifiers as a guide to update the search process to yield the optimal subset of features. However, they have used a wrapper method based on a single classifier where there is no guarantee if a new dataset is used that the method will generate the desired level of accuracy as a different single classifier has various responses to a different dataset. To overcome this limitation, in this thesis, the filter and wrapper method was combined and several classifiers instead of one classifier were incorporated in the wrapper strategy where the search process does not depend on only one classifier and any classifier able to capture better accuracy will be considered and take the place of a weak classifier. The other advantage of the proposed work is evaluating the impact of using the original features, filter, and the proposed method on the classification performance, which give a clearer understanding of the need to address the problem of features selection and reduction and assess the effect of the proposed solution on the classification accuracy (**Chapter 3**).

It is reported in the literature that features extracted from MR image have high correlation with each other and this can be considered a significant challenge for any classification system, which can negatively affect classification performance. It is also stated that "a good feature subset is one that contains features highly correlated with the class and uncorrelated with each other" (Hall, 1999). Therefore, the proposed HFSA has taken the merit of ranking the features according to the Pearson correlation measured among the features extracted from the MR images and incorporated the outcome from different classifiers and different subsets of features to guide the search process. This seeks for the optimal subset of features associated with the highest classification accuracy for glioma grades.

7.2.3.3   Classification

This thesis aims to develop a machine learning algorithm in the classification stage of the automated system and particularly a supervised machine learning algorithm was developed

in this work. This is due to the ability of supervised machine learning approach to create an effective classification model to predict new unseen samples within a reproducible methodology based on learning strategy. The other motivation for this is that the capability of the supervised machine learning algorithm to discover nonlinear relationships among distinctive features, which can lead to promising results in the classification accuracy of the automated classification system.

Several recent studies have developed single machine learning approach in the classification stage for glioma grading (Citak-Er et al., 2018, Zhang et al., 2017, Hsieh et al., 2017a, Zacharaki et al., 2009, Zacharaki et al., 2011, Hsieh et al., 2017c, Subashini et al., 2016, Li et al., 2006). However, each single classification approach has its limitation and there is no guarantee the same classifier can show high classification accuracy if a different dataset is involved. The fusion of multi-classification models would lead to enhance the classification accuracy and could overcome the drawbacks of weak classifiers. Thus, the proposed work in this thesis has integrated popular and efficient classification models within the development of two stages of learning to construct an efficient MCS for glioma grading.

For the development of the MCS, many common approaches used the majority vote (Bashir et al., 2016, Georgiadis et al., 2009) for the fusion of single classifiers such as the EBTree and ESDA (Kuncheva, 2014). However, the majority vote is not able to sense nonlinear relationships among classifier members. To tackle this limitation, the proposed design has used DNN to integrate the classifier members due to its capability to sense nonlinear and intricate relationships among classifiers. This method is also able to weight the classifiers, utilising them in further deep learning to seek for more improvement in the classification accuracy.

At present there is no general method to select the optimal number of neurons and layers of NNs while either the existing studies used few trials or random selection to tackle this issue. In attempting to solve this challenge, the advantage of performing many iterations for NNs was considered, which can potentially lead to the optimal convergence point, enabling the NNs to achieve the optimal classification accuracy. Consequently, a systematic process within the proposed DINN was developed by using a wide range of neurons and iterations, which lead to achieve the optimal classification accuracy and design of the proposed system. Two stages of learning in the MCS have been investigated, and DINN was incorporated to integrate multi-classification models and achieve the best design of the proposed system that can lead to optimal classification accuracy of glioma grades.

An overall comparison of the proposed MTMCS versus different current approaches including single classification and ensemble systems, which are critically evaluated in the thesis, is depicted in Table 7.3. This comparison is conducted in terms of different aspects as follows; within the number of learning stages, single classifiers system and traditional ensemble systems were conducted based on one stage of learning while the proposed system was extended to include two stages of learning. With regards to number of classifiers, single classifier and the traditional ensemble approaches are designed based on one classifier while the proposed MTMCS is developed based on the integration of multi-classification models. The previous-mentioned two aspects significantly lead to generate higher complexity in the proposed system design compared to the single classifier and traditional ensemble approaches. Consequently, the time spent for the training phase of the proposed system would be relatively longer than the single classifier and traditional ensemble approaches. The single classifier system is therefore considered the fastest approach in term of implementation time and has the least complexity of design and this motivated many current studies to develop a single classifier system in different applications. Although, some type of single classifier system can achieve the optimal classification accuracy under some controlled conditions such as selecting the most appropriate classifier or using supported dataset, it has a lower stability in classification accuracy achieved where it has high variation in the classification outcome if different data characteristics, input features, or different dataset are utilised. However, the proposed system can achieve the optimal classification accuracy with more robustness to the variation in input datasets and thus better stability, the high-level of classification accuracy is maintained compared to the single classifier system and traditional ensemble approaches. Furthermore, in comparison to the traditional ensemble approach the proposed system has used the most efficient strategy using DNN that is developed in the combiner design to integrate classifier members. DNN method can sense the intricate and nonlinear relationships among classifiers, outperforming the existing ensemble approaches that rely on the majority vote that lacks sensing nonlinear relationships. Consequently, this leads to obtaining further optimisation and more enhancement in the classification accuracy of the proposed system for glioma grades (Table 7.3).

Table 7.3 Overall comparison of different approaches investigated in this work for glioma grading

Aspect of comparisons	Single classification system	Ensemble system	Proposed MTMCS
Learning	One-stage	One -stage	Two-stages
Classifier Number	One classifier	One classifier	Multi classification model
Training time	Fast	Fast to moderate, subjected to the design and type of classifier involved.	Slow, depending on the number of input samples and optimisation parameters such as number of neurons, iterations and layers of NNs
Complexity	Low	Moderate	High
Stability in classification accuracy	Low, highly affected by input data	Low, subjected to the behaviour of single classifier used	Optimal regardless of the variation in input datasets
Combiner design	Not supported	Majority Vote	DNN

7.2.4 Comparative Studies

In this section the results obtained from the research findings by this thesis are presented and compared with the other recently published works (Table 7.4). The reader should note that while the first study is conducted on BRATS2015 dataset, the other researches are not based on the same dataset, therefore the finding listed in Table 7.4 are not directly comparable.

The first study presented in the table applied deep learning approach based on convolution NNs directly to MR images using BRATS2015 dataset, to classify glioma grades into low (I, II) against high glioma grades (III, IV) (Ye et al., 2017). They have used a 10-fold cross-validation technique in the evaluation of the classification performance. An up-sampling technique was used by replicating cases in the training phases so that the effect of an imbalanced dataset on the training phase can be eliminated. It worth noting that the variation in the results generated in that work is due to the use of random samples that may be repeated more than once in different folds (multi-subsets of data). Thus, they reported their result as the mean ± SD (Table 7.4). However, a replicating strategy was adopted to increase the number of samples for the minority class in the training phase.

Table 7.4 Comparison of the proposed system versus the previous studies

Ref.	Approach	Cross-validation	MRI Modality	Feature Extraction	Feature Selection	Classification Model	Classification types	No. of Patients	Accuracy %	Sensitivity %	Specificity %	Precision %
(Ye et al., 2017)	Deep learning upsampling	10-fold	T1c	-	-	-	Binary	Brats 2015 (274)	82.1± 7.1	88.9±6.2	57.0±16.9	89.2 ±4.7
(Anaraki et al., 2019)	Deep learning	Partition the data into training and testing	T1c	-	-	-	Multi-class	722 patients	90.9	-	-	-
(Citak-Er et al., 2018)	Single classifier system	10-fold	T1, T2, diffusion W, diffusion tensor, perfusion and spectroscopic imaging	Statistical measures from advanced MRI, mean of intensities of the MR regions	Filter + wrapper approach	SVM	Binary	34 patients I (3), II (12) III (8), IV (20)	93	96.4	86.7	92.9

Table 7.4 Continued

Ref.	Approach	Cross validation	MRI Modality	Feature Extraction	Feature Selection	Classification Model	Classification Types	No. of Patients	Accuracy %	Sensitivity %	Specificity %	Precision %
Proposed System	MCS	Leave-one-out	T2, T1 T1c, Flair	Texture-GLCM, and proposed FTD	Without features selection	Proposed MTMCS	Binary	BRATS2015 (274)	**93.80**	98.64	74.07	93.9
Proposed System	MCS	Leave-one-out	T2, T1 T1c, Flair	Texture-GLCM, and proposed FTD	Without features selection	Proposed MTMCS	Binary	BRATS2018 (285)	**95.09**	98.10	86.67	95.37
Proposed System	MCS	Leave-one-out	T2	Texture-GLCM	Without features selection	Proposed MTMCS	Binary	BRATS2013 (30)	**100**	100	100	100
Proposed System	MCS	Leave-one-out	T2	Texture-GLCM	Without feature Selection	Proposed MTMCS	Binary	Cancer Data (30)	**100**	100	100	100
Proposed System	MCS	Leave-one-out	T2	Texture-GLCM	Without Feature Selection	Proposed MTMCS	Multi-class	Cancer Data (30)	**100**	100	100	100

This results in a model with relatively high classification accuracy on current data, and with undesired classification results can be obtained on a new dataset, in addition to the fact that the risk of the over-fitting problem could not be avoided by cross-validation process (Zhang et al., 2017).

The proposed system in this thesis achieved effective results using four benchmark datasets, which obtained optimal classification accuracy of 100% for both the BRATS2013 and Cancer datasets. While it yields better results of 95.09% using BRATS2018 and it achieves 93.80% using BRATS2015, which is higher than the performance of deep learning approach suggested by Ye et al. (2017) based on the BRATS2015 dataset where their work has achieved 82.1% $\pm$7.1 classification accuracy for glioma grades.

Similarly, Anaraki et al. (2019) have developed a deep learning approach for the multi-class classification of brain tumour types and grades based on T1c-weighted MR modality. This approach was conducted based on convolution NNs and applied directly to the MR images using 722 patients. The classification accuracy of multi-class classification for the three grades of glioma (II, III, and IV) is 90.9%. In the same manner, Zacharaki et al. (2009) achieved 62.50% to classify glioma into three grades (II, III, IV) (Table 2.2). However, the system proposed by the current thesis has achieved 100% full classification results in the classification of the WHO three grades of glioma (II, III, and IV), outperforming the classification result of other recent work.

A further critical comparison of the proposed system against the traditional deep learning approach such as CNN is discussed in this section. The deep learning approach, and due to its advantages and the availability of large datasets in different fields, is used in different applications **(section 2.13.2)**. However, the requirement to large dataset in the training phase is the most challenging issue in achieving an effective classification model and hence the classification performance of this approach is affected, while using small samples in the training phase can lead to a low performance if new samples are tested (Hegde et al., 2019, Mazurowski et al., 2019). In general, the availability of datasets of medical images is a significant challenge (Mohsen et al., 2018) and especially to work undertaken in this thesis as not any MRI dataset is possible to consider unless there is a confirmation of the histopathology diagnosis. To elaborate, recently and particularly on medical dataset, the deep learning approach becomes popular and can achieve better result for medical dataset in segmentation systems (Havaei et al., 2017, Rodríguez Colmeiro et al., 2017). It is less popular in decision-making systems particularly for MRI based automated system for glioma

grading and this is due to the mechanism of how the approach works and how it deals with an image. In more details, for segmentation systems, the deep learning approach is dealing directly with image pixels (Kamnitsas et al., 2017, Işın et al., 2016). Furthermore, with the availability of very large number of images for each patient and each image having a large number of pixels, and since each pixel represents a sample, this means that there is a vast number of samples that can be utilised in the training phase to build the model and thereby supporting the model to classify the image pixels efficiently. In particular applications where there are possibilities to provide a vast number of samples, breakthrough advances were achieved recently by the deep learning approach e.g., more than one million labelled images in ImageNet (Russakovsky et al., 2015, Shen et al., 2017), more than ten million annotated image developed by Google Corporation (Le, 2013). On the other hand, in the decision-making system, the decision has to be made to classify patients and not the image pixels or image regions. Consequently, a patient and regardless of how many numbers of images or pixels, is represented by one sample in the features space, hence this limit the number of samples utilised to train the model as in practice the number of samples is reflected by the number of patients in the training phase of the decision-making system (Citak-Er et al., 2018). Accordingly it is not feasible to efficiently train the classification model of the deep learning approach on a limited number of patients (Mazurowski et al., 2019).

The proposed classification system based on MTMCS overcome this limitation through developing two stages of learning and utilising several common classifiers in this first stage where these classifiers have the ability to handle small and large datasets (Kuncheva, 2014, Jun and Jian, 2005, Feng et al., 2014). In addition, CNN is computationally expensive (Caicedo et al., 2019) and requires tuning huge number of parameters that need optimization, which increases risk of overfitting that lead to a low performance particularly when training the model using a small dataset (Mazurowski et al., 2019, Zhou et al., 2014) . The developed system in this thesis utilized fixed and stable parameters for the classifiers involved in the first stage of the developed classification system. The developed system also adapts imbalanced datasets through the complementary ensemble strategy of different single classifiers (Kuncheva, 2014) whereby the experimental results showed stable and superior classification results when evaluated using different datasets. However, the outcome of deep learning approach is highly affected by the class imbalanced problem that is common in medical dataset. Furthermore, it is necessary for deep learning application to demonstrate an

additional technique to cope with this problem (Buda et al., 2018) which increases the chance of overfitting that leads to low performance in the testing phase (Zhang et al., 2017).

The application of CNN is recognised as an end to end approach, which is applied directly to image pixels with ignoring the advantages of the features extraction and selection. While these stages provide significant support to the classification model, leading to better quality in the classification performance in terms of improved classification accuracy and efficient system design (El-Dahshan et al., 2014). In current thesis, smart, significant and new features were extracted, as well as the vital features and classifiers were selected, which they are investigated and developed to improve the performance of the automated classification system for glioma grades. Furthermore, the complementary of the new methods with the two stages of learning optimises the classification performance of the developed system.

Accordingly, it can draw the inference that the proposed approach, which is developed and comprehensively evaluated and validated using four benchmark datasets, is more suitable to the classification problem undertaken in this work than the deep learning approach such as CNN, leading to more effective accuracy for the classification of glioma grades.

Many studies have developed a classification system for glioma grades based on a single classifier system due to its efficiency and low complexity required in developing procedures of the classification system design. These are explored in details in (Table 2.2). However, single classifier system has been proven experimentally for lacked stability in the classification accuracy as it has various sensitivities to diverse datasets and can behave differently if tested with a different dataset. Furthermore, most of the recent works demonstrated the evaluation of the classification model using only one dataset. While using different datasets can provide a comprehensive evaluation of system behaviour. Nonetheless, the proposed system in this thesis overcomes this limitation by integrating different methods based on two stages of learning and by combining effective classification models incorporating with DNN to optimise the classification accuracy of glioma grades. Also, the developed system has been evaluated using four different datasets. Accordingly, the developed system achieved higher classification accuracy with optimal results at 100% with two datasets and superior results at 93.80% and 95.09 for the BRATS2015 and BRATS2018 dataset respectively.

7.3 Conclusion

The inconsistency of many morphological behaviours and tumour descriptors of different glioma grades leads to high complexity in glioma grading. Additionally, an increase in the malignancy growth from lower tumour grade to the higher grade is associated with high mortality. It is a critical challenge to achieve an accurate diagnosis of glioma grade, through the visual diagnosis that highly depends on the extent of experience. Low and high glioma grades follow different treatment protocol. Low glioma grades, which includes grade I, and grade II could follow an extensive resection of lesion and may need radiotherapy postoperatively (Pouratian and Schiff, 2010). However the high-grade gliomas, which include grade III and grade IV, are managed with an essential resection, chemotherapy and radiotherapy (Stupp et al., 2010). It is necessary to provide patients with the proper clinical treatment, prolonging survival and quality of life, and thus it creates the need for accurate tumour characterisation (Pantelis, 2010, Chao et al., 2006). Consequently, to overcome this challenge, this study proposes a novel automated system for classification of glioma grades, which is based on quantitative features extracted from MR images incorporated with two stages of learning. In the first stage, different popular machine learning algorithms are trained and evaluated, in the second stage; the multi-classification models are integrated based on deep neural networks.

This work has many contributions, mainly the classification approach itself; the quantitative features extracted from the MR images, which are utilised in the integration of multi-machine learning algorithms within two stages of learning (meta-learning). This is performed based on deep neural network that integrates multi-classification models to provide an automated, efficacious, robust classification approach for glioma grades, which is comparable to the state-of-the-art.

The proposed classification system has achieved an optimal classification accuracy in the classification of different WHO grades of glioma, where it has obtained the full discrimination rate of 100% in the differentiation between low-grade glioma (I, II) and high-grade glioma (III, IV) using BRATS2013 dataset. Similarly, the full accuracy rate of 100% was also achieved in the discrimination between grade IV and the lower grades (II, III), and the separation between grade II and the high-grade glioma (III, IV) using the Cancer dataset. The classification accuracy achieved was 93.8% in discriminating between low-grade glioma (I, II) and high-grade glioma (III, IV) using the BRATS2015 dataset. The proposed system also achieved the full classification rate at 100% in the multi-class classification of

glioma grades into the three WHO grades (II, III, and IV). Furthermore, the proposed system has been compared with other current approaches used widely in the classification of medical images as well as the comparison made against recent approaches developed for the classification of brain tumour types and grades. The results of the comparison confirmed that the proposed system has outperformed the other recent developed systems and has achieved optimal and robust classification accuracy for glioma grades.

The proposed system overcomes the potential effect of using a limited size of samples that can cause deficiency in the training phase, and it showed superior results and even with small sample sizes. This is due to the integration of different robust classifiers that has such advantages in the ensemble stage.

The proposed system has many strengths, and advantageous aspects, which were listed as follows:

> **Systematic:** the proposed system is systematic approach developed through predefined sequence of actions that are implemented in order irrespective of the data set. Consequently, it is a standardised framework that is not dataset-dependent.

> **Automatic**: the developed system is fully automated and does not need any intervention by an expert.

> **Robustness:** unlike many single classification approaches which are sensitive to the data distribution giving no guarantee of a model with high accuracy showing significant classification results if a different dataset is involved. The developed classification system proposed in this research is less affected by changing the dataset showing high stability in classification accuracy for the entire four benchmark datasets. Similarly, the proposed system outperforms the traditional ensemble approaches where the latter approach is highly dependent on the majority vote technique in the integration design where the majority vote is not able to sense the nonlinear relationships among classifiers. While the proposed system develops DNN instead that can sense more intricate relationships among classifiers, leading to more effective classification accuracy for glioma grades.

> **Selection best subset of classifiers:** one of the major challenging aspects in the design of MCS is how to choose the best classifier members that can result in the best results. While in this work, an efficient algorithm is proposed to tackle this challenge and select the best subset of classifiers from many different classifiers where the suggested algorithm is also beneficial in reducing the dimensions of

classifiers, which is consequently crucial for complexity reduction of the system design based on MCS.

The developed automated classification system for glioma grade will offer great help as a second opinion to support radiologist and experts in the assessment of the malignancy grades of glioma and produce an accurate, automated and objective decision generated by prediction model using highly technological methods, quantitatively designed based on a machine learning algorithm.

7.4 Research Recommendations and Future Trends

This thesis uses four different benchmark datasets publicly available for academic use, which have different acquisition settings such as various slice thickness, slice space, strength of scanner magnetic field. However, these datasets are post acquired dataset that means this work has no control over the dataset setting. While the protocol used to acquire the image dataset such as the strength of scanner magnetic field and slice thickness have a significant influence on the image resolution and thus image texture features can be affected and hence this impact the classification performance (Savio et al., 2010, Tantisatirapong, 2015). It is recommended, for example, to use a high strength magnetic field by MRI to achieve better image texture details. Future work could be conducted by controlling these image acquisition protocols and examine the effect of different ranges of the acquisition factors and demonstrate an optimisation study through experimentally finding the optimal acquisition setting that achieves a best classification of glioma grades.

Several image features for brain tumour are investigated, particularly, features associated with tumour heterogeneity, which are widely used in many different medical applications such as segmentation, and classification of brain tumours. These features were investigated and showed successful results in this study. The challenge with these features was the homogeneity of the tumour image, which can play a significant role in the amount of texture that can be recognised and thereby affecting the classification accuracy results. The inference was drawn from an image sample that the low level of homogeneity can support the texture recognition, which consequently leads to enhancing the ability of the classification system for more improved classification accuracy of glioma grades. Further studies are needed to identify the exact level of homogeneity acceptable to allow best classification accuracy with a classification system. This is beyond the scope of this study, and out of our control as all samples that are undertaken in this study are post-acquired dataset, where this study does not

involve in the creation of this dataset and all samples are provided from a public dataset. Future work can be directed by changing the level of homogeneity of image samples and track the change in classification accuracy, which will be useful to determine the tolerance range of homogeneity that can support the classification model for best and optimal classification accuracy for glioma grades based on image texture features. A suggestion would be useful that if modification criteria is used to the brain tumour images, through changing the image homogeneity, and then the effect on classification accuracy of glioma grades can be pursued and investigated.

The automated classification system was developed based on several brain tumour descriptors and image features extracted from MR images while there are many other tumour descriptors and image features can be used. Future trend could involve the combination of the image features undertaken in this thesis with other tumour descriptors or features such as tumour vascularity and cellularity and examine the impact of these features on the glioma grading which could lead to further improvement in the classification accuracy of glioma grades.

This work concentrated on the classification of different WHO glioma grades while it is possible as a future trend to design a classification system for glioma grades and types, which is also a valuable target, as each tumour type requires different treatment and prognosis.

The classification of glioma grades followed the WHO scheme where the glioma can be classified into four grades (I, II, III, and IV). This work determined the classification of four grades through the binary classification between low grades (I, II) and high grades (III, IV). However, in this research work, the multi-class classification task was accomplished for only the three grades (II, III, and IV). This is due to the limited availability of datasets that supported four separated grades with the histopathology confirmation. It is also noted that many recent studies commonly perform multi-class classification of glioma grades considering only the three grade of glioma (II, III, and IV) (Zacharaki et al., 2009, Anaraki et al., 2019, Zhang et al., 2017, Ryu et al., 2014, Sajjad et al., 2019). This is probably due to the limited availability of the data with glioma grade I, which is usually cured compared to other glioma grades (Moore and Kim, 2010) which reduces the number of samples availability for grade I and also reduces the motivation to consider it in the classification system. Future trend could consider the four WHO grades of glioma. However, this could be a big challenge as limited samples of grade I will not be sufficient for those who develop automated systems based on a machine learning strategy.

As a future trend, larger samples are needed whereby it is possible to use deep learning if sufficient data and advanced hardware are available. Future work could demonstrate different approach of deep learning in attempting to achieve the most optimal classification accuracy. It is also recommended to take into account the merit of feature extraction methods for brain tumour images.

This work focused on the integration of different popular and efficient classification models in terms of successful achievement in accuracy and handling small sample data. However, there are many alternative classifiers can be examined and developed. Future study could propose a new methodology through utilising and integrating new classifiers and investigating the impact of the developed method on the optimisation of the classification accuracy. Since this research work developed a new method to select the best subset of classifiers, this could help and guide the future research direction to answer the question regarding the selection of best classification model that will be the best choice for more optimal classification results.

The proposed system was quantitatively developed and evaluated based on comprehensive experimental design. The aim of this work is to classify different WHO glioma grades within an automated computer-based analysis. Future work could be conducted to study the effect of the developed system on improving the diagnosis performance of group of radiologists within a clinical environment. The impact of the proposed system on the diagnosis performance of the experts can be analysed for glioma grading.

The proposed system has been proven to be an efficacious classification approach for both binary classification and multi-class classification. However, the complexity of system design is the critical challenge, which requires considerable implementation time for training and optimisation where the parallel processing and advanced hardware is highly recommended. The proposed system is not dataset-dependent so as a Future work it can be applied and developed in a different application or dataset for possible enhancement and achieves an optimal classification result.

REFERENCES

Abdallah, I., Dertimanis, V., Mylonas, H., Tatsis, K., Chatzi, E., Dervilis, N., Worden, K. & Maguire, E. Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data. Proceedings of ESREL, 2018 Trondheim, Norway. Taylor & Francis Group, 3053-3061.

Abdel-Hamid, O., Mohamed, A., Jiang, H. & Penn, G. Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 25-30 March 2012 2012. 4277-4280.

Al-Waeli, A. M. H. 2017. *An Automated System for the Classification and Segmentation of Brain Tumours in Mri Images Based on the Modified Grey Level Co-Occurrence Matrix.* PhD, University of Salford, UK.

Al-Zurfi, A. N., Meziane, F. & Aspin, R. A Computer-aided Diagnosis System for Glioma Grading using Three Dimensional Texture Analysis and Machine Learning in MRI Brain Tumour. 2019 IEEE 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART), 2019 Paris, France. IEEE.

Aly, M. 2005. Survey on multiclass classification methods. *Neural Networks,* 19(1)**,** 1-9.

AlZurfi, A., Meziane, F. & Aspin, R. Automated glioma grading based on an efficient ensemble design of a multiple classifier system using deep iteration neural networks matrix. Proceedings of the 24th International Conference on Automation and Computing (ICAC'2018), 2018 Newcastle University, Newcastle upon Tyne, UK. IEEE.

Ananda Resmi, S. 2013. *Development of Techniques for the Automatic Extraction and Grade Detection of Glioma Tumors from Conventional Brain Magnetic Resonant Images.* PhD, Cochin University of Science and Technology, INDIA.

Ananda Resmi, S. & Thomas, T. 2010. Texture Description of low grade and high grade Glioma using Statistical features in Brain MRIs. *International Journal of Recent Trends in Engineering and Technology,* 4(3)**,** 27-33.

Anaraki, A. K., Ayati, M. & Kazemi, F. 2019. Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybernetics and Biomedical Engineering,* 39(1)**,** 63-74.

Aragao, M. d. F. V., Law, M., de Almeida, D. B., Fatterpekar, G., Delman, B., Bader, A., Pelaez, M., Fowkes, M., de Mello, R. V. & Valenca, M. M. 2014. Comparison of perfusion, diffusion, and MR spectroscopy between low-grade enhancing pilocytic astrocytomas and high-grade astrocytomas. *American Journal of Neuroradiology,* 35(8)**,** 1495-1502.

Armano, G., Chira, C. & Hatami, N. A New Gene Selection Method Based on Random Subspace Ensemble for Microarray Cancer Classification. 6th IAPR International Conference of Pattern Recognition in Bioinformatics., 2011 Berlin, Heidelberg. Springer Berlin Heidelberg, 191-201.

Ashour, A. S., Guo, Y., Hawas, A. R. & Xu, G. 2018a. Ensemble of subspace discriminant classifiers for schistosomal liver fibrosis staging in mice microscopic images. *Health information science and systems,* 6(1)**,** 21.

Ashour, D. S., Abou Rayia, D. M., Maher Ata, M., Ashour, A. S. & Abd Elnaby, M. M. 2018b. Hybrid feature extraction techniques for microscopic hepatic fibrosis classification. *Microscopy Research and Technique,* 81(3)**,** 338-347.

Baaziz, N., Abahmane, O. & Missaoui, R. 2010. Texture feature extraction in the spatial-frequency domain for content-based image retrieval. *Computer Vision and pattern Recognition. arXiv:1012.5208.*

Baboo, S. S. & Sasikala, M. S. 2010. Multicategory classification using support vector machine for microarray gene expression cancer diagnosis. *Global Journal of Computer Science and Technology,* 10(15)**,** 38-44.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K. & Davatzikos, C. 2017. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data,* 4(1)**,** 170117-170117.

Baptista, F. D., Rodrigues, S. & Morgado-Dias, F. Performance comparison of ANN training algorithms for classification. Intelligent Signal Processing (WISP), 2013 IEEE 8th International Symposium on, 2013 Funchal, Portugal. IEEE, 115-120.

Barker, F. G., Davis, R. L., Chang, S. M. & Prados, M. D. 1996. Necrosis as a prognostic factor in glioblastoma multiforme. *Cancer,* 77(6)**,** 1161-1166.

Barnett, G. H. 2007. *High-grade gliomas: diagnosis and treatment,* Berlin, Springer.

Bashir, S., Qamar, U. & Khan, F. H. 2016. A multicriteria weighted vote-based classifier ensemble for heart disease prediction. *Computational Intelligence,* 32(4)**,** 615-645.

Bauer, S., Wiest, R., Nolte, L. P. & Reyes, M. 2013. A survey of MRI-based medical image analysis for brain tumor studies. *Physics in Medicine and Biology,* 58(13)**,** R97-R128.

Behin, A., Hoang-Xuan, K., Carpentier, A. F. & Delattre, J.-Y. 2003. Primary brain tumours in adults. *The Lancet,* 361(9354)**,** 323-331.

Bertoni, A., Folgieri, R. & Valentini, G. Random subspace ensembles for the bio-molecular diagnosis of tumors. NETTAB 2004 (Fourth International Workshop on Network Tools And Application in Biology), 2004 Italian.

Bertoni, A., Folgieri, R. & Valentini, G. 2005. Feature selection combined with random subspace ensemble for gene expression based diagnosis of malignancies. *In:* APOLLONI, B., MARINARO, M. & TAGLIAFERRI, R. (eds.) *Biological and Artificial Intelligence Environments:15th Italian Workshop on Neural Nets, WIRN VIETRI 2004.* Netherlands: Springer.

Birry, R. A. K. 2013. *Automated classification in digital images of osteogenic differentiated stem cells.* PhD, University of Salford.

Blink, E. J. 2004. mri: Physics. *Online PDF file* [Online]. [Accessed 2016.http://www.mri-physics.net/].

Bloch, F. 1946. Nuclear induction. *Physical review,* 70(7-8)**,** 460.

Bonilha, L., Kobayashi, E., Castellano, G., Coelho, G., Tinois, E., Cendes, F. & Li, L. M. 2003. Texture analysis of hippocampal sclerosis. *Epilepsia,* 44(12)**,** 1546-1550.

Bracewell, R. N. 2000. *The Fourier transform and its applications,* Boston, Mass. [u.a.], McGraw Hill.

Breiman, L. 1996. Bagging predictors. *Machine Learning,* 24(2)**,** 123-140.

Buda, M., Maki, A. & Mazurowski, M. A. 2018. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks,* 106249-259.

Burns, R. P. & Burns, R. 2008. *Business research methods and statistics using SPSS,* London, Sage.

Caicedo, J. C., Roth, J., Goodman, A., Becker, T., Karhohs, K. W., Broisin, M., Csaba, M., McQuin, C., Singh, S., Theis, F. & Carpenter, A. E. 2019. Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images. *bioRxiv,* 335216.

REFERENCES

Camargo, L. S. & Yoneyama, T. 2001. Specification of Training Sets and the Number of Hidden Neurons for Multilayer Perceptrons. *Neural Computation,* 13(12)**,** 2673-2680.

Carter, J. S., Koopmeiners, J. S., Kuehn-Hajder, J. E., Metzger, G. J., Lakkadi, N., Downs, L. S. & Bolan, P. J. 2013. Quantitative multiparametric MRI of ovarian cancer. *Journal of Magnetic Resonance Imaging,* 38(6)**,** 1501-1509.

Castellano, G., Bonilha, L., Li, L. M. & Cendes, F. 2004. Texture analysis of medical images. *Clinical Radiology,* 59(12)**,** 1061-1069.

Chakraborty, S., Paul, S., Sarkar, R. & Nasipuri, M. Feature Map Reduction in CNN for Handwritten Digit Recognition. 2019 Singapore. Springer Singapore, 143-148.

Chan, Y. 2003. Biostatistics 104: correlational analysis. *Singapore Med J,* 44(12)**,** 614-9.

Chao, S. T., Barnett, G. H., Liu, S. W., Reuther, A. M., Toms, S. A., Vogelbaum, M. A., Videtic, G. M. M. & Suh, J. H. 2006. Five-year survivors of brain metastases: A single-institution report of 32 patients. *International Journal of Radiation Oncology Biology Physics,* 66(3)**,** 801-809.

Chen, G. H. & Shah, D. 2018. Explaining the Success of Nearest Neighbor Methods in Prediction. *Foundations and Trends® in Machine Learning,* 10(5-6)**,** 337-588.

Chen, W.-S., Huang, R.-H. & Hsieh, L. 2009. Iris Recognition Using 3D Co-occurrence Matrix. *In:* TISTARELLI, M. & NIXON, M. S. (eds.) *Advances in Biometrics: Third International Conference, ICB 2009, Alghero, Italy, June 2-5, 2009. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Chen, W., Giger, M. L., Li, H., Bick, U. & Newstead, G. M. 2007. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magnetic Resonance in Medicine,* 58(3)**,** 562-571.

Chen, Y., Crawford, M. M. & Ghosh, J. Integrating support vector machines in a hierarchical output space decomposition framework. Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International, 2004 Anchorage, AK, USA. IEEE, 949-952.

Chevrefils, C., Périé, D., Parent, S. & Cheriet, F. 2018. To distinguish flexible and rigid lumbar curve from MRI texture analysis in adolescent idiopathic scoliosis: A feasibility study. *Journal of Magnetic Resonance Imaging,* 48(1)**,** 178-187.

Chow, K. L., Gobin, Y. P., Cloughesy, T., Sayre, J. W., Villablanca, J. P. & Viñuela, F. 2000. Prognostic factors in recurrent glioblastoma multiforme and anaplastic astrocytoma treated with selective intra-arterial chemotherapy. *American journal of neuroradiology,* 21(3)**,** 471-478.

Citak-Er, F., Firat, Z., Kovanlikaya, I., Ture, U. & Ozturk-Isik, E. 2018. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T. *Computers in Biology and Medicine,* 99(1)**,** 154-160.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L. & Prior, F. 2013. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging,* 26(6)**,** 1045-1057.

Corso, J. J., Sharon, E., Dube, S., El-Saden, S., Sinha, U. & Yuille, A. 2008. Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification. *IEEE Transactions on Medical Imaging,* 27(5)**,** 629-640.

Cover, T. & Hart, P. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory,* 13(1)**,** 21-27.

Dara, S. & Tumma, P. Feature Extraction By Using Deep Learning: A Survey. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018 Coimbatore, India. 1795-1801.

REFERENCES

Das, D., Mahanta, L. B., Ahmed, S., Baishya, B. K. & Haque, I. 2018. Study on Contribution of Biological Interpretable and Computer-Aided Features Towards the Classification of Childhood Medulloblastoma Cells. *Journal of Medical Systems,* 42(8)**,** 151.

Davnall, F., Yip, C. S., Ljungqvist, G., Selmi, M., Ng, F., Sanghera, B., Ganeshan, B., Miles, K. A., Cook, G. J. & Goh, V. 2012. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice. *Insights Imaging,* 3(6)**,** 573-89.

DeAngelis, L. M. 2001. Brain tumors. *New England Journal of Medicine,* 344(2)**,** 114-123.

Deepa, S. & Devi, B. A. 2011. A survey on artificial intelligence approaches for medical image classification. *Indian Journal of Science and Technology,* 4(11)**,** 1583-1595.

Dennie, C., Thornhill, R., Sethi-Virmani, V., Souza, C. A., Bayanati, H., Gupta, A. & Maziak, D. 2016. Role of quantitative computed tomography texture analysis in the differentiation of primary lung cancer and granulomatous nodules. *Quantitative imaging in medicine and surgery,* 6(1)**,** 6-15.

Devos, A., Simonetti, A., Van Der Graaf, M., Lukas, L., Suykens, J., Vanhamme, L., Buydens, L., Heerschap, A. & Van Huffel, S. 2005. The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification. *Journal of Magnetic Resonance,* 173(2)**,** 218-228.

Dong, H., Yang, G., Liu, F., Mo, Y. & Guo, Y. 2017. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. *In:* VALDÉS HERNÁNDEZ, M. & GONZÁLEZ-CASTRO, V. (eds.) *Medical Image Understanding and Analysis. MIUA 2017. Communications in Computer and Information Science.* Cham: Springer.

Doolittle, N. D. 2004. State of the science in brain tumor classification. *Seminars in Oncology Nursing,* 20(4)**,** 224-230.

Dougherty, G. 2009. *Digital image processing for medical applications,* UK, Cambridge University Press.

Drabycz, S., Roldán, G., De Robles, P., Adler, D., McIntyre, J. B., Magliocco, A. M., Cairncross, J. G. & Mitchell, J. R. 2010. An analysis of image texture, tumor location, and MGMT promoter methylation in glioblastoma using magnetic resonance imaging. *Neuroimage,* 49(2)**,** 1398-1405.

Dubitzky, W., Granzow, M. & Berrar, D. P. 2007. *Fundamentals of data mining in genomics and proteomics,* New york, USA, Springer Science & Business Media.

Duda, R. O., Hart, P. E. & Stork, D. G. 2012. *Pattern classification,* New York, John Wiley & Sons.

Dvořák, P., Kropatsch, W. & Bartušek, K. 2013. Automatic brain tumor detection in T2-weighted magnetic resonance images. *Measurement Science Review,* 13(5)**,** 223-230.

El-Dahshan, E.-S. A., Mohsen, H. M., Revett, K. & Salem, A.-B. M. 2014. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert Systems with Applications,* 41(11)**,** 5526-5545.

Eliat, P.-A., Olivié, D., Saïkali, S., Carsin, B., Saint-Jalmes, H. & de Certaines, J. D. 2012. Can dynamic contrast-enhanced magnetic resonance imaging combined with texture analysis differentiate malignant glioneuronal tumors from other glioblastoma? *Neurology research international,* Article ID 195176.

Erickson, B. J., Korfiatis, P., Akkus, Z. & Kline, T. L. 2017. Machine learning for medical imaging. *Radiographics,* 37(2)**,** 505-515.

Feng, W., Zhang, Q., Hu, G. & Huang, J. X. 2014. Mining network data for intrusion detection through combining SVMs with ant colony networks. *Future Generation Computer Systems,* 37(1)**,** 127-140.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. & Bray, F. 2015. Cancer incidence and mortality worldwide: Sources,

methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer,* 136(5)**,** E359-E386.

Ferreira, J. E. V., Pinheiro, M. T. S., dos Santos, W. R. S. & da Silva Maia, R. 2016. Graphical representation of chemical periodicity of main elements through boxplot. *Educación química,* 27(3)**,** 209-216.

Fetit, A. E., Novak, J., Peet, A. C. & Arvanitis, T. N. 2015. Three-dimensional textural features of conventional MRI improve diagnostic classification of childhood brain tumours. *NMR in Biomedicine,* 28(9)**,** 1174-1184.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics,* 7(2)**,** 179-188.

Florez, E., Nichols, T., Lirette, S., Howard, C. & Fatemi, A. 2018. Developing a Texture Analysis Technique using Fluid-Attenuated Inversion Recovery (FLAIR) to Differentiate Tumor from Edema for Contouring Primary Intracranial Tumors. *SM Journal of Clinical and Medical Imaging,* 4(2)**,** 1023.

Freund, Y. 1995. Boosting a weak learning algorithm by majority. *Information and computation,* 121(2)**,** 256-285.

Friedl, M. A. & Brodley, C. E. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment,* 61(3)**,** 399-409.

Fujita, H., Uchiyama, Y., Nakagawa, T., Fukuoka, D., Hatanaka, Y., Hara, T., Lee, G. N., Hayashi, Y., Ikedo, Y., Gao, X. & Zhou, X. 2008. Computer-aided diagnosis: The emerging of three CAD systems induced by Japanese health care needs. *Computer Methods and Programs in Biomedicine,* 92(3)**,** 238-248.

Ganeshan, B., Abaleke, S., Young, R., Chatwin, C. R. & Miles, K. A. 2010. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer imaging,* 10(1)**,** 137-143.

Ganeshan, B., Skogen, K., Pressney, I., Coutroubis, D. & Miles, K. 2012. Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: Preliminary evidence of an association with tumour metabolism, stage, and survival. *Clinical Radiology,* 67(2)**,** 157-164.

Geneidi, E. A. S., Habib, L. A., Chalabi, N. A. & Haschim, M. H. 2015. Potential role of quantitative MRI assessment in differentiating high from low-grade gliomas. *The Egyptian Journal of Radiology and Nuclear Medicine,* 47(1).

Georgiadis, P., Cavouras, D., Kalatzis, I., Glotsos, D., Athanasiadis, E., Kostopoulos, S., Sifaki, K., Malamas, M., Nikiforidis, G. & Solomou, E. 2009. Enhancing the discrimination accuracy between metastases, gliomas and meningiomas on brain MRI by volumetric textural features and ensemble pattern recognition methods. *Magnetic Resonance Imaging,* 27(1)**,** 120-130.

Giacinto, G., Roli, F. & Fumera, G. Design of effective multiple classifier systems by clustering of classifiers. Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, 2000 Barcelona, Spain. IEEE, 160-163.vol 2.

Gibbs, P. & Turnbull, L. W. 2003. Textural analysis of contrast-enhanced MR images of the breast. *Magnetic Resonance in Medicine,* 50(1)**,** 92-98.

Gnep, K., Fargeas, A., Gutiérrez-Carvajal, R. E., Commandeur, F., Mathieu, R., Ospina, J. D., Rolland, Y., Rohou, T., Vincendeau, S. & Hatt, M. 2017. Haralick textural features on T2-weighted MRI are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer. *Journal of Magnetic Resonance Imaging,* 45(1)**,** 103-117.

REFERENCES

Gómez, W., Pereira, W. & Infantosi, A. F. C. 2012. Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. *IEEE Transactions on Medical Imaging,* 31(10)**,** 1889-1899.

Graña, M., Termenon, M., Savio, A., Gonzalez-Pinto, A., Echeveste, J., Pérez, J. M. & Besga, A. 2011. Computer Aided Diagnosis system for Alzheimer Disease using brain Diffusion Tensor Imaging features selected by Pearson's correlation. *Neuroscience Letters,* 502(3)**,** 225-229.

Grant, L. A. & Griffin, N. 2013. *Grainger & Allison's Diagnostic Radiology Essentials,* New York, USA, Elsevier Health Sciences.

Graupe, D. 2013. *Principles of artificial neural networks,* Singapore, World Scientific.

Gu, Q., Zhu, L. & Cai, Z. 2009. Evaluation Measures of the Classification Performance of Imbalanced Data Sets. *In:* CAI, Z., LI, Z., KANG, Z. & LIU, Y. (eds.) *Computational Intelligence and Intelligent Systems, vol 51.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Gupta, M., Rajagopalan, V., Pioro, E. P. & Rao, B. V. V. S. N. P. 2017. Volumetric analysis of MR images for glioma classification and their effect on brain tissues. *Signal, Image and Video Processing,* 11(7)**,** 1337-1345.

Gupta, N., Bhatele, P. & Khanna, P. 2019. Glioma detection on brain MRIs using texture and morphological features with ensemble learning. *Biomedical Signal Processing and Control,* 47(1)**,** 115-125.

Guyon, I. & Elisseeff, A. 2003. An introduction to variable and feature selection. *Journal of machine learning research,* 31157-1182.

Hadziahmetovic, M., Shirai, K. & Chakravarti, A. 2011. Recent advancements in multimodality treatment of gliomas. *Future Oncology,* 7(10)**,** 1169-1183.

Hall, M. A. 1999. *Correlation-based feature selection for machine learning.* PhD, University of Waikato, Hamilton, NewZealand.

Hamel, L. H. 2009. *Knowledge discovery with support vector machines,* Hoboken, New Jersey., John Wiley & Sons.

Han, Y., Kim, J., Lee, K., Han, Y., Kim, J. & Lee, K. 2017. Deep Convolutional Neural Networks for Predominant Instrument Recognition in Polyphonic Music. *IEEE/ACM Transaction. Audio, Speech and Lang. Proc.,* 25(1)**,** 208-221.

Haralick, R. M., Shanmugam, K. & Dinstein, I. 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics,* 3(6)**,** 610-621.

Hasan, A., Meziane, F., Aspin, R. & Jalab, H. 2016a. Segmentation of Brain Tumors in MRI Images Using Three-Dimensional Active Contour without Edge. *Symmetry,* 8(11)**,** 132.

Hasan, A. M. & Meziane, F. 2016. Automated screening of MRI brain scanning using grey level statistics. *Computers & Electrical Engineering,* 53(1)**,** 276-291.

Hasan, A. M., Meziane, F. & Jalab, H. A. Performance of grey level statistic features versus Gabor wavelet for screening MRI brain tumors: A comparative study. Information Communication and Management (ICICM), International Conference on, 2016b Hatfield, UK. IEEE, 136-140.

Hasan, A. M., Meziane, F. & Kadhim, M. A. Automated Segmentation of Tumours in MRI Brain Scans. Proceeding BIOSTEC 2016 Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologie, 2016c Rome, Italy. 55-62.

Hastie, T., Tibshirani, R. & Friedman, J. 2001. *The elements of statistical learning. 2001,* USA, Springer.

REFERENCES

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M. & Larochelle, H. 2017. Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis,* 3518-31.

Hegde, R. B., Prasad, K., Hebbar, H. & Singh, B. M. K. 2019. Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. *Biocybernetics and Biomedical Engineering,* 39(2)**,** 382-392.

Helbren, E., Fanshawe, T. R., Phillips, P., Mallett, S., Boone, D., Gale, A., Altman, D. G., Taylor, S. A., Manning, D. & Halligan, S. 2015. The effect of computer-aided detection markers on visual search and reader performance during concurrent reading of CT colonography. *European Radiology,* 25(6)**,** 1570-1578.

Herlidou, S., Rolland, Y., Bansard, J. Y., Le Rumeur, E. & de Certaines, J. D. 1999. Comparison of automated and visual texture analysis in MRI: Characterization of normal and diseased skeletal muscle. *Magnetic Resonance Imaging,* 17(9)**,** 1393-1397.

Holli, K. K., Harrison, L., Dastidar, P., Wäljas, M., Liimatainen, S., Luukkaala, T., Öhman, J., Soimakallio, S. & Eskola, H. 2010. Texture analysis of MR images of patients with mild traumatic brain injury. *BMC Medical Imaging,* 10(1)**,** 1.

Hsieh, K. L.-C., Chen, C.-Y. & Lo, C.-M. 2017a. Quantitative glioma grading using transformed gray-scale invariant textures of MRI. *Computers in Biology and Medicine,* 83(1)**,** 102-108.

Hsieh, K. L.-C., Lo, C.-M. & Hsiao, C.-J. 2017b. Computer-aided grading of gliomas based on local and global MRI features. *Computer Methods and Programs in Biomedicine,* 139(1)**,** 31-38.

Hsieh, K. L.-C., Tsai, R.-J., Teng, Y.-C. & Lo, C.-M. 2017c. Effect of a computer-aided diagnosis system on radiologists' performance in grading gliomas with MRI. *PLOS ONE,* 12(2)**,** e0171342.

Hutter, A., Schwetye, K. E., Bierhals, A. J. & McKinstry, R. C. 2003. Brain neoplasms: epidemiology, diagnosis, and prospects for cost-effective imaging. *Neuroimaging Clinics of North America,* 13(2)**,** 237-250.

Hwang, J.-N. & Hu, Y. H. 2001. *Handbook of neural network signal processing,* Boca Raton, Florida, CRC press.

Illán, I. A., Górriz, J. M., López, M. M., Ramírez, J., Salas-Gonzalez, D., Segovia, F., Chaves, R. & Puntonet, C. G. 2011. Computer aided diagnosis of Alzheimer's disease using component based SVM. *Applied Soft Computing,* 11(2)**,** 2376-2382.

Inano, R., Oishi, N., Kunieda, T., Arakawa, Y., Yamao, Y., Shibata, S., Kikuchi, T., Fukuyama, H. & Miyamoto, S. 2014. Voxel-based clustered imaging by multiparameter diffusion tensor images for glioma grading. *NeuroImage: Clinical,* 5396-407.

Iram, S., Al Jumeily, D., Fergus, P. & Hussain, A. Exploring the Hidden Challenges Associated with the Evaluation of Multi-class Datasets Using Multiple Classifiers. Complex, Intelligent and Software Intensive Systems (CISIS), 2014 Eighth International Conference on, 2014 Birmingham, UK. IEEE, 346-352.

Işın, A., Direkoğlu, C. & Şah, M. 2016. Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods. *Procedia Computer Science,* 102317-324.

Jafari, P. & Azuaje, F. 2006. An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors. *BMC Medical Informatics and Decision Making,* 6(1)**,** 27.

REFERENCES

Javed, U., Riaz, M. M., Ghafoor, A. & Cheema, T. A. 2013. MRI brain classification using texture features, fuzzy weighting and support vector machine. *Progress In Electromagnetics Research B,* 5373-88.

Jiang, J., Trundle, P. & Ren, J. 2010. Medical image analysis with artificial neural networks. *Computerized Medical Imaging and Graphics,* 34(8)**,** 617-631.

Johnson, K. J. & Synovec, R. E. 2002. Pattern recognition of jet fuels: comprehensive GC×GC with ANOVA-based feature selection and principal component analysis. *Chemometrics and Intelligent Laboratory Systems,* 60(1)**,** 225-237.

Jun, T. & Jian, Y. Developing an intelligent data discriminating system of anti-money laundering based on SVM. 2005 International Conference on Machine Learning and Cybernetics, 18-21 Aug. 2005 2005. 3453-3457 Vol. 6.

Just, N. 2011. Histogram analysis of the microvasculature of intracerebral human and murine glioma xenografts. *Magnetic Resonance in Medicine,* 65(3)**,** 778-789.

Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D. & Glocker, B. 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis,* 3661-78.

Kang, Y., Choi, S. H., Kim, Y.-J., Kim, K. G., Sohn, C.-H., Kim, J.-H., Yun, T. J. & Chang, K.-H. 2011. Gliomas: Histogram Analysis of Apparent Diffusion Coefficient Maps with Standard- or High-b-Value Diffusion-weighted MR Imaging—Correlation with Tumor Grade. *Radiology,* 261(3)**,** 882-890.

Karpathy, A. 2016. *Connecting images and natural language.* Ph. D. thesis, Stanford University.

Kassner, A. & Thornhill, R. 2010. Texture analysis: a review of neurologic MR imaging applications. *American Journal of Neuroradiology,* 31(5)**,** 809-816.

Khalid, A. & Noureldien, N. A. 2014. Determining the Efficient Structure of Feed-Forward Neural Network to Classify Breast Cancer Dataset. *International Journal of Advanced Computer Science and Applications,* 1(2014)**,** 87-90.

Kharrat, A., Gasmi, K., Messaoud, M. B., Benamrane, N. & Abid, M. 2010. A hybrid approach for automatic classification of brain MRI using genetic algorithm and support vector machine. *Leonardo Journal of Sciences,* 17(1)**,** 71-82.

Khawaldeh, S., Pervaiz, U., Rafiq, A. & Alkhawaldeh, R. S. 2017. Noninvasive Grading of Glioma Tumor Using Magnetic Resonance Imaging with Convolutional Neural Networks. *Applied Sciences,* 8(1)**,** 27.

Kitange, G. J., Templeton, K. L. & Jenkins, R. B. 2003. Recent advances in the molecular genetics of primary gliomas. *Current Opinion in Oncology,* 15(3)**,** 197-203.

Kjaer, L., Ring, P., Thomsen, C. & Henriksen, O. 1995. Texture analysis in quantitative MR imaging: tissue characterisation of normal brain and intracranial tumours at 1.5 T. *Acta Radiologica,* 36(2)**,** 127-135.

Kono, K., Inoue, Y., Nakayama, K., Shakudo, M., Morino, M., Ohata, K., Wakasa, K. & Yamada, R. 2001. The role of diffusion-weighted imaging in patients with brain tumors. *American Journal of Neuroradiology,* 22(6)**,** 1081-1088.

Kothari, C. R. 2004. *Research methodology: Methods and techniques,* New Delhi, New Age International.

Kotsiantis, S. B., Zaharakis, I. & Pintelas, P. 2007. Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologie. Vol 160, 3-24.* Netherlands: IOS press.

REFERENCES

Kovalev, V. & Kruggel, F. 2007. Texture Anisotropy of the Brain's White Matter as Revealed by Anatomical MRI. *IEEE Transactions on Medical Imaging,* 26(5)**,** 678-685.

Kulkarni, S. R., Lugosi, G. & Venkatesh, S. S. 1998. Learning pattern classification-a survey. *IEEE Transactions on Information Theory,* 44(6)**,** 2178-2206.

Kumar, M. & Singh, K. M. 2018. Retrieval of head–neck medical images using Gabor filter based on power-law transformation method and rank BHMT. *Signal, Image and Video Processing,* 12(5)**,** 827-833.

Kumar, S., Ghosh, J. & Crawford, M. M. 2002. Hierarchical Fusion of Multiple Classifiers for Hyperspectral Data Analysis. *Pattern Analysis & Applications,* 5(2)**,** 210-220.

Kumar, S., Moni, R. & Rajeesh, J. 2013. An automatic computer-aided diagnosis system for liver tumours on computed tomography images. *Computers & Electrical Engineering,* 39(5)**,** 1516-1526.

Kuncheva, L. 2014. *Combining Pattern Classifiers Methods and Algorithms,* Hoboken, New Jersey, USA, John Wiley & Sons.

Labani, M., Moradi, P., Ahmadizar, F. & Jalili, M. 2018. A novel multivariate filter method for feature selection in text classification problems. *Engineering Applications of Artificial Intelligence,* 70(1)**,** 25-37.

Lai, P. H., Ho, J. T., Chen, W. L., Hsu, S. S., Wang, J. S., Pan, H. B. & Yang, C. F. 2002. Brain abscess and necrotic brain tumor: discrimination with proton MR spectroscopy and diffusion-weighted imaging. *American Journal of Neuroradiology,* 23(8)**,** 1369-1377.

Larroza, A., Bodí, V. & Moratal, D. 2016. Texture Analysis in Magnetic Resonance Imaging: Review and Considerations for Future Applications. *In:* CONSTANTINIDES, C. (ed.) *Assessment of Cellular and Organ Function and Dysfunction using Direct and Derived MRI Methodologies.* InTech.

Lasocki, A., Tsui, A., Tacey, M., Drummond, K., Field, K. & Gaillard, F. 2015. MRI Grading versus Histology: Predicting Survival of World Health Organization Grade II–IV Astrocytomas. *American Journal of Neuroradiology,* 36(1)**,** 77-83.

Latif, G., Iskandar, D. N. F. A., Alghazo, J. M. & Mohammad, N. 2019. Enhanced MR Image Classification Using Hybrid Statistical and Wavelets Features. *IEEE Access,* 7(1)**,** 9634-9644.

Law, M., Oh, S., Babb, J. S., Wang, E., Inglese, M., Zagzag, D., Knopp, E. A. & Johnson, G. 2006. Low-Grade Gliomas: Dynamic Susceptibility-weighted Contrast-enhanced Perfusion MR Imaging—Prediction of Patient Clinical Response 1. *Radiology,* 238(2)**,** 658-667.

Law, M., Yang, S., Babb, J. S., Knopp, E. A., Golfinos, J. G., Zagzag, D. & Johnson, G. 2004. Comparison of cerebral blood volume and vascular permeability from dynamic susceptibility contrast-enhanced perfusion MR imaging with glioma grade. *AJNR. American journal of neuroradiology,* 25(5)**,** 746.

Law, M., Yang, S., Wang, H., Babb, J. S., Johnson, G., Cha, S., Knopp, E. A. & Zagzag, D. 2003. Glioma grading: sensitivity, specificity, and predictive values of perfusion MR imaging and proton MR spectroscopic imaging compared with conventional MR imaging. *American Journal of Neuroradiology,* 24(10)**,** 1989-1998.

Le, Q. V. Building high-level features using large scale unsupervised learning. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013 Vancouver, BC, Canada. IEEE, 8595-8598.

Lekutai, G. 1997. *Adaptive self-tuning neuro wavelet network controllers.* Virginia Tech.

REFERENCES

Lerski, R. A., de Certaines, J. D., Duda, D., Klonowski, W., Yang, G., Coatrieux, J. L., Azzabou, N. & Eliat, P.-A. 2015. Application of texture analysis to muscle MRI: 2 – technical recommendations. *EPJ Nonlinear Biomedical Physics,* 3(1)**,** 2.

Li, G.-Z., Yang, J., Ye, C.-Z. & Geng, D.-Y. 2006. Degree prediction of malignancy in brain glioma using support vector machines. *Computers in Biology and Medicine,* 36(3)**,** 313-325.

Li, X., Zhu, Y., Kang, H., Zhang, Y., Liang, H., Wang, S. & Zhang, W. 2015. Glioma grading by microvascular permeability parameters derived from dynamic contrast-enhanced MRI and intratumoral susceptibility signal on susceptibility weighted imaging. *Cancer Imaging,* 15(1)**,** 4.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B. & Sánchez, C. I. 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis,* 42(1)**,** 60-88.

Liu, R., Elhalawani, H., Fuller, C. & Zhu, H. 2018. *Stability Analysis of CT Radiomics Features With Respect to the Variation of Manual Segmentation in Oropharyngeal Cancer.* 100.

Loizou, C. P., Pantziaris, M., Seimenis, I. & Pattichis, C. S. Brain MR image normalization in texture analysis of multiple sclerosis. 9th International Conference on Information Technology and Applications in Biomedicine, 2009 Larnaca, Cyprus. IEEE, 1-5.

Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., Scheithauer, B. W. & Kleihues, P. 2007. The 2007 WHO Classification of Tumours of the Central Nervous System. *Acta Neuropathologica,* 114(2)**,** 97-109.

Lu, D. & Weng, Q. 2007. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing,* 28(5)**,** 823-870.

Lu, J., Plataniotis, K. N. & Venetsanopoulos, A. N. 2005. Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. *Pattern Recognition Letters,* 26(2)**,** 181-191.

Luts, J., Heerschap, A., Suykens, J. A. & Van Huffel, S. 2007. A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection. *Artificial Intelligence in Medicine,* 40(2)**,** 87-102.

Ly, A., Marsman, M. & Wagenmakers, E. J. 2018. Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica,* 72(1)**,** 4-13.

Maani, R., Kalra, S. & Yang, Y.-H. 2016. A review of texture classification methods and their applications in medical image analysis of the brain. *Handbook of pattern recognition and computer vision.* Singapore: World Scientific.

Maggiori, E., Tarabalka, Y., Charpiat, G. & Alliez, P. 2017. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing,* 55(2)**,** 645-657.

Mahmoud-Ghoneim, D., Toussaint, G., Constans, J.-M. & de Certaines, J. D. 2003. Three dimensional texture analysis in MRI: a preliminary evaluation in gliomas. *Magnetic Resonance Imaging,* 21(9)**,** 983-987.

Mamoshina, P., Vieira, A., Putin, E. & Zhavoronkov, A. 2016. Applications of Deep Learning in Biomedicine. *Molecular Pharmaceutics,* 13(5)**,** 1445-1454.

Marshkole, N., Singh, B. K. & Thoke, A. 2011. Texture and shape based classification of brain tumors using linear vector quantization. *International Journal of Computer Applications,* 30(11)**,** 21-23.

Materka, A. & Strzelecki, M. 1998. Texture analysis methods–a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels,* 9-11.

REFERENCES

Mazurowski, M. A., Buda, M., Saha, A. & Bashir, M. R. 2019. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *Journal of Magnetic Resonance Imaging,* 49(4)**,** 939-954.

Mehra, N. & Gupta, S. 2013. Survey on multiclass classification methods. *International Journal of Computer Science and Information Technologies,* 4(4)**,** 572-576.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M. A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., Corso, J. J., Criminisi, A., Das, T., Delingette, H., Ç, D., Durst, C. R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K. M., Jena, R., John, N. M., Konukoglu, E., Lashkari, D., Mariz, J. A., Meier, R., Pereira, S., Precup, D., Price, S. J., Raviv, T. R., Reza, S. M. S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H. C., Shotton, J., Silva, C. A., Sousa, N., Subbanna, N. K., Szekely, G., Taylor, T. J., Thomas, O. M., Tustison, N. J., Unal, G., Vasseur, F., Wintermark, M., Ye, D. H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M. & Leemput, K. V. 2015. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging,* 34(10)**,** 1993-2024.

Moenninghoff, C., Maderwald, S., Theysohn, J. M., Kraff, O., Ladd, M. E., El Hindy, N., van de Nes, J., Forsting, M. & Wanke, I. 2010. Imaging of adult astrocytic brain tumours with 7 T MRI: preliminary results. *European radiology,* 20(3)**,** 704-713.

Mohan, G. & Subashini, M. M. 2018. MRI based medical image analysis: Survey on brain tumor grade classification. *Biomedical Signal Processing and Control,* 39139-161.

Mohsen, H., El-Dahshan, E.-S. A., El-Horbaty, E.-S. M. & Salem, A.-B. M. 2018. Classification using deep learning neural networks for brain tumors. *Future Computing and Informatics Journal,* 3(1)**,** 68-71.

Molina, D., Pérez-Beteta, J., Luque, B., Arregui, E., Calvo, M., Borrás, J. M., López, C., Martino, J., Velasquez, C., Asenjo, B., Benavides, M., Herruzo, I., Martínez-González, A., Pérez-Romasanta, L., Arana, E. & Pérez-García, V. M. 2016. Tumour heterogeneity in glioblastoma assessed by MRI texture analysis: a potential marker of survival. *The British Journal of Radiology,* 89(1064)**,** 20160242.

Moore, K. & Kim, L. 2010. Primary Brain Tumors: Characteristics, Practical Diagnostic and Treatment Approaches. *In:* RAY, K. S. (ed.) *Glioblastoma: Molecular Mechanisms of Pathogenesis and Current Therapeutic Strategies.* New York, NY: Springer.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J. & Initiative, A. s. D. N. 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage,* 104(1)**,** 398-412.

Nabizadeh, N. & Kubat, M. 2015. Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features. *Computers & Electrical Engineering,* 45(1)**,** 286-301.

Nakagawa, M., Nakaura, T., Namimoto, T., Kitajima, M., Uetani, H., Tateishi, M., Oda, S., Utsunomiya, D., Makino, K., Nakamura, H., Mukasa, A., Hirai, T. & Yamashita, Y. 2018. Machine learning based on multi-parametric magnetic resonance imaging to differentiate glioblastoma multiforme from primary cerebral nervous system lymphoma. *European Journal of Radiology,* 108147-154.

Negnevitsky, M. 2005. *Artificial intelligence: a guide to intelligent systems,* England, Pearson Education.

Ng, F., Kozarski, R., Ganeshan, B. & Goh, V. 2013. Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? *European journal of radiology,* 82(2)**,** 342-348.

REFERENCES

Nielsen, B., Albregtsen, F. & Danielsen, H. E. 2008. Statistical nuclear texture analysis in cancer research: a review of methods and applications. *Critical Reviews™ in Oncogenesis,* 14(2-3).

Nixon, M. A., A. 2008. *Feature Extraction Image Processing,* Hungary, Elsevier.

Nyúl, L. G. & Udupa, J. K. 1999. On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine,* 42(6)**,** 1072-1081.

Othman, M. F. & Basri, M. A. M. Probabilistic neural network for brain tumor classification. Intelligent Systems, Modelling and Simulation (ISMS), 2011 Second International Conference on Intelligent Systems, Modelling and Simulation, 2011 Kuala Lumpur, Malaysia. IEEE, 136-138.

Oza, N. C. & Tumer, K. 2008. Classifier ensembles: Select real-world applications. *Information Fusion,* 9(1)**,** 4-20.

Özkan, C. & Erbek, F. S. 2003. The comparison of activation functions for multispectral Landsat TM image classification. *Photogrammetric Engineering & Remote Sensing,* 69(11)**,** 1225-1234.

Pan, Y., Huang, W., Lin, Z., Zhu, W., Zhou, J., Wong, J. & Ding, Z. Brain tumor grading based on Neural Networks and Convolutional Neural Networks. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 25-29 Aug. 2015 2015 Milan, Italy. IEEE, 699-702.

Pantelis, G. 2010. *Computer Assisted Diagnosis of Brain Tumors based on Statistical Methods and Pattern Recognition Techniques.* PhD, UNIVERSITY OF PATRAS.

Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B. & Swami, A. The Limitations of Deep Learning in Adversarial Settings. 2016 IEEE European Symposium on Security and Privacy (EuroS&P), 2016 Saarbrucken, Germany. 372-387.

Patel, D., Vankawala, F. & Bhatt, B. A Survey on Identification of Glioblastoma Multiforme and Low-Grade Glioma Brain Tumor Type. 2019 International Conference on Communication and Signal Processing (ICCSP), 4-6 April 2019 2019 Chennai, India. 0335-0339.

Petrou, M. 2010. Texture in biomedical images. *Biomedical Image Processing.* New York: Springer.

Polikar, R. 2006. Ensemble based systems in decision making. *IEEE Circuits and systems magazine,* 6(3)**,** 21-45.

Ponti Jr, M. P. Combining classifiers: from the creation of ensembles to the decision fusion. Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials, 2011 Alagoas, Brazil. IEEE, 1-10.

Porto, L., Jurcoane, A., Schwabe, D. & Hattingen, E. 2014. Conventional magnetic resonance imaging in the differentiation between high and low-grade brain tumours in paediatric patients. *European Journal of Paediatric Neurology,* 18(1)**,** 25-29.

Pouratian, N. & Schiff, D. 2010. Management of Low-Grade Glioma. *Current Neurology and Neuroscience Reports,* 10(3)**,** 224-231.

Prajapati, G. L. & Patle, A. On Performing Classification Using SVM with Radial Basis and Polynomial Kernel Functions. 3rd International Conference on Emerging Trends in Engineering and Technology, 19-21 Nov. 2010 2010 Goa, India. 512-515.

Purcell, E. M., Torrey, H. C. & Pound, R. V. 1946. Resonance absorption by nuclear magnetic moments in a solid. *Physical review,* 69(1-2)**,** 37.

Qian, S. & Chen, D. 1993. Discrete gabor transform. *IEEE transactions on signal processing,* 41(7)**,** 2429-2438.

Qian, Y., Bi, M., Tan, T. & Yu, K. 2016. Very Deep Convolutional Neural Networks for Noise Robust Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* 24(12)**,** 2263-2276.

Rajan, S. & Ghosh, J. An empirical comparison of hierarchical vs. two-level approaches to multiclass problems. *In:* ROLI, F., KITTLER, J. & WINDEATT, T., eds. International Workshop on Multiple Classifier Systems, 2004 Berlin, Heidelberg. Springer, 283-292.

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B. & Yang, G. 2017. Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics,* 21(1)**,** 4-21.

Rodríguez Colmeiro, R. G., Verrastro, C. A. & Grosges, T. Multimodal Brain Tumor Segmentation Using 3D Convolutional Networks. International MICCAI Brain lesion Workshop, 2017 Cham. Springer International Publishing, 226-240.

Rose, C. J., O'Connor, J. P. B., Cootes, T. F., Taylor, C. J., Jayson, G. C., Parker, G. J. M. & Waterton, J. C. 2014. Indexed distribution analysis for improved significance testing of spatially heterogeneous parameter maps: Application to dynamic contrast-enhanced MRI biomarkers. *Magnetic Resonance in Medicine,* 71(3)**,** 1299-1311.

Roshdy, N., Shahin, M., Kishk, H., El-Khouly, S., Mousa, A. & Elsalekh, I. 2010. Role of New Magnetic Resonance Imaging Modalities in Diagnosis of Orbital Masses: A Clinicopathologic Correlation. *Middle East African Journal of Ophthalmology,* 17(2)**,** 175-179.

Roy, S., Nag, S., Maitra, I. K. & Bandyopadhyay, S. K. 2013. A Review on Automated Brain Tumor Detection and Segmentation from MRI of Brain. *arXiv preprint arXiv:1312.6150* [Online].

Ruoslahti, E. 2002. Specialization of tumour vasculature. *Nature Reviews Cancer,* 2(2)**,** 83-90.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. & Fei-Fei, L. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision,* 115(3)**,** 211-252.

Ryu, Y. J., Choi, S. H., Park, S. J., Yun, T. J., Kim, J.-H. & Sohn, C.-H. 2014. Glioma: Application of Whole-Tumor Texture Analysis of Diffusion-Weighted Imaging for the Evaluation of Tumor Heterogeneity. *PLoS ONE,* 9(9)**,** e108335. .

Saad, N. M., Bakar, S. A. R. S. A., Muda, A. S. & Mokji, M. M. 2015. Review of brain lesion detection and classification using neuroimaging analysis techniques. *Jurnal Teknologi,* 74(6)**,** 73-85.

Saeys, Y., Inza, I. & Larrañaga, P. 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics,* 23(19)**,** 2507-2517.

Sahiner, B., Chan, H.-P. & Hadjiiski, L. 2008. Classifier performance estimation under the constraint of a finite sample size: Resampling schemes applied to neural network classifiers. *Neural Networks,* 21(2)**,** 476-483.

Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A. & Baik, S. W. 2019. Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *Journal of Computational Science,* 30(1)**,** 174-182.

Sanghani, P., Ang, B. T., King, N. K. K. & Ren, H. 2018. Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. *Surgical Oncology,* 27(4)**,** 709-714.

Savio, S. J., Harrison, L. C., Luukkaala, T., Heinonen, T., Dastidar, P., Soimakallio, S. & Eskola, H. J. 2010. Effect of slice thickness on brain magnetic resonance image texture analysis. *BioMedical Engineering OnLine,* 9(1)**,** 60.

Schwartzbaum, J. A., Fisher, J. L., Aldape, K. D. & Wrensch, M. 2006. Epidemiology and molecular pathology of glioma. *Nat Clin Pract Neuro,* 2(9)**,** 494-503.

Shen, D., Wu, G. & Suk, H.-I. 2017. Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering,* 19(1)**,** 221-248.

Siker, M. L., Chakravarti, A. & Mehta, M. P. 2006. Should concomitant and adjuvant treatment with temozolomide be used as standard therapy in patients with anaplastic glioma? *Critical Reviews in Oncology/Hematology,* 60(2)**,** 99-111.

Skogen, K., Ganeshan, B., Good, C., Critchley, G. & Miles, K. 2013. Measurements of heterogeneity in gliomas on computed tomography relationship to tumour grade. *Journal of Neuro-oncology,* 111(2)**,** 213-219.

Skogen, K., Schulz, A., Dormagen, J. B., Ganeshan, B., Helseth, E. & Server, A. 2016. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. *European Journal of Radiology,* 85(4)**,** 824-829.

Soh, L.-K. & Tsatsoulis, C. 1999. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing,* 37(2)**,** 780-795.

Song, S., Nones, K., Miller, D., Harliwong, I., Kassahn, K. S., Pinese, M., Pajic, M., Gill, A. J., Johns, A. L., Anderson, M., Holmes, O., Leonard, C., Taylor, D., Wood, S., Xu, Q., Newell, F., Cowley, M. J., Wu, J., Wilson, P., Fink, L., Biankin, A. V., Waddell, N., Grimmond, S. M. & Pearson, J. V. 2012. qpure: A Tool to Estimate Tumor Cellularity from Genome-Wide Single-Nucleotide Polymorphism Profiles. *PLOS ONE,* 7(9)**,** e45835.

Song, Y.-y. & Lu, Y. 2015. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry,* 27(2)**,** 130-135.

Song, Y. S., Choi, S. H., Park, C.-K., Yi, K. S., Lee, W. J., Yun, T. J., Kim, T. M., Lee, S.-H., Kim, J.-H., Sohn, C.-H., Park, S.-H., Kim, I. H., Jahng, G.-H. & Chang, K.-H. 2013. True Progression versus Pseudoprogression in the Treatment of Glioblastomas: A Comparison Study of Normalized Cerebral Blood Volume and Apparent Diffusion Coefficient by Histogram Analysis. *Korean Journal of Radiol,* 14(4)**,** 662-672.

Stupp, R., Tonn, J.-C., Brada, M., Pentheroudakis, G. & Group, O. b. o. t. E. G. W. 2010. High-grade malignant glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology,* 21(suppl_5)**,** v190-v193.

Subashini, M. M., Sahoo, S. K., Karthikeyan, S. P. & Raglend, I. J. 2015. A Grade Prediction Methodology for Astrocytoma Using Modified K-Clustering Network. *In:* KAMALAKANNAN, C., SURESH, L. P., DASH, S. S. & PANIGRAHI, B. K. (eds.) *Power Electronics and Renewable Energy Systems.* New Delhi: Springer.

Subashini, M. M., Sahoo, S. K., Sunil, V. & Easwaran, S. 2016. A non-invasive methodology for the grade identification of astrocytoma using image processing and artificial intelligence techniques. *Expert Systems with Applications,* 43(1)**,** 186-196.

Sugahara, T., Korogi, Y., Kochi, M., Ikushima, I., Shigematu, Y., Hirai, T., Okuda, T., Liang, L., Ge, Y., Komohara, Y., Ushio, Y. & Takahashi, M. 1999. Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas. *Journal of Magnetic Resonance Imaging,* 9(1)**,** 53-60.

Swietojanski, P., Ghoshal, A. & Renals, S. 2014. Convolutional Neural Networks for Distant Speech Recognition. *IEEE Signal Processing Letters,* 21(9)**,** 1120-1124.

Swinscow, T. D. V. & Campbell, M. J. 2002. *Statistics at Square One,* London, Bmj

Tang, X. 1998. Texture information in run-length matrices. *IEEE Transactions on Image Processing,* 7(11)**,** 1602-1609.

REFERENCES

Tantisatirapong, S. 2015. *Texture analysis of multimodal magnetic resonance images in support of diagnostic classification of childhood brain tumours.* PhD, University of Birmingham,UK.

Taouli, B., Vilgrain, V., Dumont, E., Daire, J.-L., Fan, B. & Menu, Y. 2003. Evaluation of Liver Diffusion Isotropy and Characterization of Focal Hepatic Lesions with Two Single-Shot Echo-planar MR Imaging Sequences: Prospective Study in 66 Patients 1. *Radiology,* 226(1)**,** 71-78.

Theeler, B. J. & Groves, M. D. 2011. High-Grade Gliomas. *Current Treatment Options in Neurology,* 13(4)**,** 386-399.

Thomas, P., Bril El Haouzi, H., Suhner, M.-C., Thomas, A., Zimmermann, E. & Noyel, M. 2018. Using a classifier ensemble for proactive quality monitoring and control: The impact of the choice of classifiers types, selection criterion, and fusion process. *Computers in Industry,* 99(1)**,** 193-204.

Thust, S. C., Heiland, S., Falini, A., Jäger, H. R., Waldman, A. D., Sundgren, P. C., Godi, C., Katsaros, V. K., Ramos, A., Bargallo, N., Vernooij, M. W., Yousry, T., Bendszus, M. & Smits, M. 2018. Glioma imaging in Europe: A survey of 220 centres and recommendations for best clinical practice. *European Radiology,* 28(8)**,** 3306-3317.

Tin Kam, H. 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 20(8)**,** 832-844.

Toennies, K. D. 2017. *Guide to Medical Image Analysis,* London, Springer.

Tonarelli, L. 2013. Magnetic resonance imaging of brain tumor. *CEwebsource.com* [Online]. [Accessed 2016. available https://www.semanticscholar.org/paper/Magnetic-Resonance-Imaging-of-Brain-Tumor-Tonarelli/37994c218084367dce4a61b9069885b0f9da2c40].

Tozer, D. J., Jäger, H. R., Danchaivijitr, N., Benton, C. E., Tofts, P. S., Rees, J. H. & Waldman, A. D. 2007. Apparent diffusion coefficient histograms may predict low-grade glioma subtype. *NMR in Biomedicine,* 20(1)**,** 49-57.

Tsirogiannis, G. L., Frossyniotis, D., Nikita, K. S. & Stafylopatis, A. A meta-classifier approach for medical diagnosis. Hellenic Conference on Artificial Intelligence, 2004 Berlin, Heidelberg. Springer, 154-163.

Walnut, D. F. 2013. *An introduction to wavelet analysis*, Springer Science & Business Media.

Weller, M. 2011. Novel diagnostic and therapeutic approaches to malignant glioma. *Swiss Medical Weekly,* 141(1)**,** w13210.

Wibmer, A., Hricak, H., Gondo, T., Matsumoto, K., Veeraraghavan, H., Fehr, D., Zheng, J., Goldman, D., Moskowitz, C., Fine, S. W., Reuter, V. E., Eastham, J., Sala, E. & Vargas, H. A. 2015. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *European Radiology,* 25(10)**,** 2840-2850.

William, W., Basaza-Ejiri, A. H., Obungoloch, J. & Ware, A. A Review of Applications of Image Analysis and Machine Learning Techniques in Automated Diagnosis and Classification of Cervical Cancer from Pap-smear Images. IST-Africa Week Conference (IST-Africa), 2018 Gaborone, Botswana. IEEE, Page 1 of 11-Page 11 of 11.

Wintermark, M., Sesay, M., Barbier, E., Borbély, K., Dillon, W. P., Eastwood, J. D., Glenn, T. C., Grandin, C. B., Pedraza, S. & Soustiel, J.-F. 2005. Comparative overview of brain perfusion imaging techniques. *Stroke,* 36(9)**,** e83-e99.

REFERENCES

Woo, S., Cho, J. Y., Kim, S. Y. & Kim, S. H. 2014. Histogram analysis of apparent diffusion coefficient map of diffusion-weighted MRI in endometrial cancer: a preliminary correlation study with histological grade. *Acta Radiologica,* 55(10)**,** 1270-1277.

Woźniak, M., Graña, M. & Corchado, E. 2014. A survey of multiple classifier systems as hybrid systems. *Information Fusion,* 16(1)**,** 3-17.

Wu, X., Yang, J. & Wang, S. 2018. Tea category identification based on optimal wavelet entropy and weighted k-Nearest Neighbors algorithm. *Multimedia Tools and Applications,* 77(3)**,** 3745-3759.

Xi, Y.-b., Kang, X.-w., Wang, N., Liu, T.-t., Zhu, Y.-q., Cheng, G., Wang, K., Li, C., Guo, F. & Yin, H. 2019. Differentiation of primary central nervous system lymphoma from high-grade glioma and brain metastasis using arterial spin labeling and dynamic contrast-enhanced magnetic resonance imaging. *European Journal of Radiology,* 11259-64.

Xu, L., Krzyzak, A. & Suen, C. Y. 1992. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics,* 22(3)**,** 418-435.

Yang, X., Tridandapani, S., Beitler, J. J., Yu, D. S., Yoshida, E. J., Curran, W. J. & Liu, T. 2012. Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity. *Medical physics,* 39(9)**,** 5732-5739.

Yazdi, M., Adelpour, Z., Bahraini, B. & Jahromi, Y. K. 2007. Novel ridge orientation based approach for fingerprint identification using co-occurrence matrix. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering,* 1(11)**,** 3414-3418.

Ye, C.-Z., Yang, J., Geng, D.-Y., Zhou, Y. & Chen, N.-Y. 2002. Fuzzy rules to predict degree of malignancy in brain glioma. *Medical and Biological Engineering and Computing,* 40(2)**,** 145-152.

Ye, F., Pu, J., Wang, J., Li, Y. & Zha, H. Glioma grading based on 3D multimodal convolutional neural network and privileged learning. IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017 Kansas City, MO, USA. 759-763.

Zacharaki, E. I., Kanas, V. G. & Davatzikos, C. 2011. Investigating machine learning techniques for MRI-based classification of brain neoplasms. *International Journal of Computer Assisted Radiology and Surgery,* 6(6)**,** 821-828.

Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R. & Davatzikos, C. 2009. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magnetic Resonance in Medicine,* 62(6)**,** 1609-1618.

Zhang, X., Yan, L.-F., Hu, Y.-C., Li, G., Yang, Y., Han, Y., Sun, Y.-Z., Liu, Z.-C., Tian, Q., Han, Z.-Y., Liu, L.-D., Hu, B.-Q., Qiu, Z.-Y., Wang, W. & Cui, G.-B. 2017. Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features. *Oncotarget,* 8(29)**,** 47816-47830.

Zhang, Y., Lu, S., Zhou, X., Yang, M., Wu, L., Liu, B., Phillips, P. & Wang, S. 2016. Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine. *SIMULATION,* 92(9)**,** 861-871.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A. & Oliva, A. Learning deep features for scene recognition using places database. Advances in neural information processing systems, 2014. 487-495.

REFERENCES

Zhou, Z.-H. & Liu, X.-Y. 2006. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering,* 18(1)**,** 63-77.

BIBLIOGRAPHY

MATLAB and Image Processing Toolbox Release R2018a, The MathWorks, Inc., Natick, Massachusetts, United States.

Uppuluri, Avinash (2008). GLCM Texture Features (https://uk.mathworks.com/matlabcentral/fileexchange/22187-glcm-texture-features), MATLAB Central File Exchange. Retrieved February 12, 2016.

Carl Philips and Daniel Li (2008). 3D statistical texture algorithm (https://uk.mathworks.com/matlabcentral/fileexchange/19058-cooc3d). Retrieved February 5, 2016.

APPENDIX A

The classification accuracy results for different number of neurons within the proposed MTMCS based on three-layer NNs, and with applying the diversity in feature space. The highest accuracy is achieved when 6, 18, 6 in the first, second, third layers respectively at the 10th iteration. Some iterations results are reported to show the difference in output results according to different number of neurons in the layers of NNs. To simplify the view of the results, precision of two decimal places are used. Layer 1, layer 2, layer 3 indicate the first layer, second layer, and third layer of NNs respectively. ACC indicates the classification accuracy

Neurons in Layer 1	Neurons in Layer 2	Neurons in Layer 3	ACC	Neurons in Layer 1	Neurons in Layer 2	Neurons in Layer 3	ACC	Neurons in Layer 1	Neurons in Layer 2	Neurons in Layer 3	ACC	Neurons in Layer 1	Neurons in Layer 2	Neurons in Layer 3	ACC
1	1	1	0.91	2	4	1	0.87	3	7	1	0.91	4	10	1	0.91
2	1	1	0.89	3	4	1	0.89	4	7	1	0.91	5	10	1	0.89
3	1	1	0.89	4	4	1	0.89	5	7	1	0.9	6	10	1	0.91
4	1	1	0.89	5	4	1	0.91	6	7	1	0.91	7	10	1	0.91
5	1	1	0.91	6	4	1	0.91	7	7	1	0.89	8	10	1	0.88
6	1	1	0.89	7	4	1	0.9	8	7	1	0.9	9	10	1	0.89
7	1	1	0.91	8	4	1	0.9	9	7	1	0.9	1	11	1	0.85
8	1	1	0.9	9	4	1	0.89	1	8	1	0.86	2	11	1	0.9
9	1	1	0.89	1	5	1	0.89	2	8	1	0.89	3	11	1	0.89
1	2	1	0.91	2	5	1	0.9	3	8	1	0.88	4	11	1	0.9
2	2	1	0.91	3	5	1	0.91	4	8	1	0.9	5	11	1	0.89
3	2	1	0.9	4	5	1	0.91	5	8	1	0.9	6	11	1	0.91
4	2	1	0.88	5	5	1	0.91	6	8	1	0.91	7	11	1	0.89
5	2	1	0.89	6	5	1	0.89	7	8	1	0.91	8	11	1	0.9
6	2	1	0.89	7	5	1	0.89	8	8	1	0.91	9	11	1	0.89
7	2	1	0.89	8	5	1	0.91	9	8	1	0.9	1	12	1	0.87
8	2	1	0.91	9	5	1	0.9	1	9	1	0.86	2	12	1	0.9
9	2	1	0.89	1	6	1	0.84	2	9	1	0.91	3	12	1	0.9
1	3	1	0.89	2	6	1	0.88	3	9	1	0.9	4	12	1	0.88
2	3	1	0.88	3	6	1	0.89	4	9	1	0.89	5	12	1	0.91
3	3	1	0.89	4	6	1	0.91	5	9	1	0.88	6	12	1	0.9
4	3	1	0.91	5	6	1	0.91	6	9	1	0.9	7	12	1	0.91
5	3	1	0.87	6	6	1	0.9	7	9	1	0.9	8	12	1	0.9
6	3	1	0.89	7	6	1	0.91	8	9	1	0.89	9	12	1	0.9
7	3	1	0.91	8	6	1	0.91	9	9	1	0.91	1	13	1	0.88
8	3	1	0.91	9	6	1	0.91	1	10	1	0.86	2	13	1	0.89
9	3	1	0.9	1	7	1	0.87	2	10	1	0.89	3	13	1	0.88
1	4	1	0.91	2	7	1	0.89	3	10	1	0.89	4	13	1	0.89

Neurons in Layer 1	Neurons in Layer 2	Neurons in Layer 3	ACC %	Neurons in Layer 1	Neurons in Layer 2	Neurons in Layer 3	ACC %	Neurons in Layer 1	Neurons in Layer 2	Neurons in Layer 3	ACC %	Neurons in Layer 1	Neurons in Layer 2	Neurons in Layer 3	ACC %
5	13	6	0.88	4	17	6	0.90	3	21	6	0.91	2	25	6	0.88
6	13	6	0.90	5	17	6	0.89	4	21	6	0.89	3	25	6	0.88
7	13	6	0.89	6	17	6	0.92	5	21	6	0.91	4	25	6	0.89
8	13	6	0.89	7	17	6	0.86	6	21	6	0.91	5	25	6	0.89
9	13	6	0.89	8	17	6	0.90	7	21	6	0.88	6	25	6	0.89
1	14	6	0.81	9	17	6	0.90	8	21	6	0.86	7	25	6	0.88
2	14	6	0.91	1	18	6	0.82	9	21	6	0.91	8	25	6	0.90
3	14	6	0.89	2	18	6	0.89	1	22	6	0.80	9	25	6	0.90
4	14	6	0.88	3	18	6	0.88	2	22	6	0.87	1	26	6	0.84
5	14	6	0.89	4	18	6	0.89	3	22	6	0.89	2	26	6	0.87
6	14	6	0.89	5	18	6	0.90	4	22	6	0.89	3	26	6	0.88
7	14	6	0.89	6	18	6	0.93	5	22	6	0.90	4	26	6	0.90
8	14	6	0.89	7	18	6	0.91	6	22	6	0.89	5	26	6	0.89
9	14	6	0.88	8	18	6	0.89	7	22	6	0.89	6	26	6	0.89
1	15	6	0.85	9	18	6	0.89	8	22	6	0.91	7	26	6	0.90
2	15	6	0.86	1	19	6	0.85	9	22	6	0.91	8	26	6	0.89
3	15	6	0.90	2	19	6	0.89	1	23	6	0.82	9	26	6	0.91
4	15	6	0.89	3	19	6	0.90	2	23	6	0.89	1	27	6	0.84
5	15	6	0.89	4	19	6	0.88	3	23	6	0.89	2	27	6	0.88
6	15	6	0.90	5	19	6	0.89	4	23	6	0.88	3	27	6	0.88
7	15	6	0.91	6	19	6	0.91	5	23	6	0.88	4	27	6	0.88
8	15	6	0.91	7	19	6	0.91	6	23	6	0.88	5	27	6	0.92
9	15	6	0.91	8	19	6	0.91	7	23	6	0.91	6	27	6	0.89
1	16	6	0.81	9	19	6	0.90	8	23	6	0.88	7	27	6	0.89
2	16	6	0.88	1	20	6	0.84	9	23	6	0.90	8	27	6	0.91
3	16	6	0.88	2	20	6	0.89	1	24	6	0.81	9	27	6	0.88
4	16	6	0.88	3	20	6	0.87	2	24	6	0.88	1	28	6	0.80
5	16	6	0.88	4	20	6	0.90	3	24	6	0.86	2	28	6	0.86
6	16	6	0.90	5	20	6	0.89	4	24	6	0.88	3	28	6	0.89
7	16	6	0.88	6	20	6	0.89	5	24	6	0.89	4	28	6	0.89
8	16	6	0.90	7	20	6	0.88	6	24	6	0.88	5	28	6	0.88
9	16	6	0.88	8	20	6	0.90	7	24	6	0.89	6	28	6	0.88
1	17	6	0.83	9	20	6	0.90	8	24	6	0.90	7	28	6	0.89
2	17	6	0.89	1	21	6	0.85	9	24	6	0.90	8	28	6	0.92
3	17	6	0.89	2	21	6	0.86	1	25	6	0.84	9	28	6	0.89

APPENDIX B

B.1 Statistical Textural Descriptors

Several statistical predictors that are recommended and widely used to recognise the image textural feature are utilised in this research work. These textural predictors are measured from the co-occurrence matrix of brain tumour images, which represent the local texture analysis of image patterns. They are successfully utilised to discriminate between different textural patterns (Gómez et al., 2012, Yang et al., 2012, Al-Waeli, 2017, Tantisatirapong, 2015) , and therefore they are used to measure image textural feature to discriminate between low and high glioma grades (Hsieh et al., 2017b, Hsieh et al., 2017c, Patel et al., 2019) . Further details and equations are explained in the followings subsection.

B.1.1 Autocorrelation

The autocorrelation predictor is used as an indicator for the variation in texture features. Images of a coarse texture will reveal a higher correlation than an image of a fine texture, indicating high values for high-grade tumours and low values for low-grade tumours. The autocorrelation function is defined by Eq. B.1 (Gómez et al., 2012, Nielsen et al., 2008).

$$autocorrlation = autoc = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} (i,j)P(i,j) \qquad B.1$$

where P is the probability co-occurrence matrix. i and j are the cell coordinates in the P, and M is the number of grey levels used.

B.1.2 Contrast

The contrast predictor presents local variations between a pixel and its neighbour. The increase of this variation favours the distribution being away from the diagonal of the co-occurrence. This predictor indicates low variations for low-grade tumour and higher weights for higher grades of malignancy of a tumour. It is defined by Eq. B.2 (Haralick et al., 1973).

$$contrast = contr = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} (i-j)^2 P(i,j) \qquad B.2$$

B.1.3 Correlation

The correlation predictor is used to measure the linear dependency between a pixel and its neighbours. This predictor tends to have relatively high values for normal lesions and lower values for increased malignancy grades of the tumour. The predictor function is defined by Eq. B.3 (Gómez et al., 2012, Yang et al., 2012).

$$Correlation = corrm = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} P(i.j)\frac{(i-u_x)(j-u_y)}{\sigma_x\sigma_y} \qquad B.3$$

Where σ_x, σ_y, u_x and u_y, are the standard deviations and means of P_x, and P_y, respectively, which are defined by Eqs. B.4-B.9.

Let M represent the grey level selected, while $P_x(i)$ is the ith element of the marginal probability matrix, obtained by summing the rows of P (i, j) and given by the Eq. B.4.

$$P_x(i) = \sum_{j=0}^{M-1} P(i,j) \qquad B.4$$

Similarly, P_y(j) is defined as j^{th} element of the marginal probability matrix, obtained by summing the columns of $P(i,j)$ and given by Eq. B.5.

$$P_y(j) = \sum_{i=0}^{M-1} P(i,j) \qquad B.5$$

$$u_x = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} i \cdot P(i,j) \qquad B.6$$

$$u_y = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} j \cdot P(i,j) \qquad B.7$$

$$\sigma_x^2 = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} (i-u_x)^2 P(i,j) \qquad B.8$$

$$\sigma_y^2 = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} (j-u_y)^2 P(i,j) \qquad B.9$$

B.1.4 Cluster Prominence

This predictor measures a peak of data distribution of the probability co-occurrence matrix. A low magnitude of this predictor indicates a small variation in the grey-spatial levels of the co-occurrence matrix. It weights low values for low-grade tumours and high values for high-grade tumours. The function of this predictor is defined by the Eq. B.10 (Gómez et al., 2012, Ananda Resmi and Thomas, 2010).

$$Cluster\ prominence = cprom = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1}\left(i+j-u_x+u_y\right)^4 P(i,j) \qquad B.10$$

B.1.5 Cluster Shade

This predictor is used to measure the symmetry of data distribution; it is also used to measure the skewness of the GLCM. A high magnitude of this predictor indicates asymmetry of the GLCM. It indicates low values for low-grade tumours and high values for high grades of malignant tumour. This predictor is defined by Eq. B.11 (Gómez et al., 2012, Ananda Resmi, 2013).

$$Cluster\ shade = cshad = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1}\left(i+j-u_x+u_y\right)^3 P(i,j) \qquad B.11$$

B.1.6 Dissimilarity

This predictor is used to measure the difference between the grey-level intensities. It indicates a low magnitude for low-grade tumours and high values for high-grade tumours. This predictor function is defined by the Eq. B.12 (Gómez et al., 2012, Molina et al., 2016).

$$Dissimilarity = dissi = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1}|i-j|P(i,j) \qquad B.12$$

B.1.7 Energy

This predictor is used as an indicator for the uniformity of the texture in an image. It is also known as an angular second moment and is used to measure the homogeneity of an image. A homogeneous texture will include only a little grey level, so that the GLCM will have few but relatively high magnitudes of the probability of the GLCM. Hence, the energy will be high when the image is homogeneous. It tends to weight high values for low-grade tumours and low values for high-grade tumours. It is defined by Eq. B.13 (Yang et al., 2012).

$$Energy = energ = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P(i,j) \qquad\qquad B.13$$

B.1.8 Entropy

This predictor is used to measure the randomness of the grey-level distribution. The entropy shows relatively smaller values when entries in the GLCM are unequal and it is highest when the probabilities in the GLCM are equal. Therefore, an inhomogeneous region will result in a higher entropy magnitude, while a homogeneous image will result in a lower entropy value. It tends to weight low values for the low grades of malignant tumour and high values for higher grades of malignancy. It is defined by Eq. B.14 (Haralick et al., 1973, Molina et al., 2016).

$$Entropy = entro = -\sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P(i,j) \log P(i,j) \qquad\qquad B.14$$

B.1.9 Homogeneity

The homogeneity predictor is used to measure the closeness of the distribution of entries in the GLCM to the diagonal of the probability matrix. The relative increase of the distribution away from the diagonal indicates a lower value of homogeneity. It tends to indicate relatively higher values for homogenous images. The predictor function is defined by Eq. B.15 (Soh and Tsatsoulis, 1999, Yang et al., 2012).

$$Homogeneity = homom = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \frac{1}{1+(i-j)^2} \cdot P(i,j) \qquad\qquad B.15$$

B.1.10 Maximum Probability

The maximum probability predictor is the maximum value of the GLCM. This predictor is used to indicate the most probability of a predominant pixel pair of the GLCM. It is defined by Eq. B.16 (Gómez et al., 2012).

$$Maximum\ probability = maxpr = MAX_{i,j} P(i,j) \qquad\qquad B.16$$

B.1.11 Sum of Squares

This predictor is also known as the variance; it places relatively high weights on the elements that differ from the mean value of the GLCM. It increases when the grey spatial values differ from their averages. Therefore, a higher value of this predictor indicates a higher heterogeneity of texture. The function of this predictor is defined by Eq. B.17 (Haralick et al., 1973, Gnep et al., 2017).

$$Sum\ of\ square = \ sosq = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} (i-u)^2 . P(i,j) \qquad\qquad B.17$$

B.1.12 Sum Average, Sum Entropy, Sum Variance

These predictors are used to show the general indications to reflect the heterogeneity regions of an image. They tend to have relatively low values for low-grade tumours and slightly higher values for a high-grade tumour. The sum average, sum entropy and sum variance are defined by Eqs. B.18-B.20 respectively (Gómez et al., 2012, Pantelis, 2010).

$$sum\ average = savgh = \sum_{i=1}^{2M-1} i.P_{x+y}(i) \qquad\qquad B.18$$

$$Sum\ entropy = \ senth = -\sum_{i=1}^{2M-1} P_{x+y}(i).logP_{x+y}(i) \qquad\qquad B.19$$

$$Sum\ variance = \ svarh = \sum_{i=1}^{2M-1} (1 - sum\ entropy)^2 . P_{x+y}(i) \qquad\qquad B.20$$

where P_{x+y} is defined by Eq. B.21.

$$P_{x+y}(k) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P(i,j) \qquad ,i+j = k \quad and \quad k = 0,1,2,\ldots\ldots.2M \qquad B.21$$

B.1.13 Information Measure of Correlation 1 and Correlation 2

These predictors are used to measure the linear dependency between a grey spatial tone and its neighbours. They are also used to measure deformation in the texture regions. These predictors are defined by Eq. B.22 and Eq. B.23 (Gómez et al., 2012, Gnep et al., 2017).

$$\text{Information measure of correlation } 1 = \ inf1h = \frac{HXY - HXY1}{MAX(HX, HY)} \qquad B.22$$

$$\begin{aligned}
\text{Information measure of correlation } 2 &= inf2h \\
&= (1 - exp[-2(HXY2 - HXY)])^{1/2}
\end{aligned} \qquad B.23$$

Where *HX*, *HY*, *HXY*, *HXY1*, and *HXY2* are defined by Eqs. B.24 - B.28 respectively.

$$HX = -\sum_{i=0}^{M-1} P_x(i).logP_x(i) \qquad B.24$$

$$HY = -\sum_{i=0}^{M-1} P_y(i)P_y(i) \qquad B.25$$

$$HXY = -\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} P(i,j).logP(i,j) \qquad B.26$$

$$HXY1 = -\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} P(i,j).log\left(P_x(i)P_y(j)\right) \qquad B.27$$

$$HXY2 = -\sum_{i=0}^{M-1}\sum_{i=0}^{M-1} P_x(i).P_y(j).log\left(P_x(i)P_y(j)\right) \qquad B.28$$

B.1.14 Inverse Difference Normalised and Inverse Difference Moment Normalised

These predictors are used to reflect the homogeneity of a textural region. The inverse difference normalised indicates the smoothness of the texture. The inverse difference moment normalised is inversely related to both energy and contrast. The functions of these predictors are defined by Eq. B.29 and Eq. B.30 (Gómez et al., 2012, Tantisatirapong, 2015).

$$\text{Inverse difference normalized predictor} = \ indnc = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} \frac{P(i,j)}{1 + \frac{|i-j|}{M}} \qquad B.29$$

$$\text{Inverse difference moment normalized} = \ idmnc = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} \frac{P(i,j)}{1 + \frac{(i-j)^2}{M}} \qquad B.30$$

APPENDIX C

Table C.1 Comparative results of different classifiers incorporated with the full set of features associated with 3DGLCM using BRTAS2018 dataset.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
DT	Class0	43	32	57.33	72.9	64.17	83.16
	Class1	16	194	92.38	85.8	88.99	
LDA	Class0	61	14	81.33	58.7	68.15	80.00
	Class1	43	167	79.52	92.3	85.42	
SVML	Class0	51	24	68.00	77.3	72.34	**86.32**
	Class1	15	195	92.86	89.0	90.90	
SVMQ	Class0	50	25	66.67	73.5	69.93	84.91
	Class1	18	192	91.43	88.5	89.92	
SVMCUB	Class0	46	29	61.33	68.7	64.78	82.46
	Class1	21	189	90.00	86.7	88.31	
SVMG	Class0	41	34	54.67	78.8	64.56	84.21
	Class1	11	199	94.76	85.4	89.84	
KNNF	Class0	43	32	57.33	58.9	58.10	78.25
	Class1	30	180	85.71	84.9	85.30	
KNNM	Class0	45	30	60.00	77.6	67.66	84.91
	Class1	13	197	93.81	86.8	90.16	
KNNCOS	Class0	47	28	62.67	71.2	66.66	83.51
	Class1	19	191	90.95	87.2	89.04	
KNNCUB	Class0	44	31	58.67	73.3	65.18	83.51
	Class1	16	194	92.38	86.2	89.19	
KNNW	Class0	47	28	62.67	73.4	67.62	84.21
	Class1	17	193	91.90	87.3	89.55	
EBTree	Class0	45	30	60.00	71.4	65.21	83.16
	Class1	18	192	91.43	86.5	88.88	
ESDA	Class0	49	26	65.33	67.1	66.21	82.46
	Class1	24	186	88.57	87.7	88.15	

Table C.2 Comparative results of different classifiers incorporated with the selected set of features associated with 3DGLCM after applying ANOVA using BRTAS2018 dataset.

| Classifier | Actual class | Confusion matrices Predicted class | | Sensitivity% | Precision % | F-measure % | Accuracy% |
		Class0	Class1				
DT	Class0	43	32	57.33	72.9	64.17	83.16
	Class1	16	194	92.38	85.8	88.99	
LDA	Class0	61	14	81.33	58.1	67.77	79.65
	Class1	44	166	79.05	92.2	85.12	
SVML	Class0	51	24	68.00	79.7	73.38	**87.02**
	Class1	13	197	93.81	89.1	91.41	
SVMQ	Class0	47	28	62.67	72.3	67.14	83.86
	Class1	18	192	91.43	87.3	89.30	
SVMCUB	Class0	47	28	62.67	69.1	65.73	82.81
	Class1	21	189	90.00	87.1	88.52	
SVMG	Class0	39	36	52.00	78.0	62.40	83.51
	Class1	11	199	94.76	84.7	89.43	
KNNF	Class0	41	34	54.67	59.4	56.94	78.25
	Class1	28	182	86.67	84.3	85.44	
KNNM	Class0	44	31	58.67	75.9	66.16	84.21
	Class1	14	196	93.33	86.3	89.70	
KNNCOS	Class0	47	28	62.67	69.1	65.73	82.81
	Class1	21	189	90.00	87.1	88.52	
KNNCUB	Class0	44	31	58.67	74.6	65.67	83.86
	Class1	15	195	92.86	86.3	89.44	
KNNW	Class0	47	28	62.67	73.4	67.62	84.21
	Class1	17	193	91.90	87.3	89.55	
EBTree	Class0	47	28	62.67	69.1	65.73	82.81
	Class1	21	189	90.00	87.1	88.52	
ESDA	Class0	50	25	66.67	68.5	67.56	83.16
	Class1	23	187	89.05	88.2	88.62	

Table C.3 Comparative results of different classifiers incorporated with the selected set of features associated with 3DGLCM after applying the proposed hybrid features selection method using BRTAS2018 dataset.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity%	Precision %	F-measure %	Accuracy%
		Class0	Class1				
DT	Class0	44	31	58.67	75.9	66.16	84.21
	Class1	14	196	93.33	86.3	89.70	
LDA	Class0	59	16	78.67	57.8	66.66	79.30
	Class1	43	167	79.52	91.3	84.98	
SVML	Class0	55	20	73.33	79.7	76.38	**88.07**
	Class1	14	196	93.33	90.7	92.01	
SVMQ	Class0	47	28	62.67	71.2	66.66	83.51
	Class1	19	191	90.95	87.2	89.04	
SVMCUB	Class0	47	28	62.67	71.2	66.66	83.51
	Class1	19	191	90.95	87.2	89.04	
SVMG	Class0	37	38	49.33	77.1	60.16	82.81
	Class1	11	199	94.76	84.0	89.03	
KNNF	Class0	41	34	54.67	60.3	57.34	78.60
	Class1	27	183	87.14	84.3	85.71	
KNNM	Class0	42	33	56.00	72.4	63.15	82.81
	Class1	16	194	92.38	85.5	88.78	
KNNCOS	Class0	46	29	61.33	71.9	66.18	83.51
	Class1	18	192	91.43	86.9	89.09	
KNNCUB	Class0	44	31	58.67	75.9	66.16	84.21
	Class1	14	196	93.33	86.3	89.70	
KNNW	Class0	44	31	58.67	68.8	63.30	82.11
	Class1	20	190	90.48	86.0	88.16	
EBTree	Class0	49	26	65.33	72.1	68.53	84.21
	Class1	19	191	90.95	88.0	89.46	
ESDA	Class0	49	26	65.33	72.1	68.53	84.21
	Class1	19	191	90.95	88.0	89.46	

Table C.4 Optimal selected set of features by the proposed HFSA with their corresponding angles using BRATS2018 dataset. The names of the features are referred by its abbreviations that were defined in the expression of the equations (APPENDIX B, B.1.1-B.1.14)

Features	Angles	Features	Angles
cprom2d0, dissi2d0, homom2d0 maxpr2d0, sosvh2d0, svarh2d0 inf1h2d0, inf2h2d0, indnc2d0	(0^0)	*autoc3d450, corrm3d450 cprom3d450, cshad3d450 dissi3d450, energ3d450 entro3d450, homom3d450 sosvh3d450, savgh3d450 svarh3d450, senth3d450 indnc3d450*	$(45^0, 0^0)$
autoc2d45, contr2d45, corrm2d45 prom2d45, cshad2d45, energ2d45 entro2d45, homom2d45, sosvh2d45 svarh2d45, senth2d45, inf1h2d45 indnc2d45	(45^0)	*autoc3d-450, contr3d-450 corrm3d-450, cprom3d-450 cshad3d-450, energ3d-450 homom3d-450, sosvh3d-450 savgh3d-450 , svarh3d-450 senth3d-450 , inf1h3d-450 inf2h3d-450*	$(-45^0, 0^0)$
corrm2d90, cprom2d90, dissi2d90 nerg2d90, homom2d90, maxpr2d90, sosvh2d90, inf1h2d90, inf2h2d90 indnc2d90	90^0	*autoc3d4545, contr3d4545 cprom3d4545, cshad3d4545 energ3d4545, entro3d4545 sosvh3d4545, savgh3d4545 svarh3d4545, senth3d4545 idmnc3d4545*	$(45^0, 45^0)$
contr2d135, cprom2d135, cshad2d135, dissi2d135, energ2d135, entro2d135, homom2d135, sosvh2d135 svarh2d135, senth2d135, inf2h2d135	135^0	*autoc3d-45-45, contr3d-45-45 cprom3d-45-45, cshad3d-45-45 energ3d-45-45, entro3d-45-45 homom3d-45-4,5maxpr3d-45-45 sosvh3d-45-45, svarh3d-45-45 senth3d-45-45, indnc3d-45-45 idmnc3d-45-45*	$(-45^0, -45^0)$
autoc3d045, contr3d045 ,cprom3d045cshad3d045, energ3d045, homom3d045 sosvh3d045, svarh3d045, senth3d045, inf1h3d045, idmnc3d045	$(0^0, 45^0)$	*autoc3d45-45, cprom3d45-45 cshad3d45-45, energ3d45-45 entro3d45-45, sosvh3d45-45 savgh3d45-45, svarh3d45-45 senth3d45-45, idmnc3d45-45*	$(45^0, -45^0)$

Table C.4 Continued

corrm3d00,cprom3d00, maxpr3d00	$(0^0,0^0)$	*autoc3d-4545, contr3d-4545* *cprom3d-4545, cshad3d-4545* *dissi3d-4545, energ3d-4545* *entro3d-4545, homom3d-4545* *sosvh3d-4545, savgh3d-4545* *svarh3d-4545, senth3d-4545* *indnc3d-4545, idmnc3d-4545*	$(-45^0,45^0)$
autoc3d0-45,contr3d0-45, *cprom3d0-45, energ3d0-45* *entro3d0-45, homom3d0-45* *sosvh3d0-45, savgh3d0-45* *svarh3d0-45, senth3d0-45* *inf1h3d0-45, indnc3d0-45* *idmnc3d0-45*	$(0^0,-45^0)$		

Table C.5 Comparative evaluation results showing the full set of the proposed features FTD extracted from the tumour descriptors incorporating different machine learning algorithms using the BRATS2018 dataset.

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	56	19	74.67	83.58	78.87	89.47
	Class1	11	199	94.76	91.28	92.99	
LDA	Class0	54	21	72.00	84.38	77.69	89.12
	Class1	10	200	95.24	90.50	92.80	
SVML	Class0	56	19	74.67	91.80	82.35	91.58
	Class1	5	205	97.62	91.52	94.47	
SVMQ	Class0	56	19	74.67	87.50	80.57	90.53
	Class1	8	202	96.19	91.40	93.73	
SVMCUB	Class0	55	20	73.33	83.33	78.01	**89.12**
	Class1	11	199	94.76	90.87	92.77	
SVMG	Class0	53	22	70.67	89.83	79.10	90.18
	Class1	6	204	97.14	90.27	93.57	
KNNF	Class0	53	22	70.67	76.81	73.61	86.67
	Class1	16	194	92.38	89.81	91.07	
KNNM	Class0	52	23	69.33	92.86	79.38	90.53
	Class1	4	206	98.10	89.96	93.84	
KNNCOS	Class0	54	21	72.00	85.71	78.26	89.47
	Class1	9	201	95.71	90.54	93.05	
KNNCUB	Class0	52	23	69.33	94.55	80.00	90.88
	Class1	3	207	98.57	90.00	94.09	
KNNW	Class0	52	23	69.33	86.67	77.03	89.12
	Class1	8	202	96.19	89.78	92.87	
EBTree	Class0	58	17	77.33	86.57	81.69	90.88
	Class1	9	201	95.71	92.20	93.92	
ESDA	Class0	53	22	70.67	89.83	79.10	90.18
	Class1	6	204	97.14	90.27	93.57	

Table C.6 Comparative evaluation results for the selected set of features FTD using the ANOVA method, incorporating different classifiers using the BRATS2018 dataset. The selected features were (Nec_M, tC_M, Nec_R, Edm_R, tC_R).

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	56	19	74.67	83.58	78.87	89.47
	Class1	11	199	94.76	91.28	92.99	
LDA	Class0	54	21	72.00	84.38	77.69	89.12
	Class1	10	200	95.24	90.50	92.80	
SVML	Class0	56	19	74.67	90.32	81.75	91.23
	Class1	6	204	97.14	91.48	94.22	
SVMQ	Class0	58	17	77.33	90.63	83.45	91.93
	Class1	6	204	97.14	92.31	94.66	
SVMCUB	Class0	56	19	74.67	90.32	81.75	**91.23**
	Class1	6	204	97.14	91.48	94.22	
SVMG	Class0	53	22	70.67	89.83	79.10	90.18
	Class1	6	204	97.14	90.27	93.57	
KNNF	Class0	55	20	73.33	76.39	74.83	87.02
	Class1	17	193	91.90	90.61	91.25	
KNNM	Class0	55	20	73.33	93.22	82.08	91.58
	Class1	4	206	98.10	91.15	94.49	
KNNCOS	Class0	53	22	70.67	82.81	76.25	88.42
	Class1	11	199	94.76	90.05	92.34	
KNNCUB	Class0	55	20	73.33	91.67	81.48	91.23
	Class1	5	205	97.62	91.11	94.25	
KNNW	Class0	56	19	74.67	86.15	80.00	90.18
	Class1	9	201	95.71	91.36	93.48	
EBTree	Class0	57	18	76.00	85.07	80.28	90.18
	Class1	10	200	95.24	91.74	93.45	
ESDA	Class0	53	22	70.67	88.33	78.51	89.82
	Class1	7	203	96.67	90.22	93.33	

Table C.7 Comparative evaluation results for the selected set of features by the proposed HFSA, incorporating different classifiers using the BRATS2018 dataset. The selected features were (*tC_R, Edm_R, Nec_M, and Nec_R*).

Classifier	Actual class	Confusion matrices		Sensitivity %	Precision %	F-measure %	Accuracy %
		Predicted class					
		Class0	Class1				
DT	Class0	56	19	74.67	83.58	78.87	89.47
	Class1	11	199	94.76	91.28	92.99	
LDA	Class0	53	22	70.67	84.13	76.81	88.77
	Class1	10	200	95.24	90.09	92.59	
SVML	Class0	55	20	73.33	90.16	80.88	90.88
	Class1	6	204	97.14	91.07	94.00	
SVMQ	Class0	58	17	77.33	89.23	82.85	91.58
	Class1	7	203	96.67	92.27	94.41	
SVMCUB	Class0	58	17	77.33	96.67	85.92	**93.33**
	Class1	2	208	99.05	92.44	95.63	
SVMG	Class0	50	25	66.67	90.91	76.92	89.47
	Class1	5	205	97.62	89.13	93.18	
KNNF	Class0	53	22	70.67	80.30	75.17	87.72
	Class1	13	197	93.81	89.95	91.84	
KNNM	Class0	53	22	70.67	92.98	80.30	90.88
	Class1	4	206	98.10	90.35	94.06	
KNNCOS	Class0	56	19	74.67	78.87	76.71	88.07
	Class1	15	195	92.86	91.12	91.98	
KNNCUB	Class0	54	21	72.00	91.53	80.59	90.88
	Class1	5	205	97.62	90.71	94.03	
KNNW	Class0	54	21	72.00	87.10	78.83	89.82
	Class1	8	202	96.19	90.58	93.30	
EBTree	Class0	57	18	76.00	87.69	81.42	90.88
	Class1	8	202	96.19	91.82	93.95	
ESDA	Class0	51	24	68.00	86.44	76.11	88.77
	Class1	8	202	96.19	89.38	92.66	

Table C.8 Comparative evaluation results of the proposed MTMCS based on one-layer NNs versus all other classification approaches and the majority vote. This test is based on the full set of features of 3DGLCM and FTD that derived from BRATS2018 dataset.

Classifier	Actual class	Confusion matrices Predicted class		Sensitivity %	Precision %	F-measure %	Accuracy %
		Class0	Class1				
DT	Class0	59	16	78.67	84.3	81.37	90.53
	Class1	11	199	94.76	92.6	93.64	
LDA	Class0	61	14	81.33	59.2	68.53	80.35
	Class1	42	168	80.00	92.3	85.71	
SVML	Class0	60	15	80.00	92.3	85.71	92.98
	Class1	5	205	97.62	93.2	95.34	
SVMQ	Class0	59	16	78.67	86.8	82.51	91.23
	Class1	9	201	95.71	92.6	94.14	
SVMCUB	Class0	58	17	77.33	81.7	79.45	89.47
	Class1	13	197	93.81	92.1	92.92	
SVMG	Class0	47	28	62.67	85.5	72.30	87.37
	Class1	8	202	96.19	87.8	91.81	
KNNF	Class0	54	21	72.00	72.0	72.00	85.26
	Class1	21	189	90.00	90.0	90.00	
KNNM	Class0	51	24	68.00	81.0	73.91	87.37
	Class1	12	198	94.29	89.2	91.66	
KNNCOS	Class0	53	22	70.67	79.1	74.64	87.37
	Class1	14	196	93.33	89.9	91.58	
KNNCUB	Class0	49	26	65.33	81.7	72.59	87.02
	Class1	11	199	94.76	88.4	91.49	
KNNW	Class0	52	23	69.33	80.0	74.28	87.37
	Class1	13	197	93.81	89.5	91.62	
EBTree	Class0	57	18	76.00	83.8	79.72	89.82
	Class1	11	199	94.76	91.7	93.20	
ESDA	Class0	59	16	78.67	80.8	79.72	89.47
	Class1	14	196	93.33	92.5	92.89	
Majority Vote	Class0	56	19	74.67	86.2	80.00	90.18
	Class1	9	201	95.71	91.4	93.48	
Proposed MTMCS	Class0	65	10	86.67	94.2	90.27	**95.09**
	Class1	4	206	98.10	95.4	96.71	

APPENDIX D

D.1 Implementation and Further Technical Details

This includes the technical steps and implementation process. Matlab 2018 student licence has been used to write the code, build the function, and process the implementation. Matlab learner application is used initially to generate the code for learning models then the code has been modified and updated. The Matlab learner application is limited to accomplish the training and testing for more than 50 samples based on LOO-cross validation, while the code has been developed to adapt any number of samples by LOO- cross-validation. SPSS IBM software VER. 24 is utilised for measuring ANOVA and Pearson correlation techniques. Dia software is used as a tool to generate the UML graphical design of the developed system. UML class diagram is used to add a further illustration to the implementation details of the developed methods in the thesis. The results and the outcome of the methods developed in the research work are presented in each chapter of the thesis. The results are also summarised and discussed in chapter 7.

D.2 System Validation

In this research work, four MR images datasets are used to evaluate and validate the novel methods proposed in this work. Two stages of learning were established to develop the automated classification system for glioma grades. To validate the developed system, leave-one-out cross-validation technique has been utilised in all stages. In the first stage of the proposed MTMCS, the dataset was divided into two parts; these are the training part that is used to train the classification model and the testing part that is used to test the trained model. In this stage the dataset was splitted into two parts to validate the trained model on new data that has not been seen during the training phase. The reason behind splitting the data into two parts only is that all single classifiers involved were developed based on fixed parameters and there is no any tuning process is conducted for the parameters for all experiments, and therefore there is no need to divide the data into additional part that is required to optimise the model before testing it. This is different from the second stage of the MTMCS where the dataset is divided into three parts namely training subset, validation subset and testing subset. The training subset is used to train the model, while the validation subset is utilised to optimise the parameters of the DNNs such as weights and biases, and to select the best design choices of the DNN, which has first divergence from the global

minimum and thereby selecting the best DNN model. Then the optimised model is used to test the testing subset to evaluate the results of the developed classification system.

D.3 MRI-Preparation

This stage is implemented to prepare the ROI from MR brain tumour images, which includes bringing in ROI and cropping the images to provide image with only ROI of brain tumour for the next stages. Each patient in the dataset has two sets of images; the first set is the original MRI series and the second set is the labelled MRI series. The labelled MRI images, which is denoted in this work as the label identification layer (SEG), has four labels to indicate four tumour regions (Figure 2.3). In the labelled image, each pixel is labelled and pre-assigned by the dataset as follows:

Label '0' indicates background.

Label '1' indicates tumour necrosis.

Label '2' indicates tumour edema.

Label '3' indicates tumour non-contrast enhancement.

Label '4' indicates tumour enhancement.

The label identification layer is utilised to bring in only the ROI of brain tumour images. To elaborate, the ROI represents the active portion of the tumour, which includes necrosis, enhanced, and non-enhanced tumour. To achieve this, masking process is applied which is designed to bring in only tumour portions that include only tumour necrosis, non-contrast enhancement, and enhancement and discard edema portion. This is implemented by designing additional images matrix and is called a mask. This mask is designed to include only logical values (only ones and zeros) and depending on the label identification layer to bring in only ROI. This mask have ones corresponding to the pixels that have label '1', label '3', and label '4' and zeros to other regions of the image. Then a logical multiplication between this mask and the original MRI image is determined. As a result images with only the ROI are obtained.

The next step is to eliminate the redundant zeros-pixel located in column and rows out of the ROI. This is achieved through applying automated search process starts from four directions inside the image, then a comparison between each two neighbouring columns for searching the x-axis is conducted and the same process is repeated for the y-axis. If any difference in the values from this comparison is captured, the search process is terminated on the column and rows larger than the area of the ROI by one column and row. As a result, the cropped image with only ROI is produced (Figure 3.2).

D.4 Features Extraction

This stags is dedicated to extract the image features from the ROI brain tumour image. These features include the textural features and the new features associated with tumour descriptors (FTD). For the texture extraction, the implementation starts with building the co-occurrence matrix from each ROI acquired from T2-MR series of each patient (section 3.4.1). This construction of GLCM is based on 13 angles, distance equal to 1, and the quantisation level is (0 to 255). Thirteen angles are used, result in thirteen matrices are produced after implementing the GLCM. From each GLCM matrix, 18 statistical predictors are measured; the formulas of these predictors are presented in (APPENDIX B, B.1.1-B.1.14). After the predictors are implemented, each patient is represented by 18 measures × 13 matrices, and hence each patient is represented by 234 feature values. The first four angles of GLCM are used to represent the 2DGLCM, while all the 13 angles of GLCM are utilized to indicate the 3DGLCM. At the final step of feature extraction, the output feature vectors are saved in excel file, and the corresponding grade target (the truth label for each patient) is also assigned to each vector in same excel sheet. Label 0 and 1 were also assigned to the feature vectors, indicating low and high grades samples respectively and according to the pre-diagnosis information provided by the dataset. The contents of the target vector are then used in the training phase of the supervised classification process where both the target vector and the feature vectors are provided to learn the classification model. The Target vector is also used in the evaluation process through comparing it by the output probability produced by the classification model during the testing phase.

Regarding the extraction of the new features proposed in chapter 4, the areas of the presence of the tumour descriptors were measured by utilizing the labelled pixel of the input image and calculating the total number of the labelled pixels for each labelled region in all images. For examples to measure the presence area of tumour necrosis, a search process is conducted to measure the total number of image pixels assigned by labelled '1'. Similarly to get the presence of tumour edema, the number of pixels labelled with '2' is accounted. The same strategy is applied to calculate other tumour presences non-enhanced and enhanced tumour. After that the total number is divided on the number of images that have ROI to produce the features *Nec_M, Edm_M, tnC_M, tC_M*. New relationships are constructed from these areas using Eqs 4.1-4.5. As a results eight features are produced to present each patient in the dataset, then these features are also saved in the excel file and assigned with grade labels (Target).

Algorithm steps for feature extraction:

Step 1:   *Start*

Step 2:   *Read ROI images*

Step 3:   *Normalize intensity of image to the range [0 to 255]*

Step 4:   *Build the Gray Level Co-Occurrence Matrix for the image*

Step 5:   *Utilising the values from the matrix, measure the textural predictors using Eqs.*
            (APPENDIX B, B.1.1-B.1.14)

Step 6:   *Given the labelled pixels of images, compute the FTD features using Eqs (4.1-*
            4.5)

Step 7:   *Repeat the steps from Step 2 to Steps 5 to list out the features for all images in*
            the dataset

Step 8:   *End*

D.5 Building the Classification Model

Several single and ensemble classification models were built and implemented in this work, including eleven single classifiers and three ensemble classification systems. The same configuration and design choices of the classification models (Table 3.1) were applied for the four datasets. The reason behind this is to achieve consistent performance and robust comparison between the models evaluated on different datasets. Building the model starts with configuration the classification models then feeding the data into the models to implement the training phase. Then the trained model is validated using LOO cross-validation technique.

Algorithm steps for feature classification:

Step 1:   *Start*

Step 2:   *Read extracted features of input data*

Step 3:   *For i=1 to: the number of patients*

Step 4:   *Remove the features of one patient from the input data and store it in the testing Set*

Step 5:   *Train the classification model*

Step 6:   *Test the trained model using the features stored in the testing set*

Step 7:   *Store the output decision of testing the model in A*

Step 8:   *Add the removed sample to the training set and chose a different one*

Step 9:   *End for*

Step 10:   *Compute Confusion Matrix*

Step 11:   *End*

D.6 Unified Modelling Language

The Unified Modelling Language (UML) is a tool to draw the code classes, methods, attributes and the relationship between classes. Class diagram is a graphical representation to show the relationship between classes. Dia software is used as a visualisation tool to draw the UML diagram for the code classes of all the methods developed in the thesis. Particularly, the classes show the implementation procedure, which are explained as follows:

D.6.1 Proposed Hybrid Features Selection Algorithm

The UML diagram of the proposed Hybrid Features Selection Algorithm (HFSA) that is used to select the most significant features (Figure D.1) is started with the initialisation stage that is explained in details in section 3.5. The UML diagram has the ***Main Class*** that represents the starting point of the implementation process and it has all initial values for reading the input data. The input data represents the feature vectors that are extracted from MR images. These vectors are arranged in a matrix or excel sheet. As a result each column represents a statistical predictor, and each row indicates a patient. The last column indicates the target list. In this class, the input feature is splitted into two lists. The first list includes the *TargetData* that represents the ground truth label of glioma grades, and the second list is *inputData* that includes the input feature vectors. This class is designed to call all other classes, and return the results from other classes that are explored as follows:

1. ***Splitting_features***: The ***Main Class*** call this class and it is used to split the features through removing one feature (*Remove_oneFeature*) and then the produced features are saved in (*NewSetofFeatures*). The feature elimination starts with the one that has highest correlation with the reference predictor and then this process is repeated in descending order to exam all features vectors in the input data.

307
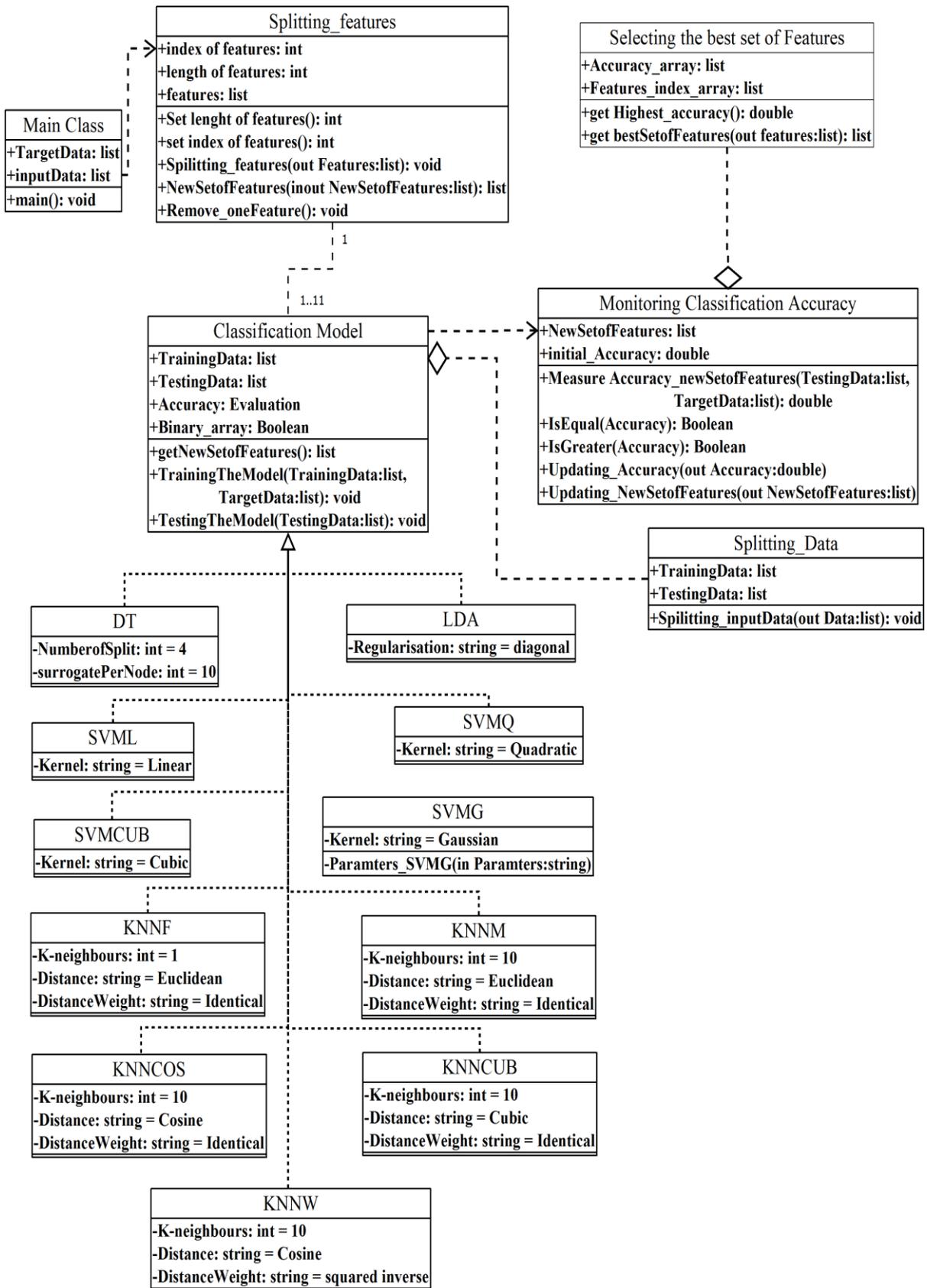
Figure D.1 Implementation of UML class diagram for the hybrid feature selection
algorithm using Dia software tools.

2. ***Classification Model*:** This class is utilised to implement, train and test eleven classification models using different subset of features where the ***Splitting_features*** class is connected with one or more classification models in this class. This class and through each classification models use the class ***splitting_Data*** that is built to split the input data into two subsets these are the *TrainingData* and the *TestingData* where this splitting of the dataset are demonstrated based on LOO cross-validation technique. The *TrainingData* and the *TargetData* are used to train the classification model while the trained model is then used to test the *TestingData*. Afterwards the classification performance is evaluated by calculating the confusion matrix that can be used to determine the evaluation metrics such as the classification accuracy, sensitivity and specificity. Eleven classifiers are implementing this class through train and test each classification models. The classification models are represented by the classes: ***DT, LDA, SVML, SVMQ, SVMCUB, SVMG, KNNF, KNNM, KNNCOS, KNNCUB and KNNW***.

3. ***Monitoring the Classification Accuracy*:** this class is called by the ***Classification Model*** class, it is responsible for tracking the classification accuracy based on examining different subset of features through the implementation of different classification models. The classification accuracy is compared and updated for each classifier and then the new subset of features (*Updating_NewSetofFeatures*) is also updated according the results of the tests. This class has the ***Selecting the best set of Features*** class that is used to pick up the highest classification accuracy and the best subset of features that reflects that maximum classification accuracy. The method is ended when the full length of features is examined; afterwards the highest classification accuracy with the corresponding subset of features is selected.

D.6.2 First Stage of the Proposed MTMCS

The UML diagram of the first stage of the proposed MTMCS is illustrated in Figure D.2, the classes related to train, test the classification models have followed the same methodology and class diagram presented in Figure D.1 but without the classes designed for the features selection. The *Classification Model* class is called by the *Main Class* and it is used to implement, train, test and evaluate the eleven classification models. This class also has the class of *Splitting_Data* and the class of *Generating Output Decision matrix_ODM*. The latter class is utilised to generate the output decisions (ODs) from testing the trained models using the *TestingData* and then saving them in new *Binary_Matrix*. The ODs are binary numbers that includes 1 and 0 indicate high and low grades respectively. The *TargetData* is assigned in the *Binary_Matrix* to represent the Target of the samples and to prepare them for the next stage of learning. The output of this stage is the output decision matrix ODM (*ODM_data*) that has the data associated with patient's index arranged in row-wise and the data produced from the classifiers arranged in column wise. The size of this matrix is the number of patient's × the number of classifiers.

D.6.3 Second Stage of the Proposed MTMCS

The UML diagram of the second stage of the proposed MTMCS is depicted in Figure D.3. In the *Main Class* of this diagram, the process starts by reading ODM (*ODM_data*) generated by the first stage then this data is splitted into two lists: the *TargetData* and *inputData*. Further details of the classes are explored as follows:

1. *Neural Network Model*: this class is called by the *Main Class*. This class is used to build, train and validate the NN model. In this class, the backpropagation method (Matlab function: trainscg) is implemented to optimise the weights and biases of the NN network through training the NN model using the *TrainingData* and *TargetData* and validating the model through using the *ValidationData* and *TargetData*. This class is connected with the *Network Layer* class that is used to get number of layer. The *Network Layer* is connected with the *Network Neuron* class that is utilised to provide number of neurons per layer and to activate each neuron by the activation Matlab function (tansig). The *Neural Network Model* class is also connected with the *Network Iteration* class to assign number of iterations for the NN model.

310
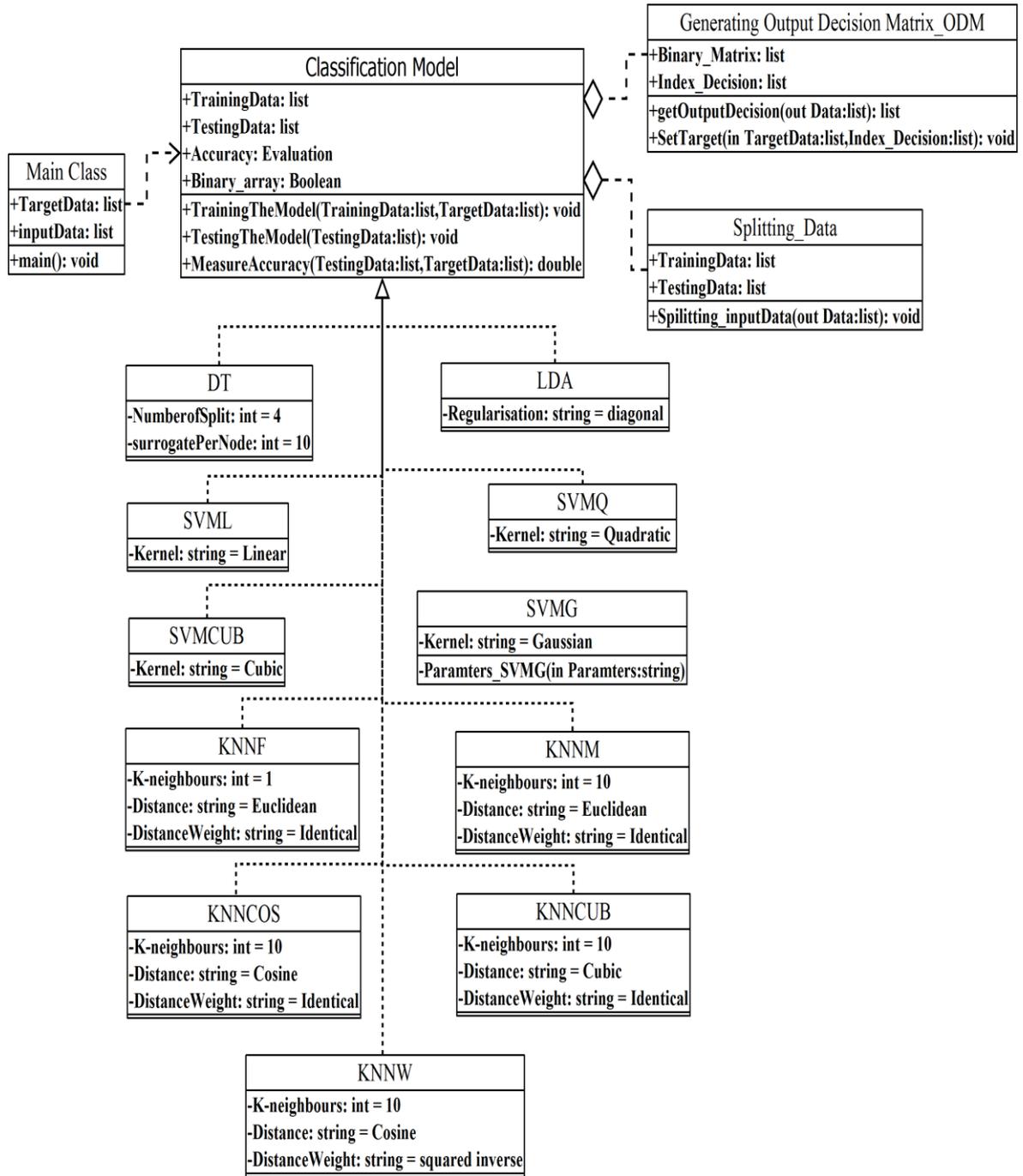
Figure D.2 Implementation of UML class diagram for the first stage of the MTMCS using Dia software tools.
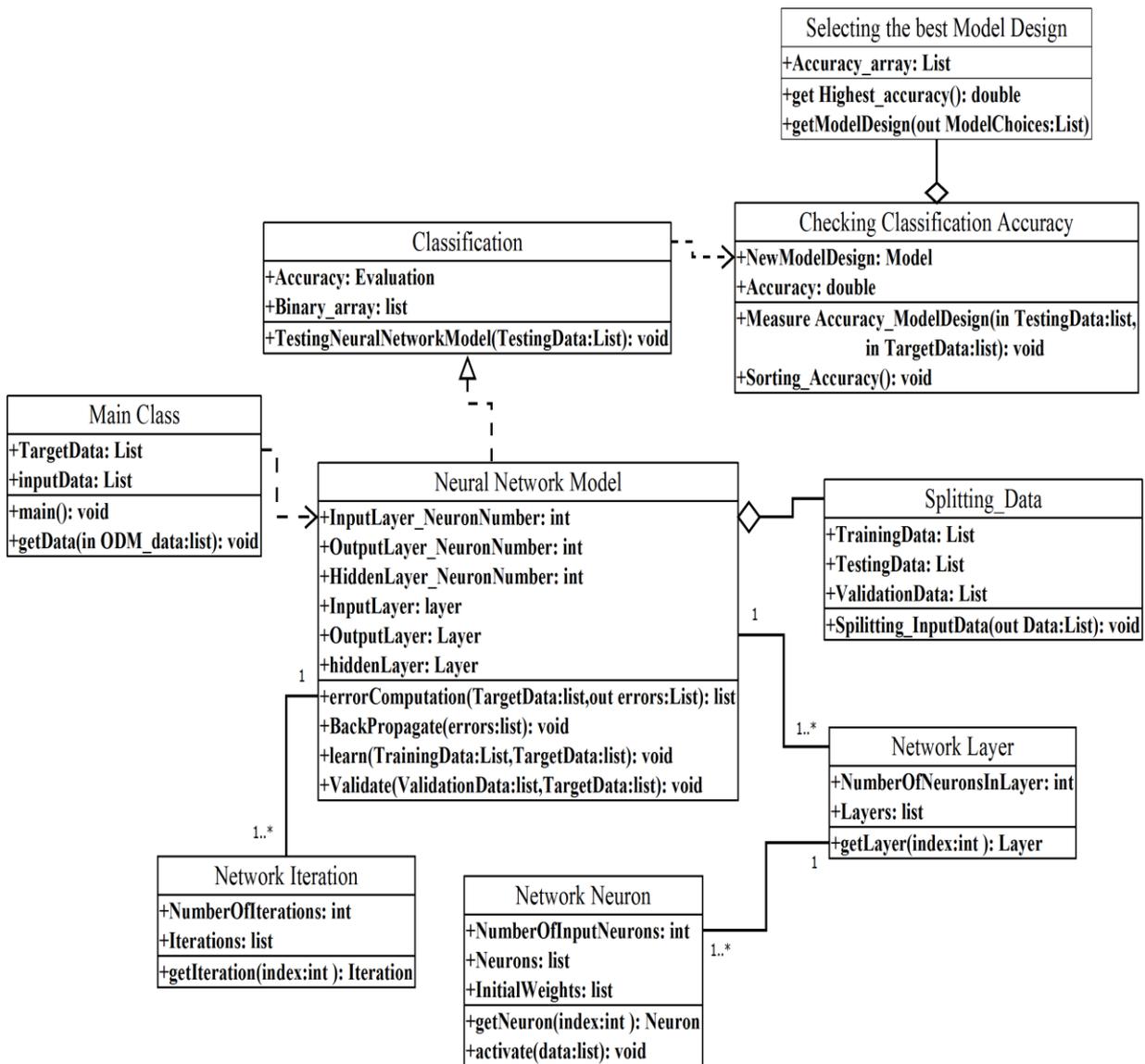
Figure D.3 Implementation of UML class diagram for the second stage of the MTMCS using Dia software tools.

2. **Splitting_Data:** the **Neural Network Model** class uses this class to split the input data into three different subsets: the *TrainingData, ValidationData* and *TestingData*. Two subsets are used in the training phase where the training data is splitted into the *TrainingData* of 85% and *ValidationData* of 15%. The testing phase uses LOO technique whereby the trained model is tested using new samples that are not included in the training phase.

3. **Classification**: the **Neural Network Model** class is implementing this class to test the trained model using the TestingData, and then results are evaluated.

4. ***Checking Classification Accuracy***: this class is called by the ***Classification*** class and it is used to measure the classification accuracy using the *TestingData* and *TargetData*. This class has the ***Selecting the best Model design*** class that is used to select model design that reflects the highest classification accuracy.

D.6.4 Proposed Algorithm for Selecting the Best Set of Classifiers

The UML diagram of the proposed algorithm for selecting the best set of classifiers (SCA) is presented in Figure D.4. Further details of diagram design are explained as follows:

1. ***Main Class***: the input data to this class is the ODM. Starting with splitting the input data into two lists: the *TargetData* and the *inputData*. This class is started with call the ***Spliting_Classifiers*** class.

2. ***Splitting_Classifiers:*** this class is used to generate new set of classifiers based on removing classifiers one by one. The direction of this process is implemented automatically from the classifier that has lowest classification accuracy and moving through examine all classifiers until reaching the highest one (section 5.5). As a results a new set of classifiers is produced.

3. ***MTMCS***: the ***Main Class*** call this class, it includes two stages of learning and the UML of this class is presented in Figure D.2 and Figure D.3. In this class, the classification model is trained, tested and evaluated based on updating different subset of classifiers. The ***Splitting_Classifiers*** class is providing this class with the new subset of classifiers.

4. ***Monitoring Classification Accuracy:*** this class is called by the ***MTMCS*** class and it is used to measure the classification accuracy based the new subset of classifiers. This class also includes comparing the classification accuracy with the initial or the previous state. It has the ***Selecting the best set of classifiers*** class, which is designed to find the best subset of classifiers that reflects the maximum classification accuracy.
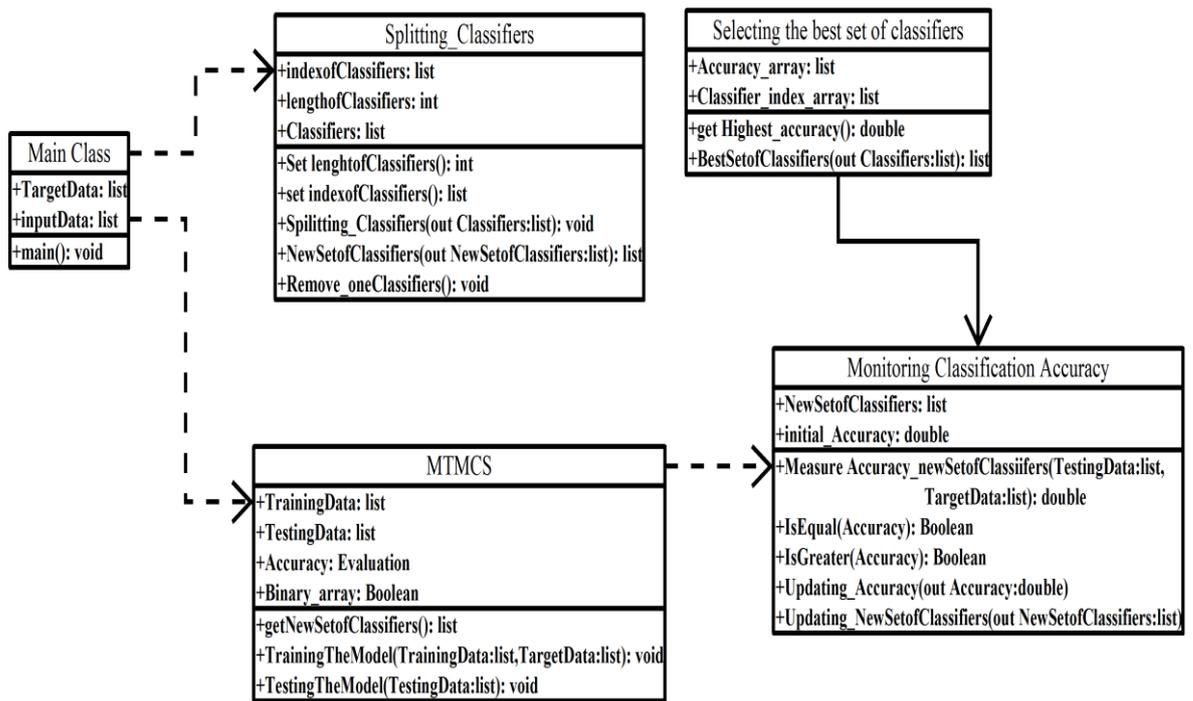
Figure D.4 Implementation of UML class diagram for selecting the best set of classifiers algorithm (SCA) using Dia software tools.

APPENDIX E

The following code has been generated by the author to implement the proposed HFSA and the classification models. The code includes the Matlab functions with further development and modifications. The code of the classification models has been developed to adapt any number of samples in the implementation of LOO-cross validation. The classification models have been designed to automatically implement the training, testing, evaluation phases, and measuring the final classification results based on LOO-cross validation technique. An additional option has been added in establishing the classification model by providing a trained model using the whole dataset. If this option is used, the input dataset needs to be divided in advance into training and testing subsets, and then the training subset is delivered to the code, thereafter the testing phase is conducted to fulfil the cross-validation technique.

```
A=FS1; % FS1 represents the selected index of features from the initial stage
A1=A;
%IC is the initial accuracy set experimentally by testing the full set of features
maxacc_1=IC;
For i=1:length(FS1-1)
AN=A(end-i+1)
A1(A1==AN)=[];
% BT represents the input data with the full set of features
Bm=BT;
Bm=Bm(:,[A1,end]);
BIN=Bm;
 %% 1
rng('default')
[ACC,cm,ind,Label_out,classificationTree_ALL]=DT_CROSSV(BIN);
DT(i).ACC=ACC;
DT(i).confusion=cm;
DT(i).indx=ind;
DT(i).labelout=Label_out;
DT(i).model=classificationTree_ALL;
%% 2
rng('default')
[ACC,cm,ind,Label_out,classificationLDA_ALL]=LDA_CROSSV(BIN);
LDA(i).ACC=ACC;
LDA(i).confusion=cm;
LDA(i).indx=ind;
LDA(i).labelout=Label_out;
LDA(i).model=classificationLDA_ALL;
%% 3
rng('default')
[ACC,cm,ind,Label_out,classificationSVML_ALL]=SVML_CROSSV(BIN);
```

```matlab
SVML(i).ACC=ACC;
SVML(i).confusion=cm;
SVML(i).indx=ind;
SVML(i).labelout=Label_out;
SVML(i).model=classificationSVML_ALL;
%% 4
rng('default')
[ACC,cm,ind,Label_out,classificationQA_ALL]=SVMQA_CROSSV(BIN);
SVMQA(i).ACC=ACC;
SVMQA(i).confusion=cm;
SVMQA(i).indx=ind;
SVMQA(i).labelout=Label_out;
SVMQA(i).model=classificationQA_ALL;
%% 5
rng('default')
[ACC,cm,ind,Label_out,classificationSVMCUB_ALL]=SVMCUB_CROSSV(BIN);
SVMCUB(i).ACC=ACC;
SVMCUB(i).confusion=cm;
SVMCUB(i).indx=ind;
SVMCUB(i).labelout=Label_out;
SVMCUB(i).model=classificationSVMCUB_ALL;
%% 6
rng('default')
[ACC,cm,ind,Label_out,classificationSVMG_ALL]=SVMG_CROSSV(BIN);
SVMG(i).ACC=ACC;
SVMG(i).confusion=cm;
SVMG(i).indx=ind;
SVMG(i).labelout=Label_out;
SVMG(i).model=classificationSVMG_ALL;
%% 7
rng('default')
[ACC,cm,ind,Label_out,classificationKNNF_ALL]=KNNF_CROSSV(BIN);
KNNF(i).ACC=ACC;
KNNF(i).confusion=cm;
KNNF(i).indx=ind;
KNNF(i).labelout=Label_out;
KNNF(i).model=classificationKNNF_ALL;
%% 8
rng('default')
[ACC,cm,ind,Label_out,classificationKNNM_ALL]=KNNM_CROSSV(BIN);
KNNM(i).ACC=ACC;
KNNM(i).confusion=cm;
KNNM(i).indx=ind;
KNNM(i).labelout=Label_out;
KNNM(i).model=classificationKNNM_ALL;
%% 9
rng('default')
[ACC,cm,ind,Label_out,classificationKNNW_ALL]=KNNW_CROSSV(BIN);
KNNW(i).ACC=ACC;
KNNW(i).confusion=cm;
```

```
KNNW(i).indx=ind;
KNNW(i).labelout=Label_out;
KNNW(i).model=classificationKNNW_ALL;
%% 10
rng('default')
[ACC,cm,ind,Label_out,classificationKNNCOS_ALL]=KNNCOS_CROSSV(BIN);
KNNCOS(i).ACC=ACC;
KNNCOS(i).confusion=cm;
KNNCOS(i).indx=ind;
KNNCOS(i).labelout=Label_out;
KNNCOS(i).model=classificationKNNCOS_ALL;
%% 11
rng('default')
[ACC,cm,ind,Label_out,classificationKNNCUB_ALL]=KNNCUB_CROSSV(BIN);
KNNCUB(i).ACC=ACC;
KNNCUB(i).confusion=cm;
KNNCUB(i).indx=ind;
KNNCUB(i).labelout=Label_out;
KNNCUB(i).model=classificationKNNCUB_ALL;
LIST(i,:)=[DT(i).ACC,LDA(i).ACC,SVML(i).ACC,SVMQA(i).ACC,SVMCUB(i).ACC,S
VMG(i).ACC,...
KNNF(i).ACC,KNNM(i).ACC,KNNCOS(i).ACC,KNNCUB(i).ACC,KNNW(i).ACC];
list=LIST(i,:)
maxacc_new=max(list(:))
DT(i).max_res=maxacc_new;
DT(i).listA=list;
%%
if (maxacc_new < maxacc_1)% test [ if max_old - max_new  > 0]
B_active(i)=A(end-i+1)
B_active(B_active==0)=[];
A2=A1(1:end-length(B_active)+1);
A3=A(end-i+1);
A1=[A2,A3, A1(end-length(B_active)+2:end)];
else
maxacc_1=maxacc_new;
B_del(i)= A(end-i+1)
B_del(B_del==0)=[];
end
end
%% the supported functions in single files for classification models----------
% each model train, test and evaluate the input data
%% ---- DT----------------------------------------------------------------
function[ACC,cm,ind,Label_out,classificationTree_ALL]=DT_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;
Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[]; %
response=trainingData(:,end);
```

```matlab
predictors=trainingData(:,1:end-1);
rng('default')

% Train a classifier
% This code specifies all the classifier parameters and trains the classifier.
classificationTree = fitctree(...
    predictors, ...
    response, ...
    'SplitCriterion', 'gdi', ...
    'MaxNumSplits', 4, ...
    'Surrogate', 'off', ...
    'ClassNames', [0; 1]);
[label,score,cost] = predict(classificationTree,Xnew)
Label_out(i).label=label;
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
trainingData=Bm2;
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%% training for the whole dataset,
% THIS IS an additional option to train all samples
classificationTree_ALL = fitctree(...
    predictors, ...
    response, ...
    'SplitCriterion', 'gdi', ...
    'MaxNumSplits', 4, ...
    'Surrogate', 'off', ...
    'ClassNames', [0; 1]);
end
%% LDA
function[ACC,cm,ind,Label_out,classificationDiscriminant_ALL]=LDA_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;
Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[];
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
rng('default')
classificationDiscriminant = fitcdiscr(...
    predictors, ...
    response, ...
    'DiscrimType', 'diagLinear', ...
    'FillCoeffs', 'off', ...
```

318

```matlab
    'SaveMemory', 'on', ...
    'ClassNames', [0; 1]);

[label,score,cost] = predict(classificationDiscriminant,Xnew)
Label_out(i).label=label;
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
%% Training for the whole dataset
trainingData=Bm2;
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%%  built the model
rng('default')
rng(1);
classificationDiscriminant_ALL = fitcdiscr(...
    predictors, ...
    response, ...
    'DiscrimType', 'diagLinear', ...
    'FillCoeffs', 'off', ...
    'SaveMemory', 'on', ...
    'ClassNames', [0; 1]);
end
%% SVML
function[ACC,cm,ind,Label_out,classificationSVM_All]=SVML_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;

Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[];
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
rng('default')
%% built the model
% Train a classifier
% This code specifies all the classifier options and trains the classifier.
classificationSVM = fitcsvm(...
    predictors, ...
    response, ...
    'KernelFunction', 'linear', ...
    'PolynomialOrder', [], ...
    'KernelScale', 'auto', ...
    'BoxConstraint', 1, ...
    'Standardize', true, ...
```

```
      'ClassNames', [0; 1]);

[label,score,cost] = predict(classificationSVM,Xnew)
Label_out(i).label=label;
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
%% Training for the whole dataset
trainingData=Bm2;
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%%  built the model
rng('default')
rng(1);
classificationSVM_All = fitcsvm(...
   predictors, ...
   response, ...
   'KernelFunction', 'linear', ...
   'PolynomialOrder', [], ...
   'KernelScale', 'auto', ...
   'BoxConstraint', 1, ...
   'Standardize', true, ...
   'ClassNames', [0; 1]);
end
%% SVMQA
function[ACC,cm,ind,Label_out,classificationSVM_ALL]=SVMQA_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;
Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[];
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
rng('default')
classificationSVM = fitcsvm(...
   predictors, ...
   response, ...
   'KernelFunction', 'polynomial', ...
   'PolynomialOrder', 2, ...
   'KernelScale', 'auto', ...
   'BoxConstraint', 1, ...
   'Standardize', true, ...
   'ClassNames', [0; 1]);
[label,score,cost] = predict(classificationSVM,Xnew)
Label_out(i).label=label;
```

```matlab
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
%% Training for the whole dataset
trainingData=Bm2;
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%%  built the model
rng('default')
rng(1);
classificationSVM_ALL = fitcsvm(...
    predictors, ...
    response, ...
    'KernelFunction', 'polynomial', ...
    'PolynomialOrder', 2, ...
    'KernelScale', 'auto', ...
    'BoxConstraint', 1, ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
end
%% SVMCUB
function[ACC,cm,ind,Label_out,classificationSVM_ALL]=SVMCUB_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;
Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[];
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
rng('default')
classificationSVM = fitcsvm(...
    predictors, ...
    response, ...
    'KernelFunction', 'polynomial', ...
    'PolynomialOrder', 3, ...
    'KernelScale', 'auto', ...
    'BoxConstraint', 1, ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
[label,score,cost] = predict(classificationSVM,Xnew);
Label_out(i).label=label;
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
```

```matlab
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
trainingData=Bm2;
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%%  built the model
rng('default')
rng(1);
classificationSVM_ALL = fitcsvm(...
    predictors, ...
    response, ...
    'KernelFunction', 'polynomial', ...
    'PolynomialOrder', 3, ...
    'KernelScale', 'auto', ...
    'BoxConstraint', 1, ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
end
%% SVMG
function[ACC,cm,ind,Label_out,classificationSVM_ALL]=SVMG_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;
Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[];
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%%  built the model
rng('default')
classificationSVM = fitcsvm(...
    predictors, ...
    response, ...
    'KernelFunction', 'gaussian', ...
    'PolynomialOrder', [], ...
    'KernelScale', 'auto', ...
    'BoxConstraint', 1, ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
[label,score,cost] = predict(classificationSVM,Xnew)
Label_out(i).label=label;
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
trainingData=Bm2;
```

```matlab
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%% built the model
rng('default')
rng(1);
classificationSVM_ALL = fitcsvm(...
    predictors, ...
    response, ...
    'KernelFunction', 'gaussian', ...
    'PolynomialOrder', [], ...
    'KernelScale','auto' , ...
    'BoxConstraint', 1, ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
end
%% KNNF
function[ACC,cm,ind,Label_out,classificationKNN_ALL]=KNNF_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;
Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[];
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
rng('default')
classificationKNN = fitcknn(...
    predictors, ...
    response, ...
    'Distance', 'Euclidean', ...
    'Exponent', [], ...
    'NumNeighbors', 1, ...
    'DistanceWeight', 'Equal', ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);

[label,score,cost] = predict(classificationKNN,Xnew)
Label_out(i).label=label;
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
%% Training for the whole dataset
trainingData=Bm2;
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%% built the model
```

```matlab
rng('default')
rng(1);
classificationKNN_ALL = fitcknn(...
    predictors, ...
    response, ...
    'Distance', 'Euclidean', ...
    'Exponent', [], ...
    'NumNeighbors', 1, ...
    'DistanceWeight', 'Equal', ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
end
%% KNNM
function[ACC,cm,ind,Label_out,classificationKNN_all]=KNNM_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;
Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[];
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
rng('default')
classificationKNN = fitcknn(...
    predictors, ...
    response, ...
    'Distance', 'Euclidean', ...
    'Exponent', [], ...
    'NumNeighbors', 10, ...
    'DistanceWeight', 'Equal', ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
[label,score,cost] = predict(classificationKNN,Xnew)
Label_out(i).label=label;
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
%% Training for the whole dataset
trainingData=Bm2;
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%%  built the model
rng('default')
rng(1);
classificationKNN_all = fitcknn(...
    predictors, ...
```

```matlab
    response, ...
    'Distance', 'Euclidean', ...
    'Exponent', [], ...
    'NumNeighbors', 10, ...
    'DistanceWeight', 'Equal', ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
end
%%
function[ACC,cm,ind,Label_out,classificationKNN_all]=KNNCUB_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;
Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[];
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
rng('default')
classificationKNN = fitcknn(...
    predictors, ...
    response, ...
    'Distance', 'Minkowski', ...
    'Exponent', 3, ...
    'NumNeighbors', 10, ...
    'DistanceWeight', 'Equal', ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
[label,score,cost] = predict(classificationKNN,Xnew)
Label_out(i).label=label;
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
%% Training for the whole dataset
trainingData=Bm2;
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
%%  built the model
rng('default')
classificationKNN_all = fitcknn(...
    predictors, ...
    response, ...
    'Distance', 'Minkowski', ...
    'Exponent', 3, ...
    'NumNeighbors', 10, ...
    'DistanceWeight', 'Equal', ...
```

```matlab
    'Standardize', true, ...
    'ClassNames', [0; 1]);
end
function[ACC,cm,ind,Label_out,classificationKNN_ALL]=KNNCOS_CROSSV(BIN)
Bm2=BIN;
rng('default')
parfor i=1:size(Bm2,1)
trainingData=Bm2;
Xnew=trainingData(i,1:end-1);
trainingData(i,:)=[];
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
rng('default')
classificationKNN = fitcknn(...
    predictors, ...
    response, ...
    'Distance', 'Cosine', ...
    'Exponent', [], ...
    'NumNeighbors', 10, ...
    'DistanceWeight', 'Equal', ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
[label,score,cost] = predict(classificationKNN,Xnew)
Label_out(i).label=label;
score_out(i).score_1=score;
cost_out(i).cost_1=cost;
end
pred=[Label_out.label];
Target=Bm2(:,end)';
[c,cm,ind,per] = confusion(Target,pred);
ACC=1-c;
trainingData=Bm2;
response=trainingData(:,end);
predictors=trainingData(:,1:end-1);
rng('default')
rng(1);
classificationKNN_ALL = fitcknn(...
    predictors, ...
    response, ...
    'Distance', 'Cosine', ...
    'Exponent', [], ...
    'NumNeighbors', 10, ...
    'DistanceWeight', 'Equal', ...
    'Standardize', true, ...
    'ClassNames', [0; 1]);
 End
```

More than two thousands code lines have been generated to implement the work presented in the current thesis. All the designed files and Matlab functions of this this work have been attached in DVD.