Loukil, Zainab ORCID: 0000-0003-2731-7051, Ali Mirza, Qublai Khan ORCID: 0000-0003-3403-2935, Sayers, William ORCID: 0000-0003-1677-4409 and Awan, Irfan (2023) A Deep Learning Based Scalable and Adaptive Feature Extraction Framework for Medical Images. Information Systems Frontiers. doi:10.1007/s10796-023-10391-9 (In Press)

A Deep Learning based Scalable and Adaptive Feature Extraction Framework for Medical Images

Zainab Loukil[1] · Qublai Khan Ali Mirza[1] · Will Sayers[2] · Irfan Awan[3]

Abstract

Features extraction has a fundamental value in enhancing the scalability and adaptability n of medical image processing framework. The outcome of this stage has a tremendous effect on the reliability of the medical application being developed, particularly disease classification and prediction. The challenging side of features extraction frameworks, in relation to medical images, is influenced by the anatomical and morphological structure of the image which requires a powerful extraction system that highlights high- and low- level features. The complementary of both feature types reinforces the medical image content-based retrieval and allows to access visible structures as well as an in-depth understanding of related deep hidden components. Several existing techniques have been used towards extracting high- and low-level features separately, including Deep Learning based approaches. However, the fusion of these features remains a challenging task. Towards tackling the drawback caused by the lack of features combination and enhancing the reliability of features extraction methods, this paper proposes a new hybrid features extraction framework that focuses on the fusion and optimal selection of high- and low-level features. The scalability and reliability of the proposed method is achieved by the automated adjustment of the final optimal features based on real-time scenarios resulting an accurate and efficient medical images disease classification. The proposed framework has been tested on two different datasets to include BraTS and Retinal sets achieving an accuracy rate of 97% and 98.9%, respectively.

1 Introduction

Medical image processing is a challenging step towards the efficiency enhancement of disease detection and diagnosis.

The analysis of medical images has been considered as challenging and time-consuming task, particularly, for doctors and specialists. Improving the early diagnosis of a medical disease presents a serious challenge for them. To cope with this problem, medical field is being in a massive progress to improve existing physiological analysis methods as well as medical machines for early disease detection and prediction. This topic has gained a great importance in medical innovative research, as a result it becomes an inner area for researchers including different specialities such as doctors, computer and data scientists to use medical images in several applications.

One of the interesting stages in image processing is the medical image content-based retrieval. The complicated composition of medical images makes the information extraction a challenging step. Features extraction represents an important stage towards providing relevant image content-based to result efficient medical application, for example, disease detection, medical analysis, disease prediction, ...etc. Each medical application is reflected by a focus area, namely,

✉ Zainab Loukil
  zloukil2@glos.ac.uk

  Qublai Khan Ali Mirza
  qalimirza@gmail.com

  Will Sayers
  wsayers@glos.ac.uk

  Irfan Awan
  i.u.awan@bradford.ac.uk

1 Cyber and Applied Computing, University of Gloucestershire, The Park, GL50 2RH Cheltenham, Gloucestershire, UK

2 Games Technologies, University of Gloucestershire, The Park, GL50 2RH Cheltenham, Gloucestershire, UK

3 Computer Science, Faculty of Engineering and Informatics, University of Bradford, Richmond Rd, BD7 1DP Bradford, West Yorkshire, UK

Region of Interest (RoI), which contains most of the needed features to accurately accomplishing the target task, e.g., classification.

In recent years, Artificial Intelligence (AI), approaches particularly Deep Learning (DL), have evolved significantly due to the improvement in the processing capacity of computers and the accumulation of big data Arel et al. (2010). DL proved a strong ability to identify meaningful relationships in raw data, which justifies its application to support diagnosing, treating, and predicting outcomes in many medical situations. DL approaches, have already proven their capability that has outperformed human capabilities in medical applications particularly in diagnosing and predicting diseases development. DL proved a strong ability to identify meaningful relationships in raw data, which justifies its application to support diagnosing, treating, and predicting outcomes in many medical situations. DL is transforming the practice of medicine; it is helping doctors diagnose patients more accurately, make predictions about the patient's future health, and recommend better treatments Ravì et al. (2016); Litjens et al. (2017).

DL approaches present key-methods for several medical applications including decision making, disease stage tracking, disease detection, disease diagnosis and analysis, ...etc. DL networks have shown a high sensitivity and accuracy for the detection of several diseases including breast cancer Yala et al. (2019), brain tumour Zhao et al. (2018), Diabetic Macular Oedema (DMO) Tang et al. (2021), ...etc. Its application, particularly for features extraction, has helped into controlling the progress of these diseases by improving the early detection. Figure 1 represents a diagram summarising the impact of DL on medical applications.

Based on the presented diagram, the application of DL approaches has also increased the scalability and reliability of the features' extraction methods. Convolutional Neural Network (CNN) represent one of the most used

architectures in features extraction Razzak et al. (2018), in addition to Multilayer Perceptron (MLP) Lai and Deng (2018). The implication of AI at this processing stage has shown a great improvement in the outcome of the classification and prediction tasks which represents a challenging area in medical imaging.

Despite the integration of DL in medical image processing, traditional features extraction methods have also been applied alongside it. Particularly, salient and semantic features are one of the important extracted features in medical images Gao et al. (2021); Conghua et al. (2006). These features have been used in several applications such as images fusion, and image content-based retrieval. Yoan et al. proposed a medical images fusion method based on salient feature extraction using Particle Swarm Optimisation (PSO) optimized algorithm and the fuzzy logic Gao et al. (2021). The suggested salient features extraction method is based on the non-subsampled shealet transform (NSST), where the latter helps into reducing the computational complexity of the approach. The image fusion process is based mainly on the extraction of low- and -high frequency sub-bands features through the fuzzy logic and uses the PSO algorithm for optimization. The proposed method has been tested on eight pairs of gray scale and five pairs of colour multimodal medical images.

The amount of the testing set is considered very low in order to validate the suggested method. Subsequently, this limits the scalability of its application in real-time scenarios. Semantic features have also been applied by Conghua et al. (2006). Their method is based on the space density function, where they enhanced the original method which used Bayesian Belief network (BNN) Peng and Long (2001); Conghua et al. (2005). The main idea is to transform the medical images from gray-scale to density function space. Their method has been tested on 400 pieces of images covering head, chest, abdomen and limbs of human bodies. The

Fig. 1 Deep Learning Impact On Medical Applications
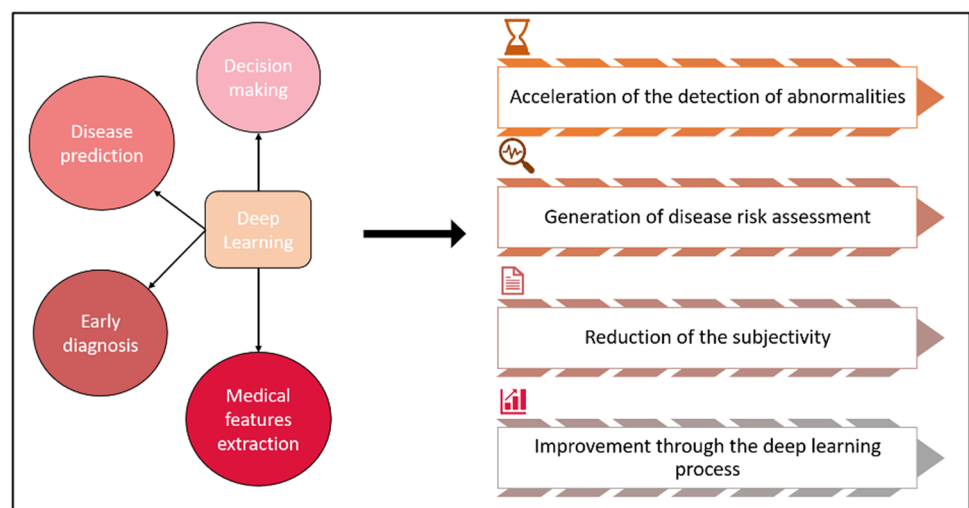
outcome precision of their method reflects a good image retrieve performance achieving 88.8%. One of the drawbacks presented by this method is the non-consideration of coloured dataset and the limited number of validation images. Similarly to Conghua et al. (2006), this leads to the scalability problem. Add to that, based on Gao et al. (2021); Conghua et al. (2006), salient features extraction is mainly dependent on labeled datasets, which decrease its reliability and responsiveness in case of unlabeled input samples. That is, these irregular features are not efficient in such scenarios. Therefore, in this work, the main focus is on the regular features categorized as high- and low-level features. Many classical and recent methods have been proposed to solve the features extraction step. Some of these methods consider single feature usage, such as texture, some others contemplate a combination of different feature levels.

In this paper, a new features extraction framework is proposed. The first contribution of the method is to select the optimal features combination according to the input dataset. The fusion involves two main types of features including high-level and low-level features. The second contribution of the proposed extraction tool is the integration of a new automated hybrid deep network for deep features extraction. The massive enhancement of the resulted classification is dedicated to the resulted optimal features fusion. The structure of the paper is as follows:

- The first section covers the related features extraction works.
- The second section highlights the identified problems.
- The following section details the suggested features extraction methodology where a deep explanation of the different types of features considered in this work, add to the features weighting and fusion stages, and finally the experimentation and evaluation done.
- The paper ends with a conclusion and further future directions.

2 Related Works

Medical image features extraction presents an important step towards resulting highly accurate analysis related, for instance, to disease detection, classification, and prediction. Extracting reflective features reinforce the efficiency rate of these particular applications. Features are categorised as two main types: High level Features (HF) and Low-level Features (LF) Chowdhary and Acharjya (2020). HF features, in particular, include texture, shape, and colour features. These features represent the fundamental features that can be extracted from medical images Mutlag et al. (2020). LF features, also named Deep Hidden Features (DHF), cover the low-level characteristics of a medical image. These include hidden information reflecting important analysis and leading to enhancing the diagnosis reliability Jeyakumar and Kanagaraj (2019).

In this context, several proposed features extraction frameworks have been applied to address both HF and LF extraction issues. However, these related works still outline some drawbacks in terms of the deployment of these features Huerga et al. (2021); Hazarika et al. (2021); Tsai et al. (2017); Rundo et al. (2019, 2021); Kavya and Padmaja (2017); Xiao et al. (2013); Zewail and Hag-ElSafi (2017); Liu and Shi (2011); Mingqiang et al. (2008).

HF and LF features, in particular, are considered the main point of interest in several features extraction methods. Multiple challenges have been highlighted in the literature to include: (1) testing models using different dataset sizes and complexities, (2) using different types of datatset to convey coloured and gray-scale based images, (3) potential of validating models using real-case scenarios, (4) testing features extraction systems responsiveness to multiple cases, and (5) the lack of sufficient extracted features in some particular scenarios.

The size of datasets applied for features extraction experimentation has an impact on the complexity of the framework outcomes. Hence, considering different dataset sizes is of great importance in experiments validation. In fact, Tahira et al. evaluated their DL-based method over challenging datasets, namely, APTOS-2019 and IDRiD Nazir et al. (2021). Both datasets have different sizes and complexities, hence the difference in the validation performances of the same model. Similar impact has been highlighted in the content-based image retrieval system proposed by Lin et al. (2009). The use of different sets of data, covering multiple aspects, proved the importance of such consideration in features extraction based models to achieve 99.2% accuracy (Acc), 72.7% average precision, and 50% average recall. That said, this factor has not been considered in several proposed features extraction frameworks. Despite the use of complex datasets, these methods lack the dataset experiments validation which, as a result, impact the reliability of such methods. For instance, a supervised Support Vector Machine (SVM) based features extraction model has been suggested by Xiao et al. (2017), providing a good model performance validation. Similar approach has also been proposed by Janakasudha and Jayashree (2020), however, different dataset has been used. Considering the same extraction method, both works proved different validation performances which, as a result, stresses the importance of considering size and complexity of datasets when it comes to building a reliable model Janakasudha and Jayashree (2020); Xiao et al. (2017). Moreover, considering the aforementioned factor will potentially add a scalablility factor to the resulted model.

Medical images can be presented in different morphological manners. These multiple representations could also impact the final outcome of the features extraction framework. Considering the latter, medical images types to convey coloured and gray-scale based images, have tremendous effect on the processing stage. Features, to include HF and LF, vary from one medical image type to another. In fact, coloured based images are a source of colour features which gray-scale based images lack of. Texture and shape features, on the other hand, can be extracted from both image types, however, some of these researches ignore the importance of coloured based datasets, instead focusing mainly on gray-scale medical images Huerga et al. (2021); Tsai et al. (2017); Rundo et al. (2019, 2021); Xiao et al. (2013); Zewail and Hag-ElSafi (2017); Janakasudha and Jayashree (2020); Xiao et al. (2017); Altaf et al. (2017); Howarth and Rüger (2004); Dara et al. (2018); Madusanka et al. (2019). This can be justified by the cost of considering coloured medical images, however, this has a drop impact on the reliability and scalability of these methods. In fact, as proposed by Liu et al. (2020). the consideration of two types of datasets gives the model a free-spcae to interpret HF and LF features; thus, removing the interpretability as a major challenge. This, however, was not the case for exiting features extraction models that focused mainly on texture and shape features extraction Madusanka et al. (2019); Janakasudha and Jayashree (2020); Xiao et al. (2017). Despite achieving interesting results in terms of accuracy, sensitivity, and specificity, these methods lack of the consideration of coloured based medical images dataset, hence, the non-scalability of their proposed model. DHF based features extraction frameworks also have been, in multiple instances, part of the above challenges particularly when it comes to processing medical images such as Computerised Tomography (CT) and Magnetic Resonance Imaging (MRI) scans proposed respectively by Dara et al. (2018), Liu et al. (2020), and Janakasudha and Jayashree (2020). However, despite the consideration of coloured based medical images dataset, DHF extraction can also lack the importance of features that can be extracted through gray-scale based datasets Nazir et al. (2021). Hence, its lack of reliability and scalability as per the above.

The potential use of features extraction frameworks on real-case scenarios also make several proposed methods under the question of their responsiveness, reliability and scalability towards particular testing experiments Altaf et al. (2017); Howarth and Rüger (2004); Liu et al. (2020). The consideration of multiple inquires helps in evaluating the consistency of the proposed model. Altaf et al. (2017), for instance, considered multiple techniques combinations. However, no experimental setup has been in place to cover multiple scenarios, hence, the lack of sufficient features extraction. Similar experimentation approach has been considered by Howarth and Rüger (2004), which limited their

evaluation mechanism leading to a drwaback in considering multiple features combinations, to include HF and LF, hence the importance of features fusion.

Texture features have been the interest point of several features extraction frameworks proposed in the literature. In particular, several methods have been applied for texture features extraction. Gray-Level Co-occurrence Matrix (GLCM), has been widely used for texture features extraction Hazarika et al. (2021). GLCM has demonstrated high efficiency in extracting discriminative features. Tsai et al. (2017) proposed a Graphical Processing Unit (GPU) based features extraction from MRI images (gray-scale). Towards accelerating processing time metric and reducing processing complexity. Based-on Region of Interests (RoIs) localised in the medical image, a set of Haralick features are derived from GLCM including: auto-correlation, dissimilarities, variance, entropy, ...etc. Despite the high-level of efficiency obtained by the suggested method, the work lacks in identifying the complexity of the used dataset which limits the potential of benchmarking their proposed method.

In the same context, Leonardo et al. proposed a new GPU-powered texture features extraction method based on the full dynamics of gray-scale levels Rundo et al. (2019). The tested dataset is composed of MRI and CT scans with no specification of related dataset size. This leads to questioning the scalability of these methods and their reliability in real-case scenarios. Recently, a CUDA-powered method for texture features extraction method has been proposed to cover unsupervised analysis of medical images, particularly CT scans Rundo et al. (2021). The suggested method relies on the mixture between GLCM and Self-organizing Map (SOM), namely CHASM. The proposed method showed high performances in terms of responsiveness over-passing the pre-suggested methods Tsai et al. (2017); Rundo et al. (2019). In addition, the proposed approach is based mainly on unsupervised extraction which covers the case of unlabeled dataset. A drawback presented by CHASM of the non-consideration of coloured dataset which can be a challenging problem when it comes to its possible application as a first or second clinical line tool. Texture features have also been extracted for medical disease detection, for instance, Glaucoma. Kavya et al. proposed a new framework for Glaucoma detection using texture features extraction Kavya and Padmaja (2017) in addition to GLCM, Gaussian Markov Random Field (GMRF) has been applied for texture extraction. The combination GLCM-GMRF reinforced the output result of the final classification task to reach 86% Acc. Despite the high accuracy of the proposed model, it did not cover the colour features of the used OCT dataset, which in turn makes the method less generalisable to cover real-case scenario. In addition, the limitation of the number of images used (50 images) is considered as not sufficient to validate the proposed method.

Shape features are also one of the most useful HF features to extract relevant information from medical images, for instance, tumour shape in case of MRI images, Optic Nerve Head (ONH) in case of OCT images,...etc Kavya and Padmaja (2017); Xiao et al. (2013); Zewail and Hag-ElSafi (2017); Liu and Shi (2011); Mingqiang et al. (2008). Kai et al. considered deformation-based features to construct a more accurate anatomical meaning from the images to represent the brain tumor Xiao et al. (2013). Their work is based on the use of MRI image, particularly the lateral ventricular part of the brain towards extracting the deformation of the shape features. The method consists of retrieving the lateral ventricular shape, then the estimation of its deformation and finally transforming it into an actual representative feature. One of the advantages of this method is the use of the supervised and unsupervised methods including K-nearest neighbors (K-nn) and conventional FCM, respectively. Their classification results shown a high sensitivity of 95.3 % in case of supervised method (K-nn), and 81.9 % in case of unsupervised method (FCM). However, the drawback of their method is that: (1) the non consideration of other features (e.g., texture) in order to improve the classification outcome, (2) the lack of covering coloured dataset, and (3) the lack of scalability due to a limitation in validating the suggested method with very few cases (i.e., 15 cases). Shape features have been also considered as key-feature in Rami et al. proposed method Zewail and Hag-ElSafi (2017). This sparse contourlet-based extraction approach is composed of mainly two methods including Second Moment Matrix (SMM), and non-subsampled Contourlet representation (NSCT). The combination NSCT-SMM is based mainly on non-maximum suppression and thresholding after the generation of shape features strength. The outcome of the proposed method showed a high Acc of 78.91% and a low Mean Average Error (MAE) to achieve 21.43%.

Combining HF features presents a supporting factor for medical applications by providing extra relevant information about the target RoIs Mutlag et al. (2020); Hazarika et al. (2021). Texture and shape features fusion has been considered in several works, particularly in Nazir et al. (2021). Bhaveneet et al. proposed a new implementation of features extraction from medical images consisting of the extraction of three features levels: (1) key-points, (2) contours, and (3) textures, storing them into a feature vector and highlighting them on the original medical image. The suggested implementation has been tested on three different datasets including MRI, Iris and Bones achieving as a classification Acc of 90% which is higher than the previous mentioned proposed approaches. That is, features fusion represents a key-stage towards enhancing images content-based retrieval. One of the disadvantages of the method is, again, the scalability and reliability by the non consideration of different datasets and lack of validation experiments, as well as the lack of extraction of colour features. The latter represents a key-element in several medical applications that are colour-based. Colour feature-based retrieve is of interest to many recent researches following the advanced engineering technologies in enhancing the medical image scans quality as well as integrating coloured options in image analysis and diagnosis, particularly in case of Optical Coherence Tomography (OCT) images Lin et al. (2009).

Traditional features extraction methods, including those to extract HF features, are still facing challenges extracting deep hidden features due to their classical composition. AI techniques, particularly DL, have shown an interesting enhancement of classification and prediction applications. Dara et al. (2018), proposed a DL-based deep features extraction method using CNN. The latter has been tested along with Deep Belief Network (DBN) and Multiple Layer Perceptron (MLP) on an MRI dataset composed of 69 subjects. CNN presented the most accurate network with 99% Acc. Despite considering the high accuracy, the suggested framework lacks of the reliability factor because of the a small number of input sample which might risk causing a thrashing problem, and suffers from achieving scalable factor due to the non-consideration of unlabeled data. Similar approach has been adopted by Nazir et al. (2021). The proposed method is based on CNN architecture, particularly Densely Connected Network-100 (DenseNet-100). The model accurately extracts hidden features and results an outstanding classification performance applied on OCT dataset for DMO detection. Despite the existence of colour features, the method eliminated the latter and focused mainly on DHF features. Subsequently, integrating colour features might of increase the final classification outcome.

The complementarity of DHF and HF features represents the main contribution of this work. In the following section, a highlight of the identified problems is presented.

3 Problems Identified

The main drawbacks identified in existing features extraction methods include responsiveness, scalability, and reliability. Several approaches achieved high accuracy as an initial evaluation of the proposed framework, however many requirements have not been met so far. An efficient features extraction method consists of the consideration of every aspect that can be retrieved from the medical image in order to reflect a certain RoI. This includes texture, shape, and colour (in case of coloured dataset). In addition, hidden features that can be extracted through DL approaches represent additional important features towards having a complimentary feature set. The elimination of one of these features is affecting the final efficiency of the medical application. The use of small set of medical images represents an inefficient

way to validate a proposed framework. In fact, features provided by a small set of images limits the evaluated method to be generic. Add to that, several problems can occur including the over-fitting where the DL model cannot successfully classify the data when it becomes higher than what it has been trained on (small dataset), and under-fitting where the DL network cannot find the accurate relationship between the dataset used and the input samples hence the non-scalability of the designed feature extraction model. In turn, this effects its reliability and initial efficient functionality.

In this paper, the addressed problems are as follows:

- Unautomated methods for medical images based features extraction
- Non-scalability of existing approaches
- The lack of the use of HF and LF features in a unique framework

4 Methodology of the Proposed Features Extraction Model

Feature extraction from image data is a crucial step for content-based retrieval. Particularly, its significant application in case of medical images is considered challenging. The variety and deepness of extracted features represent the key-points in achieving high classification and prediction performances.

The methodology presented in this research focuses mainly on the extraction and fusion of HF and LF features. HF features extraction is based on segmented images whereas LF features are derived from associated parameters provided along with the input dataset. The proposed features extraction framework is summarized in Fig. 2.

The following subsections outline the features extraction and fusion model. The first stage is to pre-process the data, second is to define the images segmentation method and associated parameters, third is to design the HF features extraction block, then is to introduce the LF features extraction stage, afterwards the features weighting method and optimal features fusion strategy.

4.1 Image Pre-Processing

Image pre-processing represents a major and essential step towards improving features extraction by eliminating unwanted noise and irrelevant regions located in the medical image. The proposed image pre-processing model consists of the following principles:

- Ground truth extraction for data training and testing stage
- Images denoising using block matching and three dimensions filtering (BM3D) method.
- Bias field correction using N4 bias field correction method.

These steps are pivotal in order to enhance the quality of the image considered, from algorithmic perspective, as a matrix of pixels/intensities. It leads into the elimination of non-essential areas that contain unwanted signals which results image quality degradation. In fact, the medical image ground truth is important for the validation of the RoIs segmentation. Denoising of medical scanned images such as OCT, MRI, CT, ...etc is also an important stage towards enhancing the outcome of medical applications including, detection, analysis, and prediction. Subsequently, denoising stage generates clean images with high "signal-to-noise ratio" as well as high spatial resolution. In this denoising model, block-matiching and BM3D method are used to denoise the input samples Zhao et al. (2019). Main steps used in BM3D are grouping, 3-dimensional discrete wavelet transformation and wavelet shrinkage. BM3D can remove the noise easily by eliminating it from the group of similar patches. The principle of denoising is to remove the additive
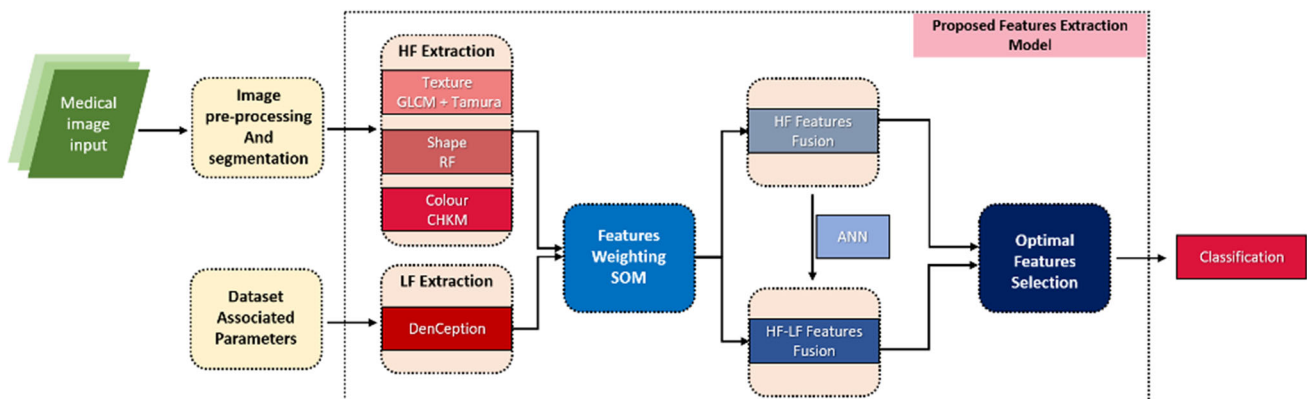

**Fig. 2** Proposed Features Extraction Methodology Framework

noise and invert the blurring at the same time Kaur et al. (2018). This method is called Wiener filter. The latter determines the optimal tradeoff between the inverse filtering and the noise smoothing.

The N4 bias field correction algorithm is a popular method for correcting low frequency intensity and the non-uniformity present in the medical image data, known as a bias or gain field. The main purpose of this stage is to ensure that the mask image and the main input image occupy the same physical space to ensure pixel to pixel correspondence. All these steps are complementary towards producing a high-quality input sample that can be effectively processed and analysed.

Figure 3 presents an example of image pre-processing applied on MRI image provided by RSNA-ASNR-MICCAI Brain Tumor Segmentation (BraTS) dataset contacting subjects with O6-methylguanine-DNA methyltransferase (MGMT) Altaf et al. (2017); Kavya and Padmaja (2017); Liu et al. (2020). The figure illustrates the results of a Flair modality for two different subjects. The top image is related to a subject with negative MGMT and the below image belongs to a subject with positive MGMT.

Fig. 3 Pre-processing Steps Applied on FLAIR Modality: (TOP) Subject with Negative MGMT, (BOTTOM) Subject with Positive MGMT
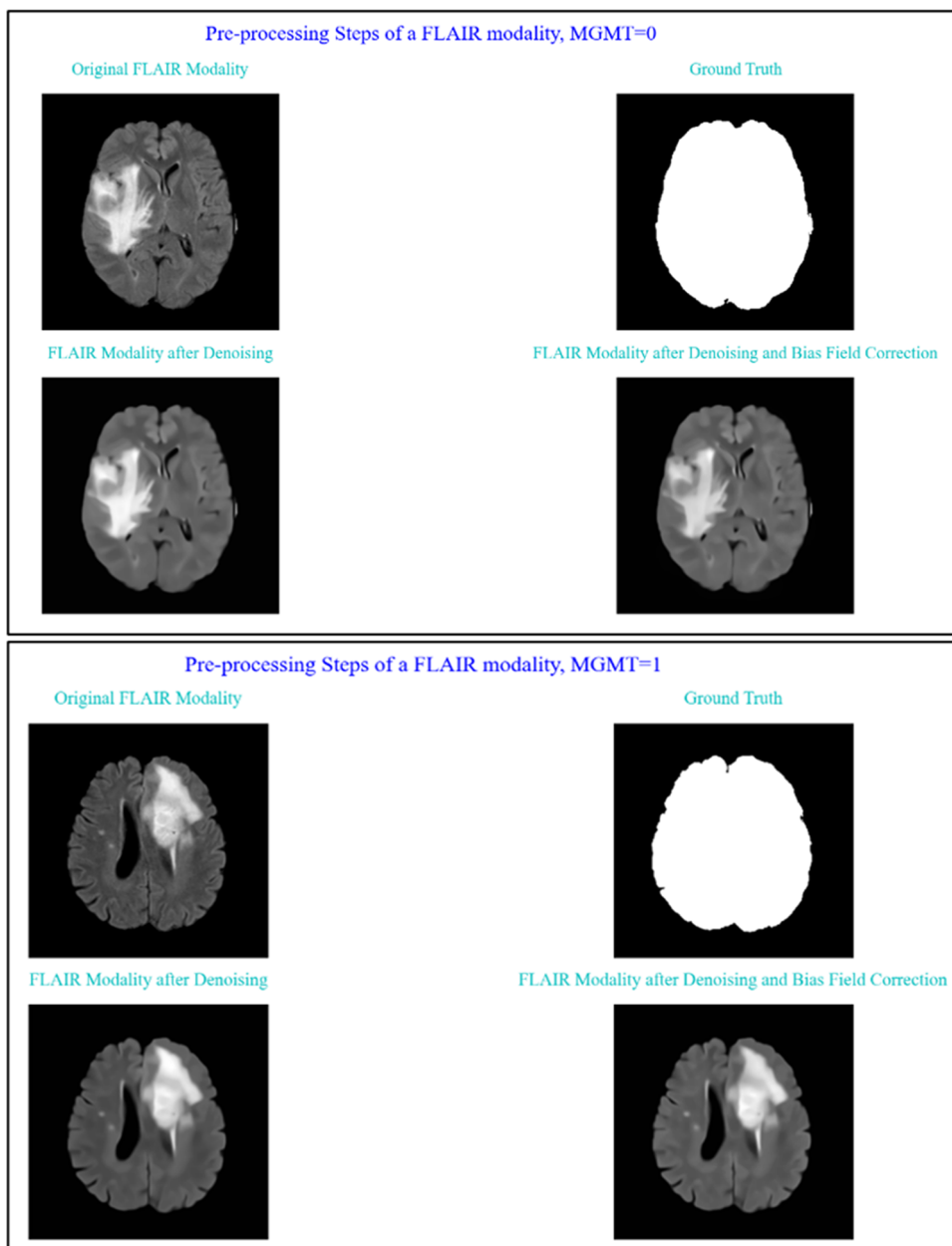
4.2 Image Segmentation and Associated Parameters

Image segmentation characteristics represent the lower level of image characteristics including pixel intensities, RoI, bounding, edges, ...etc. It is defined by semantic image segmentation through extraction of RoIs of the input samples. Medical image segmentation represents a challenging step due to its deformable characteristics. The aim of semantic segmentation is to partition the image into multiple segments in order to simplify the representation of which makes it more significant and easier to process.

The focus of this work is mainly on unsupervised segmentation algorithms. For instance, Markov Random Field (MRF) presents one of the well-used unsupervised segmentation algorithms, in addition to Expectation-Maximization (EPM) algorithm. The combination MRF-EPM iterates the posteriori probabilities and distributions of labeling in case there are no possibilities of the construction of an estimate segmentation model, i.e., no predefined classes. The segmentation process starts with randomly estimating the model parameters, then computing the conditional probabilities of a label given a random image region using naïve Bayes technique. The conditional probabilities are defined as follows (Eq. 1):

$$P\left(\frac{\lambda}{r_i}\right) = \frac{P(\frac{r_i}{\lambda})P(\lambda)}{\sum\limits_{\lambda \in L} P(\frac{r_i}{\lambda})P(\lambda)} \tag{1}$$

where L represents the set of possible labels, λ is the given label, and r_i is the region of features. Finally, MRF-EPM iterative algorithm uses the output of proceeding step in order to calculate the priori estimate of a given label, $\lambda \in L$. The computation involves a hidden estimate of the number of labels (β), knowing that the actual number of total labels is unknown. The priori estimate is defined as the following (Eq. 2):

$$P(\lambda) = \frac{\sum\limits_{\lambda \in L} P(\frac{\lambda}{r_i})}{\beta} \tag{2}$$

Figure 4 shows the result of MRF-EPM algorithm applied on the MRI image presented in Fig. 2.

The method successfully segments the RoIs in the original image. Resulted segmented images will be used as the input data for the feature extraction model. In the following, an overview and banchmarking of HF features extraction stage of the proposed framework is presented.
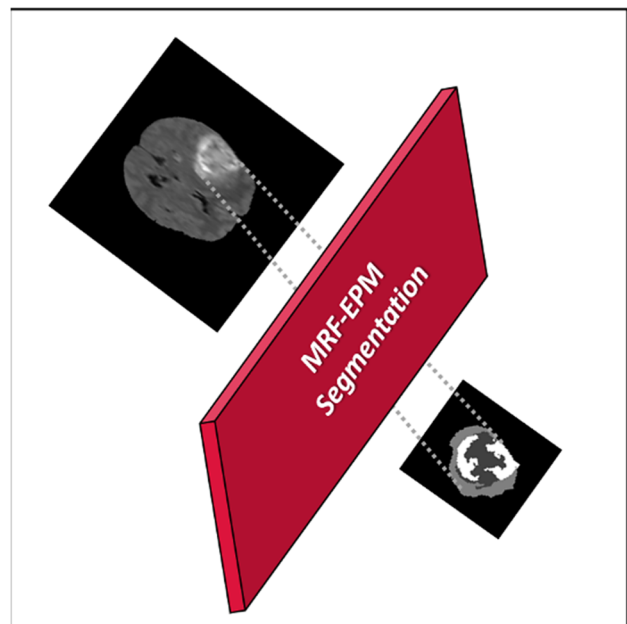
Fig. 4 MRF-EPM Segmentation: Test Done on a Sample Scan from the BraTS Dataset

4.3 High-Level Features Extraction

HF features are mainly defined by the features that can be interpreted by human brain. This is presented in the form of spectral features including texture, shape, and colour.

4.3.1 Texture Features Extraction

Texture features are based on the collection of image regions. It generally refers to a specific region within the image. Referred RoIs provide other important features such as shape and colour which will be discussed in the subsequent sections. Texture features extraction methods are generally divided into two main categories based on: (1) spatial relationship between regions, and (2) primitive attributes. The former includes (i) primitive region types presented as numbers and (ii) spatial organisation covering functional, structural, and statistical features. Primitive attributes texture features category focuses mainly on (i) gray-level and (ii) geometrical attributes. The latter covers the shape, area, ...etc, whereas, grAy-level attributes enclose average and extremum. Texture features generally highlight discriminative features that represent key-features in disease detection and prediction applications. The focus of the proposed texture features extraction method is based on statistical features as the following:

- First- and second- order features including: contrast, entropy, angular second moment, and homogeneity.

- Additional features to include Coarseness and directionality.

Texture features reflect changes that might happen in the medical image due to disease detection and progression which in turn affects the pixels intensities. Several methods can be applied to extract texture features, for instance, GMRF Kavya and Padmaja (2017), SOM Rundo et al. (2021), GLCM Rundo et al. (2019); Kavya and Padmaja (2017); Altaf et al. (2017), and tamura Mutlag et al. (2020); Umamaheswari et al. (2018) approaches. As per Table 1, multiple performance parameters of the aforementioned methods have been reported in the literature to include classification accuracy, processing speed-up and other parameters.

Despite of demonstrating a valuable speed-up performance of 0.3 times, SOM method requires additional parallel computing platform that allows it to use certain types of GPUs which represents a limitation in case of the non-existence of such resource. GLCM, at the other hand, showed an independent processing speed-up to reach 19.5 times due to its processing optimisation and image handling. It also overpassed the classification accuracy of GMRF to acheive 86%. By considering the optimal pixel direction and orientation, tamura's features application showed quite an interesting classification accuracy to reach 96% and 3.43% of the mean average precision retrieval. As per the above, in this study a combination of GLCM and tamura is proposed.

First- and Second- Order Texture Features: GLCM Technique GLCM technique is based mainly on pixels intensity and related changes. The major advantage of GLCM is that the co-occurring groups of pixels are spatially linked in multiple directions by referencing to two different factors including distance and angular second moment relationships. GLCM also highlights the busy texture regions defined by a very rapid changes of one-pixel intensity compared to its neighbours. Thus, it results a high intensity alteration of the related special frequencies. The GLCM algorithm first quantises the segmented input by specifying the value of each pixel intensity. The quantisation is specified based on a range of gray-levels included in the range of [2:256]. Second, it creates the co-occurrence matrix sized (n*n), where n presents the number of levels used in the quantisation step. The creation of co-occurrence matrix ($GLCM_f$) is based on the calculation of the number of occurrences of a pixel (p), located at (i,j) coordinates, in a pre-defined iterative window that covers the surrounding pixels. The steps are detailed as follows:

- Set p the sample considered for calculation.
- Set S the group of neighbour pixels surrounding p. The selected group is done under a centered window having as length, and height values in [3:999] interval.
- Each element (i,j) of GLCM matrix, based on S, is defined as (Eq. 3):

$$GLCM(i, j) = \sum_k occ(i, j) \qquad (3)$$

where (i,j) are the ith and j^{th} pixels intensities $\in$ [0:n-1], and occ() is the function representing the time of occurrence of (i,j) in S based on multiple direction and distance relationships, and k is the total occurrence of (i,j) in the centered window.

- Construct the symmetrical matrix of GLCM and add it to the co-occurrence matrix itself (Eq. 4):

$$GLCM_f = GLCM + GLCM_s$$
$$= \sum_k occ(i, j) + occ(j, i) = 2 \sum_k occ(i, j) \qquad (4)$$

where $GLCM_s$ is the symmetric matrix and $GLCM_f$ is the final co-occurrence matrix.

Table 1 Existing Texture Features Extraction Methods Benchmarking

Methods	Performance Parameters Classification Accuracy	Processing Speed-up	Mean Average Precision Retrieval
SOM Rundo et al. (2021)	-	10.03 times	-
GLCM Rundo et al. (2019)	-	19.5 times	-
GLCM Kavya and Padmaja (2017)	86%	-	-
GLCM Altaf et al. (2017)	79.8%	-	-
MRF Kavya and Padmaja (2017)	84%	-	-
Tamura Mutlag et al. (2020)	96%	-	-
Tamura Zhao et al. (2019)	-	-	3.43%

- Normalisation of $GLCM_f$ (Eq. 5):

$$GLCM_f(i, j) = \frac{\sum\limits_{k} occ(i, j)}{M} \qquad (5)$$

where M is the number of total elements, $M > 0$.
- Calculate first- and second-order texture features as the following:

  – Angular Second Moment (ASM): ASM is known also as Energy feature, denoted f_{ASM}, it is defined as the squared elements of $GLCM_f$, as follows (Eq. 6):

$$f_{ASM} = \sum_{j \in n} \sum_{j \in n} GLCM_f(i, j)^2 \qquad (6)$$

  – Entropy (E): It is defined as the measurement of haphazardly to be utilized in order to differentiate the texture of the segmented input sample, as follows (Eq. 7):

$$f_E = -\sum_{i \in n} \sum_{j \in n} GLCM_f(i, j) * log(GLCM_f(i, j)) \qquad (7)$$

  – Contrast (C):

Contrast is defined as the value of density contrast reference pixels and surrounding pixels, as follows (Eq. 8):

$$f_C = \sum_{j \in n} \sum_{j \in n} (i, j)^2 GLCM_f(i, j) \qquad (8)$$

where $GLCM_f$ (i,j) equals to pixel at the (i,j) location.
- Homogeneity (H): H is defined by approximately measure the $GLCM_f$ elements distribution to $GLCM_f$ diagonal (Eq. 9):

$$f_H = \sum_{j \in n} \sum_{j \in n} \frac{GLCM_f(i, j)}{1 + |j - i|} \qquad (9)$$

Additional Texture Features Tamura is also one of the well-used quantitative texture features extraction methods. It is based mainly on human visual perception and it represents an immense potential in image representation. Tamura provides a set of texture features including: roundness, directionality, line-likeness, regularity, coarseness, as well as contrast texture features. Ideally, tamura's texture features present complementary features to those extracted through GLCM approach. The main drawback let to combining GLCM and tamura are as follows:

- GLCM is a sparse matrix, containing many zero elements, which causes an increase in computational time and resource Kaur et al. (2018); Baid et al. (2021).
- Tamura performs inefficiently in case of generic (non-homogeneous) images.

The proposed feature extraction approach includes coarseness (Coa) and directionality (Dir) as additional features. Tamura's discriminative features are defined in the following.

- Coarseness: Coa is defined by iteratively find the largest size in which the tissue is present through different patterns at multiple scales. The granularity measurement is done by calculating, for each pixel (i,j), six averages for a window of size $2^Z * 2^Z$, where $Z \in [0:5]$, surrounding the pixel defined as follows (Eq. 10):

$$Coa_z = \sum_{k=i-2^{Z-1}-1}^{i+2^{Z-1}} \sum_{t=j-2^{Z-1}-1}^{j+2^{Z-1}} \frac{pix(k, t)}{2^{2Z}} \qquad (10)$$

where pix(k,t) is the pixel intensity at location (k,t). Iteratively, at each pixel, calculation of non-overlapping neighbours defined by the absolute difference $A_Z(i, j)$ in both relationships: Vertically (V) and Horizontally (H) as follows (Eqs. 11a and 11b):

$$A_{Z,V}(i, j) = |Coa(Z, V)(i, j+2^{Z-1}) - Coa(Z, V)(i, j-2^{Z-1})| \qquad (11a)$$

$$A_{H,V}(i, j) = |Coa_{H,V}(i+2^{Z-1}, j) - Coa_{Z,H}(i-2^{Z-1}, j)| \qquad (11b)$$

Finally, considering either direction (V or H), calculation of the value of Z is processed in order to maximise $A_{Z,V}(i, j)$ or $A_{Z,H}(i, j)$, respectively. The function is defined as follows (Eq. 12):

$$S_{Z,BEST}(i, j) = 2^Z \qquad (12)$$

resulting the final coarseness feature equation (Eq. 13):

$$f_{Coa_Z} = \frac{Coa(i, j)}{S_{Z,BEST}(i, j)} \qquad (13)$$

- Directionality: Dir is defined by devolving the existence of any directional pattern in an image by measuring the

overall degree of directivity (vertically, horizontally, or diagonally). This feature reflects the consistency of the region being processed. Dir consists in calculating the edge histogram (H_{Dir}). Dir texture feature is defined as follows (Eq 14):

$$f_{Dir} = 1 - N * N * m * \sum_{k=1}^{m} \sum_{\theta \in \psi_k} (\theta - \theta_k)^2 * H_{Dir}(\theta) \quad (14)$$

where:

- N: normalisation factor
- θ: quantisation angular position constructed by counting the edges of pixels with associated angles directions.
- m: number of peaks
- ψ_k: angles window associated to the kth peak.

The remaining tamura texture features are of importance but not considered in this method. The texture features extraction is implemented in Algorithm 1 and comprise two main steps:

- **Step 1**: Calculation of $GLCM_f$ matrix based on the occurrences (occ()) of pixels at a location (i,j) in the surrounding window S.
- **Step 2**: Calculation of Tamura texture features based on the best pixel direction and orientation at location (i,j) in the surrounding window S.

The hybrid composition of the proposed texture features extraction involves multiple interpretation levels of the input image which gives the system a better understanding of the image composition at different RoIs. GLCM-Tamura combination is considered as a booster to the whole feature extraction framework by: (1) speeding-up the processing time, (2) optimising the use of resources, and (3) increasing the efficiency of the final classification, as per the performance benchmarking shown in Table 1.

4.3.2 Shape Features Extraction

Based on the aforementioned related works, shape features extraction is mainly based-on geometry features including: area, slope, perimeter, centroid, irregularity index, equivalent diameter, convex area, and solidity, ...etc. Table 2 summarises the benchmarking of existing methods.

Towards efficiently using shape features, selected approaches need to meet essential key-points including: (1) identifiability, (2) transition, rotation, and scale invariance, (3) affine invariance, (4) noise resistance, (5) occultation invariance, (6) statistically independent, and (7) importantly

---

Algorithm 1 Texture Features Extraction

Data: RoIs (Region of Interests)
Result: f_{ASM}, f_E, f_C, f_H, $f_{Coa_{Z,BEST}}$, f_{Dir}
Step 1: Calculation of GLCM matrix
$n \Leftarrow$ number of levels
$M \Leftarrow n^2 > 0$
for $i \in \{0, \dots, n\}$ **do**
  for $j \in \{0, \dots, n\}$ **do**
    for $k \in \{0, \dots, M\}$ **do**
      $GLCM_f(i, j) \Leftarrow \frac{2}{M} \sum occ(i, j)$;
      $f_{ASM} \Leftarrow$
      $\sum \sum GLCM_f(i, j)^2 * log(GLCM_f(i, j))$
      $f_E \Leftarrow$
      $-\sum \sum GLCM_f(i, j) * log(GLCM_f(i, j))$
      $f_C \Leftarrow \sum \sum (i, j)^2 * GLCM_f(i, j)$
      $f_H \Leftarrow \sum \sum \frac{GLCM_f(i,j)}{1+|j-i|}$
    end for
  end for
end for
Step 2: Calculation of Tamura features
$pix(i, j) \Leftarrow$ intensity value of the pixel at location (i, j)
$S_Z \Leftarrow 2^{2Z}$ where $Z \in [0 : 5]$
$N \Leftarrow$ normalisation factor
$\theta \Leftarrow$ quantisation angular position
$m \Leftarrow$ number of peaks
$\psi_k \Leftarrow$ angles window associated to the kth peak.
$H_{Dir} \Leftarrow$ edge histogram
$M_w \Leftarrow$ measurement window
for $i, j \in \{0, \dots, n\}$ **do**
  for $k = i - 2^{Z-1} - 1$ to $i + 2^{Z-1}$ **do**
    for $k = j - 2^{Z-1} - 1$ to $j + 2^{Z-1}$ **do**
      $Coa_Z \Leftarrow \sum \frac{pix(i,j)}{M_w}$
    end for
  end for
end for
$A_{Z,V}(i, j) \Leftarrow | Coa_{Z,V}(i, j + 2^{Z-1}) - Coa_{Z,V}(i, j - 2^{Z-1}) |$
$A_{H,V}(i, j) \Leftarrow | Coa_{H,V}(i + 2^{Z-1}, j) - Coa_{Z,H}(i - 2^{Z-1}, j) |$
if $A_{Z,V}(i, j) > A_{H,V}(i, j)$ **then**
  $S_{Z,BEST}(i, j) \Leftarrow S_{Z,V}$
  $f_{Coa_{Z,V}} \Leftarrow \frac{Coa_{Z,V}}{S_{Z,BEST}}$
else
  $S_{Z,BEST}(i, j) \Leftarrow S_{Z,H}$
  $f_{Coa_{Z,H}} \Leftarrow \frac{Coa_{Z,H}}{S_{Z,BEST}}$
end if
end if
for $k = 1$ to m **do**
  for $each$ $angle$ $\theta \in \psi_k$ **do**
    $\rho \Leftarrow \sum (\theta - \theta^2) * H_{Dir}(\theta)$
  end for
end for
$f_{Dir} \Leftarrow 1 - N * m * \rho$

---

reliable. Shape features extraction approaches can be categorised as the following:

- Counter-based methods
- Region-based methods
- Space and transform domain-based methods

Table 2 Existing Shape Features Extraction Methods Benchmarking

Methods	Performance Parameters			
	Accuracy	Sen	Spec	MAE
Fourrier	80%	-	-	-
Descriptor Liu and Shi (2011)				
Region Focus (RF) Xiao et al. (2013)	95.3%	88.9%	99.2%	-
Convolutional FCM Xiao et al. (2013)	81.9%	38.9%	99.7%	-
SMM-NSCT Zewail and Hag-ElSafi (2017)	78.9%	-	-	21.43%
VBM Xiao et al. (2017)	91.18%	99%	83.33%	-
VBM Madusanka et al. (2019)	76.31%	75%	77.78%	-
VBM Janakasudha and Jayashree (2020)	93.8%	-	-	-

- Information preserving and non-information preserving-based methods

Relevant shape features are characterised by their uniqueness, abstraction, integrity, and agility. Table 3 outlines the existing valuable shape features extraction methods from mainly counter and region-based methods.

As per the results presented in Liu and Shi (2011), Fourier descriptor overpassed statistical descriptors by achieving over 80% accuracy. In fact, Fourier descriptors are highly insensitive to translation, rotation, scale changes as well as the starting processing point. It has shown high performances in case of identified objects (human face, vehicles ...etc), however, in case of medical imaging the shape of different RoIs in the input sample changes through time progression, age and gender factors as well. Therefore, this study considers the region focus (RF) as the main shape feature extraction approach by calculating the coordinates of all points belonging to a particular RoIs. It is defined as the following (Eq. 17, 18.a and 18.b):

$$f_{RF} = (\bar{x}, \bar{y}) \tag{15}$$

where:

$$\bar{x} = \frac{1}{A} \sum_{(x,y) \in RoI} x \tag{16a}$$

$$\bar{y} = \frac{1}{A} \sum_{(x,y) \in RoI} y \tag{16b}$$

where A is the region area: $A = \sum_{(x,y) \in RoI} 1$

4.3.3 Colour Features Extraction

This feature is based on coloured medical images. Several features extraction methods have been used in the literature. Table 4 summarises colour feature extraction approaches based on two different categories. This includes global descriptors defined when the whole image is considered, and local descriptors when separated portions of the image are considered.

Colour histogram of K-mean (CHKM), and Zernike chromaticity derived from chromaticity approach are considered highly robust colour features extraction methods. However, it is not the case of colour histogram method. Four main specifications are essential to consider a method as efficient and accurate:

- Storage space
- Scalability
- Rotation invariance
- Computational time required

Table 3 Shape Features Extraction Methods

Shape Features			
	Boundary-Based Features		Region-Based Features
Simple	Statistical	Fourier	Simple
Descriptor	Descriptor	Descriptor	Descriptor
Perimeter of the boundary	Mean moment	Translation	Regional area
Diameter boundary	Variance moment	Rotation	Roundness (Compactness)
Eccentricity Curvature	Higher-order moment	Scale	Regional focus

Table 4 Existing Colour Features Approaches: Global and Local

Colour Features	
Global Features	Local Features
Global Colour Histogram	Local Colour Histogram
Histogram Intersection	Colour Correlogram
Image Bitmap	Colour Difference Histogram

Based on the aforementioned criteria, CHKM has been applied for colour features extraction. CHKM considers 2^{24} different colours possibilities. The main process conveyed by CHKM is to select a colour, denoted c_p, from 2^{24} possibilities that reassemble the best to a particular pixel colour and update the latter with c_p. This step is applied on each pixel towards classifying all the pixels of an image into k clusters. The outcome of it is defined by the mean of all pixels in each cluster. The final output of CHKM feature is (Eq 17):

$$f_{CHKM} = \frac{N_K}{N} \tag{17}$$

Where N is the total number of pixels localized in the image, and N_K is the total number of pixels belonging to cluster K. This method efficiently shortens image retrieval time and improves its performance. Moreover, CHKM demonstrates a less computational time factor, a high robustness to noise and displacement invariance. Algorithm 2 demonstrates the proposed shape and colour features extraction model.

Algorithm 2 Shape and Colour Features Extraction

Data: RoIs (Region of Interests)
Result: f_{RF}, f_{CHKM}
Step1: Calculation of Region focus shape features based on the region are A
$(x, y) \Leftarrow$ coordinates of the pixel $\in$ A
for $(x, y) \in RoIs$ **do**
$\quad | \quad A \Leftarrow \sum 1$
end for
for $(x, y) \in RoIs$ **do**
$\quad | \quad \bar{x} \Leftarrow \frac{\sum x}{A}$
$\quad | \quad \bar{y} \Leftarrow \frac{\sum y}{A}$
end for
$f_{RF} \Leftarrow (\bar{x}, \bar{y})$
Step2: Calculation of colour histogram of K-mean where K represents the number of clusters
$N \Leftarrow$ total number of pixels
$N_k \Leftarrow$ total number of pixels in cluster k
$c_p \in 2^{24}$ colours possibilities
$c_{pixel} \Leftarrow$ current pixel colour
for $c_{pixel} \in cluter\ k \in K$ **do**
$\quad$ **if** $c_{pixel} \in 2^{24}$ **then**
$\quad\quad | \quad c_{pixel} \Leftarrow$ best matching colour (c_p)
$\quad$ **end if**
end for
$f_{CHKM} \Leftarrow \frac{N_K}{N}$

4.4 Low-Level Features Extraction

The idea of this paper is to extract, in addition to HF features, LF features. DHF features are uninterpretable by human brain and visually unidentifiable. Hence, it can meet the computer understanding and can be extracted by deep networks, particularly DL approaches. Recently, DL has been used in several applications, including detection, classification, prediction, ...etc. Moreover, DL is being used for deep features extraction, particularly, CNN frameworks. The working principle of CNN is to extract features maps (FMs) of each input layer, for instance, the input of nth layer if the FMs extracted from the $(n-1)^{th}$ layer. The shape of the input layer in CNN is defined as N*N*M, where N is the size of the FMs, and M represents the total number of channels considered. Figure 5 illustrates image size reduction using convolutional layers.

Several CNN networks operate on reducing image representation while going deeper into the network layers to include, Residual Neural Network (ResNet), Inception-V4, and DenseNet. The former is characterised of the use of multi-filters and auto image size reduction, however, features extraction applications using ResNet suffer from time consumption and complicated implementation. Inception-V4 applies inception blocks to avoid vanishing-gradient training problem. However, time consuming still define a problematic metric causing a high risk of information loss through reduction blocks. DenseNet, on the other hand, presents a high accurate and efficient architecture overpassing Inception-V4 and ResNet, with a low error detection due to dense connections. The architecture of DenseNet is considered as complicated processing and its efficiency decreases in case of complex datasets. For LF features extraction, an automated hybrid deep learning model has been proposed in
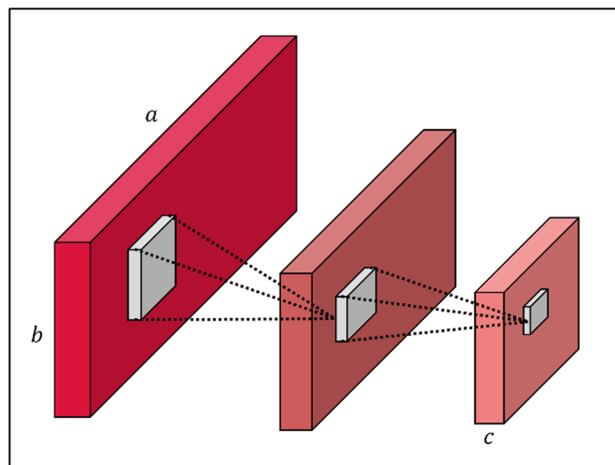

Fig. 5 Image Dimensionality Reduction Through the Convolutional Layer where (a): Feature width, (b): Feature height, (c): Number of Channels

Loukil and Salah (2020), namely, DenCeption. The idea is to construct a new dense block (DB) composed of convolutional and inception modules. This will show the effect of the concatenation operation of each convolution on the output of inception modules. The new densely connections will be translated by the dense connectivity between all inception modules within the DBs by conserving the initial dense connections between convolutional blocks (CBs). Toward minimising the size of the medical images while being processed, an integration of reduction A (Rc_A) and reduction B (Rc_B) blocks into the transition block (TB) takes place. As a result, reduction modules (Rc_A) and (Rc_B) will be densely linked to inception modules A (In_A) and B (In_B), respectively. A single inception C (In_C) module will be part of TBs as well. It will link the resulted output of (Rc_B) modules to the average pooling operation block. Figure 6 manifests the proposed hybrid network.

DenCeption efficiently shows a high DHF extraction performance compared to DenseNet-100 proposed in Nazir et al. (2021) for features extraction. Each hybrid DB produces a set of features resulted from highly dense connections

linking different internal components including Batch Normalization (BN) layer, Convolutional layer, Rectified Linear Unit (ReLU) activation function, Inception modules to include In_A and In_B. Figure 7 illustrates the hybrid DB composition.

In_A and In_B increase the dense connections opting to minimize the total channels by that it means reducing the sample representation and used for key-points rearrangements along with 3*3 convolutional layer. The final outcome of $(n+1)^{th}$ hybrid DB is $N*N*(M + \alpha * M')$, where $\alpha 2$. By increasing the dense connections composing the hybrid DB, the FMs extracted at each nth layer shows an increase as well. Therefore, the hybrid TB is taking over the outcome and reduces the extracted DHF dimension from the $(n-1)^{th}$ DB as discussed in Mutlag et al. (2020). Figure 8 presents the composition of the hybrid TB block.

The presence of reduction blocks, inherited from Inception architecture, including Rc_A and Rc_B had a fundamental value in: (1) improving the FMs representation, (2) reducing the FMs dimension, and (3) emphasizing on keeping the RoIs's DHF features (loss rate is very low). This is
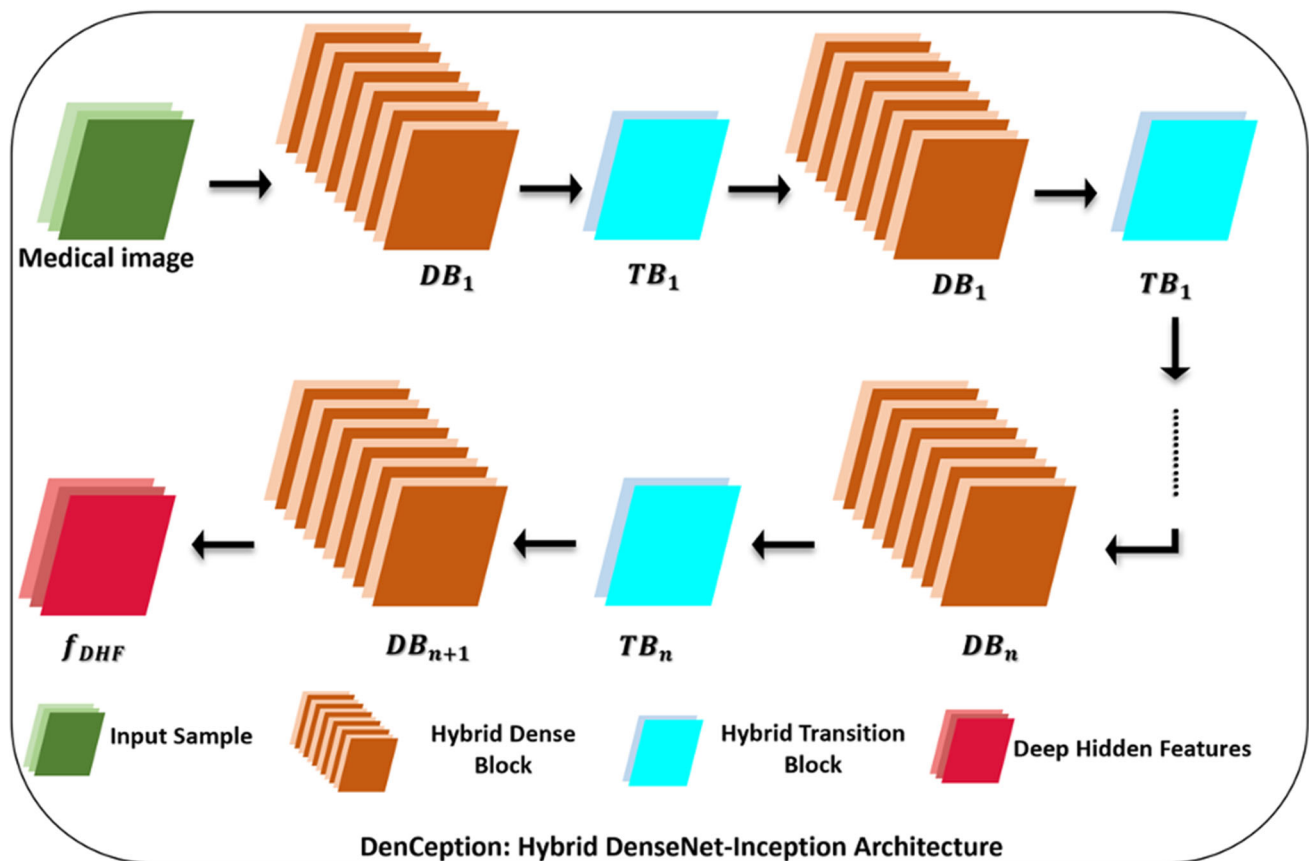

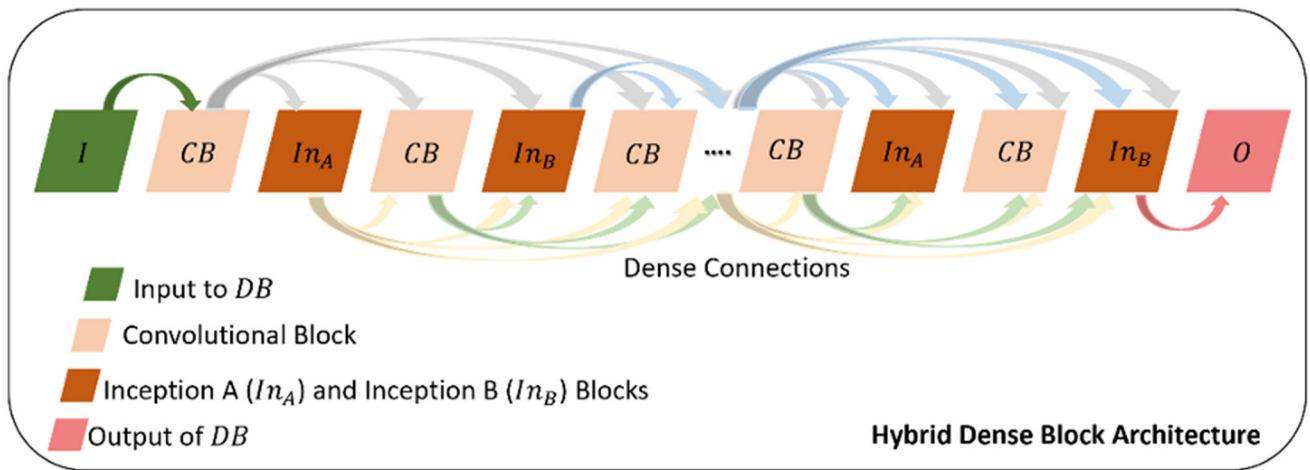**Fig. 6** Hybrid DenseNet-Inception Architecture for LF Extraction

Fig. 7 Hybrid Dense Block Architecture

another advantage of DenCeption in comparison to classical DenseNet architecture.

4.5 Features Weighting

4.5.1 Features Initialization and Normalization

Towards determining the approximate optimal degree of influence of each extracted feature, weighting presents a crucial step for relevant features selection. The weighting technique used in this work is to assign random initial weights. Let $F = [f_{ASM}, f_E, f_C, f_H, f_{Coa}, f_{Dir}, f_{RF}, f_{CHKM_f}, f_{DHF}]$ is the matrix of extracted features and $W = [W_{ASM}, W_E, W_C, W_H, W_{Coa}, W_{Dir}, W_{RF}, W_{CHKM_f}, W_{DHF}]$ is the associated weights vector. After randomly assigning weights to each particular feature, a feature normalization is applied which produces, subsequently, a normalized features weights having as values in the range of [0,1]. This conveys the following relationships (Eq 18):

$$F(x_i, w) = \sum_{j=1, w_j \in w}^{k} w_j x_{ij} \qquad (18)$$

where w_j is the weight associated to x_{ij} feature $\in k$, and k is the number of features.

4.5.2 Weights Regularisation

Several weighting techniques are introduced in the literature including, logistic regression, random forest classifier, Bayesian linear model, ...etc. Each feature's importance training model aims to update the assigned weights while training the network. The efficiency of each aforementioned technique is mainly linked to the size of the dataset being trained which most likely can cause overfitting, underfitting and vanishing problems, as mentioned earlier. Pre-defined dataset classes is also an essential requirement for most of these techniques. Therefore, the aim of this step is to use an unsupervised machine learning approach, namely, SOM. Initially, in the learning phase, SOM associates W as the random input weights vector with the artificial neurons, namely, units of the network. Then, each input feature vector f ∈ F is presented to all units in the SOM. The unit with most similar weights to the input vector becomes the best matching unit, namely BMU. Based on the Euclidian distance, BMU

Fig. 8 Hybrid Transition Block Architecture
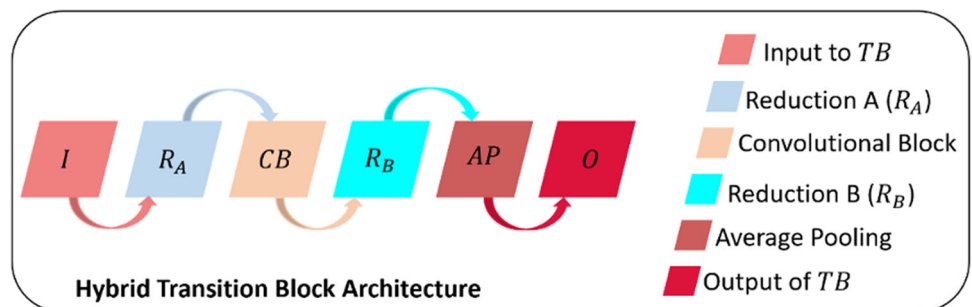
is defined as follows (Eq 19):

$$BMU = arg_i min \| f - w_i \| \tag{19}$$

Once the BMU is calculated, the weight vector is updated following the following equation (Eq 20):

$$w_i(k+1) = w_i(k) + \delta(k)\Delta_i(BMU, k)(f - w_i(k)) \tag{20}$$

The SOM training iterations ends when all features have been assigned updated weights. The feature vector with higher weight represents the feature with higher importance and vice-versa. The closer w_i is to zero, the more irrelevant the related feature is.

4.6 Features Fusion

Towards constructing a more robust features extraction outcome, capable of efficiently using multiple types of medical image and can lead to a high classification and prediction performances, the purpose of the proposed framework is to combine the HF features including those part of texture, shape, and colour, as well as fusing HF and DHF features as a following step. A set of experiments will be held to identify the optimal features combination that will feed into the disease classification stage.

4.6.1 High-Level Features Fusion

The first combination is presented by fusing texture and shape features considering their obtained weights from SOM. Therefore, there is no need to design a linear model with a fixed proportion and iteratively determining its value in order to update the fused features. Avoiding that, the features fusion is presented as the following (Eq 21):

$$F_{texture-shape} = max\left(0, \sum_i^k w_i f_{texture} + \sum_j^m w_j f_{shape} + b_1\right) \tag{21}$$

where k is the number of texture features, m is the number of shape features, and b_1 is the bias. Same operation is applied for other considered combinations including: (i) shape-colour, (ii) texture-colour, and (iii) texture-shape-colour, defined as the following (Eq 22, 23, 24):

$$F_{shape-colour} = max\left(0, \sum_i^m w_i f_{shape} + \sum_j^n w_j f_{colour} + b_2\right) \tag{22}$$

$$F_{texture-colour} = max\left(0, \sum_i^k w_i f_{texture} + \sum_j^n w_j f_{colour} + b_3\right) \tag{23}$$

$$F_{texture-shape-colour} = max\left(0, \sum_i^k w_i f_{texture} + \sum_j^m w_j f_{shape} + \sum_t^n w_t f_{colour} + b_4\right) \tag{24}$$

where n is the number of colour features and (b_2, b_3, b_4) are thebias considered for shape-colour, texture-colour, and texture-shape-colour fusion operation, respectively. The updated weights are obtained by using feedforward Artificial Neural Network (ANN). The resulted updated weights are defined as $W^{'} = [W^{'}_{f_{texture-shape}}, W^{'}_{f_{shape-colour}}, W^{'}_{f_{texture-colour}}, W^{'}_{f_{texture-shape-colour}}]$ and $F^{'} = [f_{texture-shape}, f_{shape-colour}, f_{texture-colour}, f_{texture-shape-colour}]$ is the combinations vector.

4.6.2 High- and Low-Level Features Fusion

At this stage the optimal HF features combination is considered. Let $F_optimal = f_{i-j}$ where i,j are the optimal selected HF features fusion $\in F^{'}$. Following the fusion strategy applied for HF features, $F_{optimal}$ and DHF combination is defined as the following (Eq 25):

$$F_{F_{optimal}-DHF} = max\left(0, \sum_i^t w^{'}_i F_{optimal} + \sum_j^s w_j f_{DHF} + b_5\right) \tag{25}$$

where t is the number of optimal features fusion, having as possible values k+m,m+n,or k+n, s is the number of DHF features, b_4 is the bias and $w^{'}$ is the optimal features fusion weight vector. Updated weights are obtained using ANN.

4.7 Experiments and Results Discussion

4.7.1 Dataset

In this work, two datasets have been used to test and validate the proposed features extraction framework including BraTS data obtained as part of the RSNA-ASNR-MICCAI Brain

Tumor Segmentation challenge 2021 Altaf et al. (2017); Kavya and Padmaja (2017); Liu et al. (2020), as well as an OCT dataset from Kaggle datasets, namely Retinal dataset.

BraTS dataset is considered as a simple and unlabeled set of images composed of gray-scale scans containing patients with MGMT. It is composed of 2,000 cases equivalent 8,000 MRI scans. All BraTS MRI scans are available as NIfTii files (.nii.gz). These scans are presented as four different modalities acquired with different clinical protocols and various scanners from multiple institutions Altaf et al. (2017); Kavya and Padmaja (2017); Liu et al. (2020), as follows:

- Native (T1)
- Post-contrast T1-weighted (T1Gd)
- T2-weighted (T2) and
- T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes, and

Additional parameters have been considered along side with these modalities, are as follows:

- Patient age
- Survival days
- Resection status to include: GTR (Gross Total Resection), STR (Subtotal Resection), and N/A values.

These parameters will be considered in the evaluation the efficiency of the proposed features extraction framework as a validation stage.

The Retinal dataset is considered as a labeled and complex dataset composed of more than 10000 OCT coloured scans available as .jpeg files. The dataset contains two different medical scan types, to include:

- Normal OCT scans, where there is absence of any retinal odema.
- Abnormal OCT scans, with the presence of Diabetic Macular Odema (DMO) with retinal thickening associated intra-retinal fluid.

All medical scans have been through a grading and labelling system consisting of multiple layers, done by experts in the field. The Retinal dataset also indicates a set of additional parameters that have been considered in the features extraction process, are as follows:

- Patient age
- Scanning history
- Actual stage to include Early and Advanced.

The contribution of the aforementioned parameters in the proposed features extraction framework is to validate the results of the classification stage. This would reinforce the confirmation of the efficiency of the proposed method, as well as its consistency.

The consideration of different medical scan types is an important criterion in the proposed framework as it gives the whole model the opportunity to learn from different modalities and image types. Hence, gaining the ability to be scalable towards covering different existing medical scan types.

4.7.2 The Testbed

To process and evaluate the proposed framework, GHNHSFS DR Research AI Server is used. The operating system setup is based on the Ubuntu Linux distribution with its latest Long-Term Support release. NVIDIA and CUDA drivers have been installed to utilise the GPUs available in the system. The latest versions available for the RTX 2080Ti series cards have been used. The software environment considered is Python and MATLAB, in addition to the servers' tools. The programming environment used to build the models consists of Python programming language with TensorFlow as the CNN modelling framework. The latest release of Anaconda distribution of Python with all its supporting packages have been used. The only updates to the base Anaconda distribution consist of the TensorFlow and the OpenCV image processing libraries. MATLAB installation is required to provide a runtime environment for some of the tools in development. In addition to the tools required to build the models, two packages have been provided that are required in order to serve the models as tools for the use by GHNHSFS. Docker and K8S allow for containerisation of software, allowing tools and applications to run without a view or access to any components of the system. Table 5 summarises the considered testbed.

4.7.3 Research Evaluation Mechanism

The proposed experiments evaluate the capability of the proposed framework in handling different dataset cases including labeled and unlabeled. In fact, these experiments present challenging scenarios in order to obtain an efficient features extraction outcome. Add to that, each scenario covers a different dataset representation to include: gray-level coloured medical images as input samples.

To evaluate the effectiveness of the proposed features extraction method in comparison with existing works, an evaluation scheme of various measurement parameters is considered essential. Therefore, a set of experiments have been applied to: (1) evaluate whether the proposed method has compiled all the outlined requirements for HF and LF features extraction and fusion and (2) validate its accurate functionality compared to existing features extraction methods. Existing HF features extraction methods have highlighted important drawbacks such as redundant features

Table 5 Detailed Testbed

|	Specifications	Operating System	Drivers
Hardware	- 1x ASUS ESC8000	Ubuntu 18.04.3 LTS	Nvidia Graphics
	- 2x Intel Xeon Gold 5218	(kernel: 5.0.0-32-generic	Drivers 430.50
	- 16x 32GB DDR4 2666Mhz	x86_64)	CuDa 10.1
	ECC RDIMM (512GB Total)		
	- 8x GPU RTX 2080TI 11GB Blower		
	- 1x 2TB Intel 760p M.2 PCIe NVME		
	- 8x 2TB Seagate Exos		
	7E2000 ST2000NX0433		
	- 10GbE SFP+ Network Adapter		
	- 1x PIKE II 3108-8i-16PD/1G		
	Python environment	MATLAB 2019b	Servers tools
Software	- Anaconda 2019.03	Deep learning toolbox	Docker 19.03.2
	- Conda 4.7.10		Microk8s v1.16.2
	- Python 3.7.4		
	- Tensorflow 2.0.0	Parallel computation toolbox	
	- Opencv 4.1.1.26		
	- keras (version 2.31)		
	- tensorflow-gpu (v1.2.1)	Computer Vision toolbox	
	- matplotlib (v3.1.1)		
	- opencv (v4.1.2)		
	- flask (v1.2.1)	Signal Processing toolbox	
	- scikit-learn (v0.21.3)		
	- scipy (v1.3.1)		
	- numpy (v1.17.3)	Statistics and Machine	
		Learning toolbox	
	- h5py (v.2.10.0)		
	- pytorch (v1.3.0)		

extraction, processing issues due to complex datasets, particularly unlabeled datasets. Recent LF features extraction, particularly Deep Learning methods, also have shown complicated processing especially when it comes to thrashing, underfitting, and overfitting problems. The latter, particularly, is considered as a major problem occurring in case the number of layers composing the deep network is higher than the amount of data being processed. In fact, the deeper the DL network is, the higher the computational power is, which makes these methods not applicable in real-case scenarios. In fact, existing CNN architectures lack in considering the depth of the features set at this stage, alongside with how it can be scalable over time. The evaluation mechanism of related works in the literature considered in this study conveys to four main stages defined as follows:

- **Responsiveness:** it is defined as the processing time required to accurately extract relevant features when the

dataset changes. In another word, it is the performance of the proposed features extraction model to execute any type of medical images within a given time interval considering the parallel mechanism of the proposed features extraction framework.
- **Adaptability:** the proposed features extraction method is an adaptive tool. The implemented model is able to pick relevant features without external supervision (unsupervised case).
- **Scalability:** a features extraction method is considered scalable if it can be applied on any type of medical images. The proposed model is independent of the type, size, and complexity of the input dataset. The flow of the framework makes the features extraction tool scalable to any size of dataset (covering overfitting and underfitting), which in turn, results a high advantage of its application in real-time scenarios (1-5) where it does not require any external intervention (adaptive). The low-level features DenCeption network is a key-block in the proposed

Fig. 9 Classification Outcomes of Comparative Works
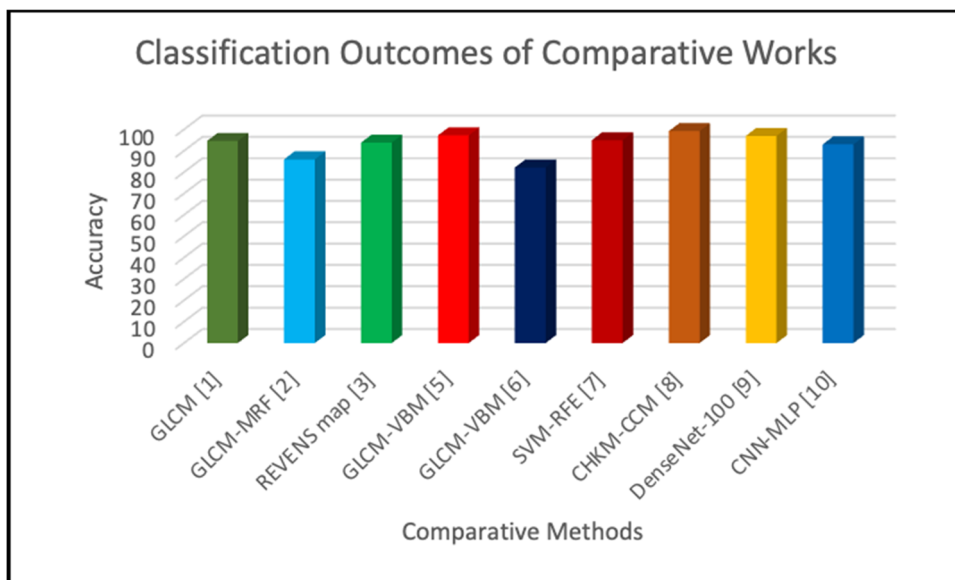
framework. In fact, it makes the process faster even in case of complex datasets due to its composition in terms of transition and reductions integrated block. Real-time scenarios evaluation method is a valuable way to making sure that the proposed model performs in the same accurate way with each scenario. Subsequently, scalability is defined as the measurement of the capability of the features extraction model to handle increase/decrease in load by scaling up/scaling out without impact on the performance responsiveness, thus, without effect on the available resources' costs including CPU, memory, disk, bandwidth, throughput, ...etc. The proposed features extraction system has the ability to meet the growing needs based on the input datasets keeping the stability of the extraction model.

- **Reliability:** it is defined when the system keeps processing correctly even when faults occur. The parallel structure of the proposed method keeps the blocks

independently processing the data. In case there will be any bug at any stage, none of the other interfaces will get affected.

Despite the high classification results achieved by the aforementioned methods, as illustrated in Fig. 9, these approaches present serious drawbacks when it comes to our suggested evaluation experiments which shows their limitations in generalisable scenarios.

Table 6 summarises the set of comparative works considered for the evaluation mechanism of the proposed features extraction method.

Based on the aforementioned comparison table, the proposed features extraction successfully covers all the applied experiments with high level of accuracy compared with other methods. The evaluation mechanism highlighted the key-elements that a features extraction model needs to include.

Table 6 Proposed Experimentation Mechanism Applied on Related Works

Features	Authors	Method	Experiment 1	Experiment 2
HF / HF-fusion	Altaf et al. (2017)	GLCM	✓	x
	Kavya and Padmaja (2017)	GLCM-MRF	x	✓
	Liu et al. (2020)	REVENS map	✓	x
	Xiao et al. (2017)	GLCM-VBM	✓	x
	Janakasudha and Jayashree (2020)	GLCM-VBM	✓	x
	Madusanka et al. (2019)	SVM-RFE	✓	x
	Lin et al. (2009)	CHKM-CCM	x	✓
LF	Nazir et al. (2021)	DenseNet-100	✓	x
	Dara et al. (2018)	CNN-MLP	✓	x
HF-LF Features Fusions		Proposed Method	✓	✓

Table 7 Compliance of Related Works with the Proposed Research Evaluation Mechanism

Method	Responsiveness	Adaptability	Scalability	Reliability
GLCM Altaf et al. (2017)	✓	-	-	-
GLCM-MRF Kavya and Padmaja (2017)	✓	-	-	✓
REVENS map Liu et al. (2020)	✓	-	-	-
GLCM-VBM Janakasudha and Jayashree (2020); Xiao et al. (2017)	✓	✓	-	✓
SVM-RFE Madusanka et al. (2019)				
CHKM-CCM Lin et al. (2009)	✓	-	-	✓
DenseNet-100 Nazir et al. (2021)	✓	✓	-	✓
CNN-MLP Dara et al. (2018)	✓	✓	-	✓

Missing one of these aforementioned characteristics leads to the inability of the method to be applicable in real-case scenarios. The responsiveness that the adaptive proposed approach demonstrated, has been complimented by its scalability and reliability. In fact, the proposed features extraction model, particularly, DenCeption, is designed specifically to handle generalizable scenarios by integrating algorithms that are compatible with most of the medical dataset being tested (e.g. MRI, OCT, ...etc.). Table 7 presents the compliance of each method to the suggested evaluation mechanism.

4.7.4 Individual Block Testing and Integration

Towards validating the accurate functionality of the proposed features extraction framework, a testing mechanism is considered to include the following two stages: Individual blocks testing Blocks integration testing Therefore, two main experiments have been considered based on real-case scenarios defined as: the first experiment (Exp 1) includes greyscale and unlabelled dataset, whether the second experiment (Exp 2) involves coloured and labelled dataset.

Individual blocks testing: In order to efficiently evaluate each block composing the proposed model, a unit testing assessment is applied based on the functionality of each block. The individual testing is composed of two main blocks to include: HF features fusion extraction block testing ($Block_1$), and LF features extraction block testing ($Block_2$). Table 8 summarises the block testing experiments done.

Based on processing time and resources usage, the application of the aforementioned experiments to test the proposed features extraction framework confirmed the optimal features fusion based on each case. In fact, HF features fusion

has three different propositions including: (1) texture-shape, (2) shape-colour, (3) texture-colour, and (4) texture-shape-colour as shown in Fig. 10.

Integration testing: Once every unit is successfully tested, integrating them is a crucial step in order to assure the efficiency of running all blocks at once. Particularly, each unit in the proposed features extraction framework is strongly dependent on previous units' outcomes. Therefore, the proposed integration testing mechanism results features fusion testing blocks to include HF-LF features fusion block ($Block_3$).

Following the independent testing of $Block_1$, the proposed experimentation mechanism is applied on the resulted features combinations of $Block_3$. Table 9 summarises the HF-LF features fusion cases to be tested and validated accordingly.

All results of the aforementioned unit testing and integration will be presented and discussed in the following section.
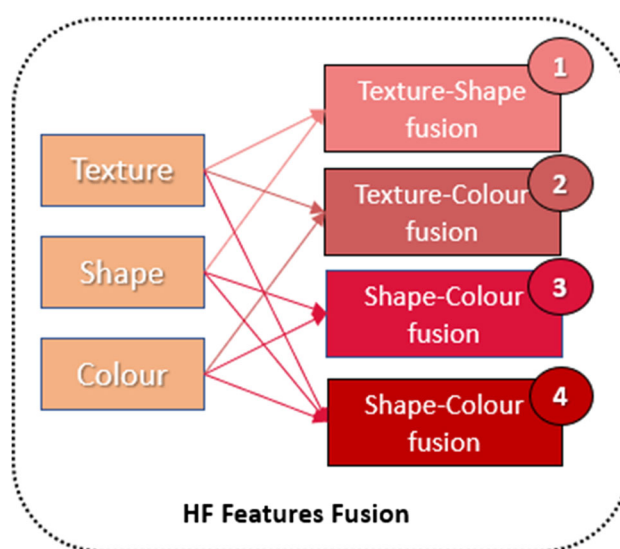
Table 8 Individual Block Testing Experimentation

Block	Unit Testing	Exp 1	Exp 2
1	HF features extraction	✓	✓
2	LF features extraction	✓	✓

Fig. 10 Proposed Integration Testing Mechanism for HF Fusion

Table 9 Integration Testing: HF-LF Cases

Case	Texture	HF Shape	LF Colour	DHF
1	✓	✓	-	✓
2	✓	-	✓	✓
3	✓	✓	✓	✓
4	-	✓	✓	✓

4.7.5 Analysis and Evaluation

Following the integration testing process, the resulted optimal features are used for classification task using SVM. The first classification task is done using BraTS dataset which covers the case of unlabeled gray-scale dataset. The second is applied using OCT dataset which conveys the case of labeled and coloured dataset. To evaluate the classification outcomes of each case a set of quantitative performance metrics is considered including: Sen, Specificity (Spe), Prevalence (Prev), Acc, and MAE, defined as follows (Eq 26, 27, 28, 29, and 30):

$$Sen = Probability(+ \mid disease) = \frac{+ \mid disease}{t_d c}$$
$$= \frac{TP}{(TP + FN) * 100} \tag{26}$$

where t_{dc} is the total number of disease cases.

$$Spe = Probability(- \mid normal) = \frac{- \mid normal}{t_{nc}}$$
$$= \frac{TN}{(TN + FP) * 100} \tag{27}$$

where t_{nc} is the total number of normal cases.

$$Prev = Probability(disease) \tag{28}$$

$$Acc = Sen * Prev + Spe * (1 - Prev)$$
$$= \frac{(TP + TN)}{(TP + TN + FP + FN) * 100} \tag{29}$$

$$MAE = \frac{1}{Q} \sum_{j=1}^{Q} \mid y_i - \hat{y}_i \mid \tag{30}$$

where disease $\in$ MGMT,DMO, (+ | disease) is the number of classified positive cases knowing that these cases actually have the disease, (- | normal) is the number of classified negative cases knowing that these cases are actually normal, TP is the true positives, FN is the false negatives, TN is the true negatives, FP is the false positives, Q is the total number of input samples, y_i is the expected classification output, and (y_i) is the actual classification output.

the experimentation process applied on both available datasets showed a clear difference in terms of responsiveness to image processing, system scalability, and reliability to reflect the parallel processing where possible.

- **BraTS Dataset:**

The specifications of the BraTS dataset have impacted the evaluation mechanism of the proposed optimal features extraction method. In fact, its responsiveness to block testing scheme has not been successful where it comes to individual block testing, except Texture-Shape fusion (see Table 10).

Moving forward into the integration testing scheme (Table 11) BraTS dataset showed a successful responsiveness, particularly in case of Texture-Shape-DHF, Texture-Colour-DHF, and Texture-Colour-Shape-DHF fusions. The lack of responsiveness presented by Shape-Colour fusion could be interpreted by the absence of Texture feature which represents a key feature in tumor disease detection and classification, which, as a result, impacted the responsiveness of the system processing.

The responsiveness gained by Texture-Coulour-Shape-DHF fusion compared to Texture-Coulour-Shape fusion is linked mainly to the increase of the sensitivity and specificity of the optimal features selection system. The gain of a deeper understanding of the input medical images impacted the performance of the system. As per this case, Texture-Colour-DHF also showed an evolutionary impact o the overall classification accuracy which helped into optimising the system processing which highlights its accuracy. Despite that, the overall system is not considered as scalable as it does not

Table 10 Compliance of Block 1 & 2 Testing with the Research Evaluation Mechanism - BraTS Dataset

Testing Block	Responsiveness	Adaptibility	Scalable	Reliable
LF based	-	✓	-	-
Texture-Shape based	✓	✓	✓	-
Shape-Colour based	-	✓	-	-
Texture-Colour based	-	✓	-	✓
Texture-Shape-Colour based	-	✓	✓	✓

Table 11 Compliance of Block 3 Testing with the Research Evaluation Mechanism - BraTS Dataset

Testing Block	Responsiveness	Adaptability	Scalable	Reliable
		LF based		
Texture-Shape based	✓	✓	✓	✓
Shape-Colour based	-	✓	-	✓
Texture-Colour based	✓	✓	-	✓
Texture-Shape-Colour based	✓	✓	✓	✓

involve the shape features, which represents, in this particular dataset, a critical parameter that reflects the evolution of the tumor and helps in optimising its detection and classification. Texture-Colour-DHF, on the other hand, is also classified as a non scalable solution for the latter reason. That said, Texture-Shape, Texture-Shape-Colour, Texture-Shape-DHF, and Texture-Shape-Colour-DHF have demonstrated a scalable solution by including a key factors that has an impact on the overall classification.

Despite being responsive, adaptive, scalable, and reliable, Texture-Shape-Colour-DHF has demonstrated aless optimal features selection system, as compared to Texture-Shape-DHF fusion (Table 12).

Subsequently, the former resulted a lower classification performance to achieve 80% Sen, 78% Spe, 79.6% Acc, and 0.21 MAE. Texture-Shape-DHF, on the other hand, presented 98% Sen, 96% Spe, 97.3% Acc, and 0.02 MAE. This confirms the high importance of Texture-Shape fusion in the case of BraTS dataset, and particularly for tumor detection. Towards validating the training stage of the proposed method, an example of testing MRI image is presented in Fig. 11. The latter shows the testing result of the tumour detection and classification alongside the indication of the resection status and the estimated survival days. The sample image belongs to a patient with GTR and 131 survival days. As per the results, Texture-Shape-DHF, and Texture-Shape-Colour-DHF based proposed systems have successfully identified the tumour with the correct specifications: GTR as resection status and an approximate survival days of 128 which is strongly comparable with the original number. However, remaining experiments have not successfully identified the validation parameters as shown in Fig. 11. Therefore, this validates the presented results in Table 13.

The high processing time presented by most of the experiments impacted by two main factors. In fact, the unlabelled input images increased the processing time of the overall system as it affects the particular processing time of the overall system as it affects the particular processing time of DHF extraction. The DenCeption model took around four hours to proceed the deep LF extraction o its own. Subsequently, this impacted remaining fusion experiments requiring parallel resources processing. In addition, the type of the processed MRI image (NifTii) is considered as a complicated input image, which has a drawback on its processing time. This, therefore, justifies the lower processing time of the experiments done on the Retinal dataset, where the latter is composed of mainly .jpeg images.

- **Retinal Dataset:**

Similar experimentation approach has been applied on Retinal dataset, As per its specificastion, the dataset include labelled and coloured OCT images. The individual block testing experiments have shown a great responsiveness of the LF as well as Texture-Colour extraction (Tables 14 and 15).

Despite the abcense of colour features in the Texture-Shape fusion, the latter showed a scalable system that has a potential ton improve the DMO classification results. However, it did not show any responsiveness or system reliability. That said, by combining it with DHF, Texture-Shape-DHF has overcomed the scalability problem but remaining non reliable, as per Shape-Colour-DHF fusion. This was not the case for Texture-Colour-DHF and Texture-Colour-Shape-DHF that showed a complete system by covering all the evaluation mechanism criteria. These results are confirmed

Table 12 Performance Results for Individual Block Testing - BraTS Dataset

Testing Block	Processing Time (h:mn:sec)	Sen	Spe	Acc	MAE
LF based	3:40:00	67.2%	53%	64.36%	0.37
Texture-Shape based	8:15:00	73%	70%	72.4%	0.28
Shape-Colour based	6:44:00	68%	67%	67.8%	0.33
Texture-Colour based	7:33:00	70%	68%	69.6%	0.31
Texture-Shape-Colour based	11:20:00	75%	71%	74.2%	0.26

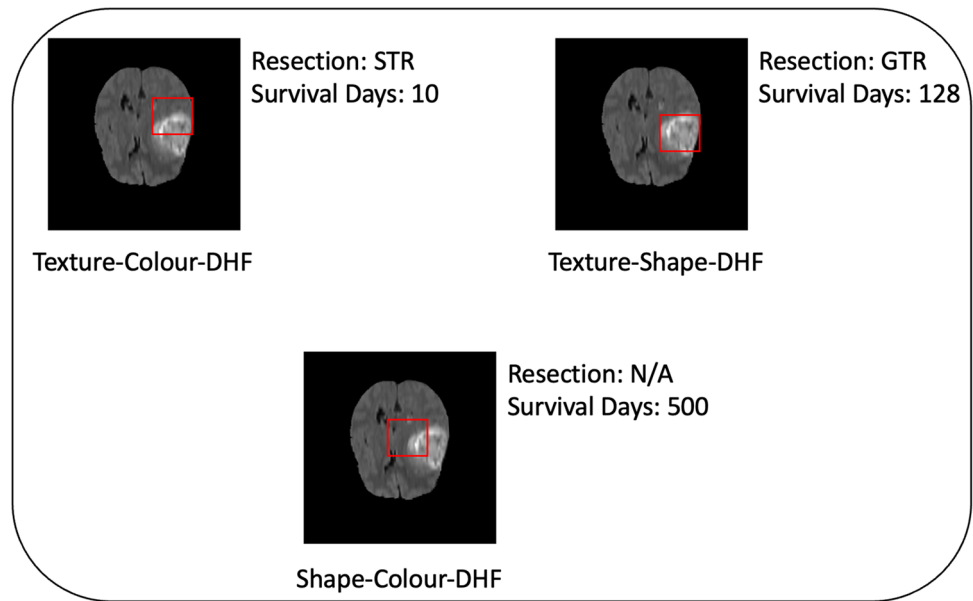**Fig. 11** Critical Sample Testing on Block 3 Methods - BraTS Dataset

Table 13 Performance Results for Integration Block Testing - BraTS Dataset

Testing Block	Processing Time (h:mn:sec)	Sen	Spe	Acc	MAE
		LF based			
Texture-Shape based	10:30:00	98%	96%	97%	0.02
Shape-Colour based	8:35:00	73%	70%	72.4%	0.28
Texture-Colour based	9:43:00	77%	76%	76.2%	0.24
Texture-Shape-Colour based	14:10:00	80%	78%	79.6%	0.21

Table 14 Compliance of Block 1 & 2 Testing with The Research Evaluation Mechanism - Retinal Dataset

Testing Block	Responsiveness	Adaptability	Scalable	Reliable
LF based	✓	✓	-	-
Texture-Shape based	-	✓	✓	-
Shape-Colour based	✓	✓	-	-
Texture-Colour based	✓	✓	✓	-
Texture-Shape-Colour based	-	✓	✓	✓

Table 15 Compliance of Block 3 Testing with The Research Evaluation Mechanism - Retinal Dataset

Testing Block	Responsiveness	Adaptability	Scalable	Reliable
	LF based			
Texture-Shape based	✓	✓	✓	-
Shape-Colour based	✓	✓	✓	-
Texture-Colour based	✓	✓	✓	✓
Texture-Shape-Colour based	✓	✓	✓	✓

Table 16 Performance Results for Individual Block Testing - Retinal Dataset

Testing Block	Processing Time (h:mn:sec)	Sen	Spe	Acc	MAE
LF based	1:30:00	68%	60%	66.4%	0.34
Texture-Shape based	5:30:00	71%	69%	70.6%	0.3
Shape-Colour based	2:50:00	69%	65%	68%	0.32
Texture-Colour based	3:00:00	77%	73%	76.2%	0.24
Texture-Shape-Colour based	4:20:00	81%	80%	80.8%	0.19

Table 17 Performance Results for Integration Block Testing - Retinal Dataset

Testing Block	Processing Time (h:mn:sec)	Sen	Spe	Acc	MAE
		LF based			
Texture-Shape based	3:30:00	84%	81%	83.4%	0.17
Shape-Colour based	2:55:00	78%	71%	76.6%	0.24
Texture-Colour based	4:45:00	91%	88%	90.4%	0.03
Texture-Shape-Colour based	5:13:00	99%	98%	98.9%	0.01
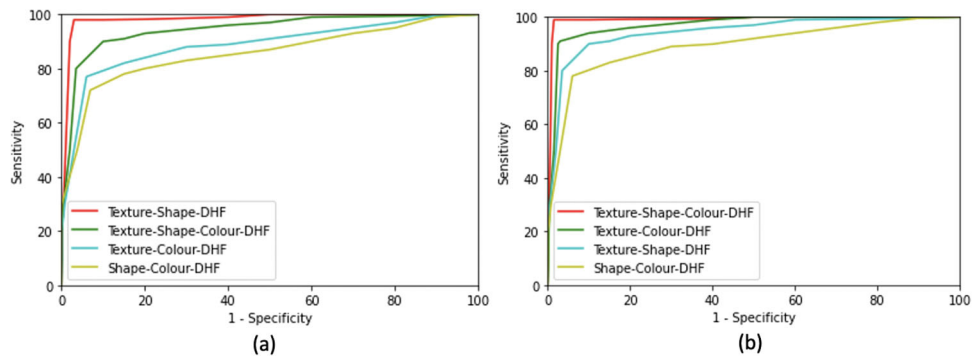
Fig. 12 ROC Graph of Block 3 Testing and Validation: (a) BraTS Dataset, (b) Retinal Dataset
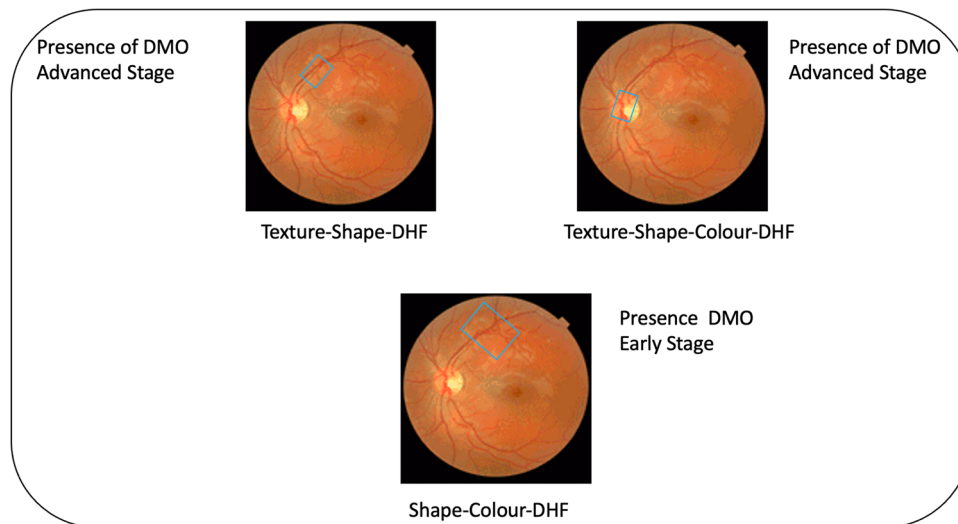

Fig. 13 Critical Sample Testing on Block 3 Methods - Retinal Dataset

by the training and testing process done on each of the aforementioned experiments (Tables 16 and 17), where Texture-Colour-DHF acheived 91% Sen, 88% Spe, 90.4% Acc, and 0.03 MAE. On the other hand, Texture-Shape-Colour-DHF demonstrated outstanding results to reach 99% Sen, 98% Spe, 98.0% Acc and 0.01 MAE. These results could be reflected by the receiver operating characteristic curve (ROC) Graph shown in Fig. 12 along with a comparison to the results obtain fron BraTS dataset.

Towards validating the obtained results in relation to the Retinal dataset, a critical sample of OCT scans is shown in Fig. 13 The latter shows the DMO detection and stage estimation of the disease (in this case Advanced stage). As per the Figure, Texture-Shape-Colour-DHF and Texture-Colour-DHF have successfully identified and detected the stage of the DMO present in the scan, which was not the case for remaining experiments. This proves the importance of the extraction of key features that can support the system in the identification of critical samples and decreases the FN rates.

5 Conclusion

Disease classification requires specialist's expertise in locating inner areas of interest from medical images, particularly, gray-scale images (MRI) and coloured fundus images OCT. That is, manual features extraction can be time consuming which might have side effects on the diagnosis and analysis process. To cope with this challenge, an automated features extraction and selection method is proposed. The framework is based on combining HF and DHF features towards achieving a high quality of medical analysis with less time consumption. A new hybrid deep learning framework, namely DenCeption, has been applied for DHF extraction alongside with high-level features extraction techniques. The optimal combination of texture, shape, colour, and DHF has been used as an input to the classification model. The main aim of the proposed method is to create a generic framework that can pick the best features combination based on the characteristics of the input dataset. Multiple experimentation have been considered to test each possible features combination and reflect that on the proposed evaluation mechanism to convey responsiveness, adaptability, scalability and reliability. The latter mechanism has also been tested on related works to validate the available bench-marking.The proposed features extraction framework achieved outstanding results on both coloured-labelled and gray-scale-unlabelled based datasets, to reach 98.9% for texture-shape-colour-DHF combination and 97% for texture-shape-DHF combination, respectively. Despite the use of other features combinations, the high impact the aforementioned combinations provided helped in intensifying the responsiveness and reliability of the pro-posed framework by minimising the false positives and false negatives that can occur.

Considering the above results, the proposed framework can be scaled to be applied in real-time experiments. Hence, the potential application of its use in second and/or first clinical line. Moreover, a disease prediction model will be designed towards testing the proposed features extraction model on scenarios other than classification. The prediction model will be mainly focused on DMO disease using OCT images.

References

Altaf, T., Anwar, S., Gul, N., Majeed, N., Majid, M. (2017). Multi-class alzheimer disease classification using hybrid features. In: *IEEE Future Technologies Conference*.

Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE computational intelligence magazine, 5*(4), 13–18.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis, 42*, 60–88.

Chowdhary, C. L., & Acharjya, D. P. (2020). Segmentation and feature extraction in medical imaging: a systematic review. *Procedia Computer Science, 167*, 26–36.

Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., & Fan, Y. (2018). A deep learning model integrating fcnns and crfs for brain tumor segmentation. *Medical image analysis, 43*, 98–111.

Conghua, X., Yuqing, S., & Jinyi, C. (2006). A new method of semantic feature extraction for medical images data. *Wuhan University Journal of Natural Sciences, 11*(5), 1152–1156.

Dara, S., Tumma, P., Eluri, N. R., & Kancharla, G. R. (2018). Feature extraction in medical images by using deep learning approach. *International Journal of Pure and Applied Mathematics, 120*(6), 305–312.

Dara, S., Tumma, P., Eluri, N. R., & Kancharla, G. R. (2018). Feature extraction in medical images by using deep learning approach. *International Journal of Pure and Applied Mathematics, 120*(6), 305–312.

Gao, Y., Ma, S., Liu, J., Liu, Y., & Zhang, X. (2021). Fusion of medical images based on salient features extraction by pso optimized fuzzy logic in nsst domain. *Biomedical Signal Processing and Control, 69*, 102852.

Hazarika, R. A., Maji, A. K., Sur, S. N., Paul, B. S., & Kandar, D. (2021). A survey on classification algorithms of brain images in

alzheimer's disease based on feature extraction techniques. *IEEE Access, 9*, 58503–58536.

Peng, H., Long, F. (2001). A bayesian learning algorithm of discrete variables for automatically mining irregular features of pattern images. In: Proceedings of the Second International Conference on Multimedia Data Mining, pp. 87–93.

Huerga, C., Morcillo, A., Alejo, L., Marín, A., Obesso, A., Travaglio, D., Bayón, J., Rodriguez, D., & Coronado, M. (2021). Role of correlated noise in textural features extraction. *Physica Medica, 91*, 87–98.

Chowdhary, C.L., Acharjya, D.P. (2020). Segmentation and feature extraction in medical imaging: a systematic review. *Procedia Computer Science* 167, 26–36.

Mutlag, W.K., Ali, S.K., Aydam, Z.M., Taher, B.H. (2020). Feature extraction methods: a review. In: *Journal of Physics: Conference Series*, vol. 1591, p. 012028. IOP Publishing.

Kaur, P., Singh, G., & Kaur, P. (2018). A review of denoising medical images using machine learning approaches. *Current medical imaging, 14*(5), 675–685.

Huerga, C., Morcillo, A., Alejo, L., Marín, A., Obesso, A., Travaglio, D., Bayón, J., Rodriguez, D., Coronado, M. (2021). Role of correlated noise in textural features extraction. *Physica Medica* 91, 87–98.

Hazarika, R. A., Maji, A. K., Sur, S. N., Paul, B. S., & Kandar, D. (2021). A survey on classification algorithms of brain images in alzheimer's disease based on feature extraction techniques. *IEEE Access, 9*, 58503–58536.

Lin, C.-H., Chen, R.-T., & Chan, Y.-K. (2009). A smart content-based image retrieval system based on color and texture feature. *Image and vision Computing, 27*(6), 658–665.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis, 42*, 60–88.

Rundo, L., Tangherloni, A., Cazzaniga, P., Mistri, M., Galimberti, S., Woitek, R., Sala, E., Mauri, G., & Nobile, M. S. (2021). A cuda-powered method for the feature extraction and unsupervised analysis of medical images. *The Journal of Supercomputing, 77*(8), 8514–8531.

Liu, L.Y.-F., Liu, Y., & Zhu, H. (2020). Masked convolutional neural network for supervised learning problems. *Stat, 9*(1), 290.

Xiao, K., Liang, A. L., Guan, H. B., & Hassanien, A. E. (2013). Extraction and application of deformation-based feature in medical images. *Neurocomputing, 120*, 177–184.

Madusanka, N., Choi, H.-K., So, J.-H., & Choi, B.-K. (2019). Alzheimer's disease classification based on multi-feature fusion. *Current Medical Imaging, 15*(2), 161–169.

Mingqiang, Y., Kidiyo, K., Joseph, R., et al. (2008). A survey of shape feature extraction techniques. *Pattern recognition, 15*(7), 43–90.

Mingqiang, Y., Kidiyo, K., Joseph, R., et al. (2008). A survey of shape feature extraction techniques. *Pattern recognition, 15*(7), 43–90.

Nazir, T., Nawaz, M., Rashid, J., Mahum, R., Masood, M., Mehmood, A., Ali, F., Kim, J., Kwon, H.-Y., & Hussain, A. (2021). Detection of diabetic eye disease from retinal images using a deep learning based centernet model. *Sensors, 21*(16), 5283.

Lin, C.-H., Chen, R.-T., & Chan, Y.-K. (2009). A smart content-based image retrieval system based on color and texture feature. *Image and vision Computing, 27*(6), 658–665.

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G.-Z. (2016). Deep learning for health informatics. *IEEE journal of biomedical and health informatics, 21*(1), 4–21.

Janakasudha, G., Jayashree, P. (2020). Early detection of alzheimer's disease using multi-feature fusion and an ensemble of classifiers. In: *Advanced Computing and Intelligent Engineering*, pp. 113–123. Springer.

Altaf, T., Anwar, S., Gul, N., Majeed, N., Majid, M. (2017). Multi-class alzheimer disease classification using hybrid features. In: *IEEE Future Technologies Conference*.

Rundo, L., Tangherloni, A., Cazzaniga, P., Mistri, M., Galimberti, S., Woitek, R., Sala, E., Mauri, G., & Nobile, M. S. (2021). A cuda-powered method for the feature extraction and unsupervised analysis of medical images. *The Journal of Supercomputing, 77*(8), 8514–8531.

Tang, F., Wang, X., Ran, A.-R., Chan, C. K., Ho, M., Yip, W., Young, A. L., Lok, J., Szeto, S., Chan, J., et al. (2021). A multitask deep-learning system to classify diabetic macular edema for different optical coherence tomography devices: a multicenter analysis. *Diabetes Care, 44*(9), 2078–2088.

Tsai, H.-Y., Zhang, H., Hung, C.-L., & Min, G. (2017). Gpu-accelerated features extraction from magnetic resonance images. *IEEE Access, 5*, 22634–22646.

Umamaheswari, C., Bhavani, R., & Sikamani, D. K. T. (2018). Texture and color feature extraction from ceramic tiles for various flaws detection classification. *International Journal on Future Revolution in Computer Science & Communication Engineering, 4*(1), 169–179.

Dara, S., Tumma, P., Eluri, N. R., & Kancharla, G. R. (2018). Feature extraction in medical images by using deep learning approach. *International Journal of Pure and Applied Mathematics, 120*(6), 305–312.

Xiao, K., Liang, A. L., Guan, H. B., & Hassanien, A. E. (2013). Extraction and application of deformation-based feature in medical images. *Neurocomputing, 120*, 177–184.

Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology, 292*(1), 60–66.

Zewail, R., & Hag-ElSafi, A. (2017). Appearance-based salient features extraction in medical images using sparse contourlet-based representation. *International Journal of Image, Graphics and Signal Processing, 9*(9), 1.

Zhao, T., Hoffman, J., McNitt-Gray, M., & Ruan, D. (2019). Ultra-low-dose ct image denoising using modified bm3d scheme tailored to data statistics. *Medical physics, 46*(1), 190–198.

Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., & Fan, Y. (2018). A deep learning model integrating fcnns and crfs for brain tumor segmentation. *Medical image analysis, 43*, 98–111.

Zainab Loukil received the MS.C degree in Information and Communication Technology Telecommunication Engineer from the Higher School of Communication of Tunis. She is currently is her final stage of her Ph.D degree in image processing and deep learning from the University of Gloucestershire. She is also a Lecturer in Computer Science in the department of Cyber and Technical Computing, School of Computing and Engineering, University of Gloucestershire, U.K. Her research interests include Software Development, Artificial Intelligence, Data Science, and Networking, largely centre on Deep Learning approaches and healthcare data analysis. She has reviewed several international journals and published in multiple IEEE conferences. She has more than 5 years of experience in software development and image processing. She is also an Experienced Academic/Researcher and a Fellow of HEA.

Qublai Khan Ali Mirza is Senior Lecturer and Programme Leader in Cyber Security with the School of Computing and Engineering, University of Gloucestershire. He received Ph.D. degree in network security and machine learning from the University of Bradford. His research interests include network security, malware analysis, artificial intelligence, network analytics, cryptography, cloud computing, and IoT security. He has reviewed several international journals and published in several IEEE conferences and journals. He has fifteen years of experience in software development, network security, and cloud migration. He is also an experienced Academic/Researcher and a Fellow of HEA.

Will Sayers is a researcher in the fields of machine learning, meta-heuristic optimisation and data science ethics. His current research projects centre around natural language processing, ethical uses of data science, and ethical issues in the data industry. He lectures on data science, programming, linear algebra, calculus, and is presently supervising several PhD students. He has undertaken exploratory data analysis and reporting professionally, as well as developing machine learning models based off that work and embedded these models into production systems . He is an editor for the Journal of Open Research software & a regular reviewer for several other journals.

Irfan Awan is a Professor of Computer Science at the University Of Bradford, UK. He holds a Ph.D. in Computer Science from the University of Bradford. His research interests include network security, communication systems and performance modelling. He has chaired and organised various international conferences and workshops and served as a technical programme committee member for several international conferences. He is a steering committee member of UKPEW workshop series, IEEE PMECT workshop series and an associate editor of several internal journals and has edited several special issues of international journals. He is a member of IEEE, BCS and fellow of HEA.