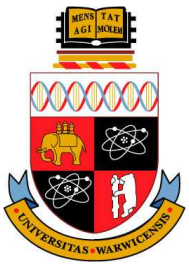
Meningioma Classification using an Adaptive Discriminant Wavelet Packet Transform



Hammad A. Qureshi

Department of Computer Science

University of Warwick

A thesis submitted for the degree of

Doctor of Philosophy

On 6th October 2009

Dedication

I would like to dedicate this thesis to my loving Father, who always believed in me and my ability and my mother who always put my education first. This thesis is dedicated to the ICT R & D Fund for posing confidence and trust in me and making it possible for me to pursue my dream. Most importantly this would not have been possible without the constant support, guidance and encouragement of my Supervisor (Dr. Nasir M. Rajpoot) and Advisor (Prof. Roland G. Wilson). I would also like to thank all my teachers who taught me over the years particularly Mr. Khizar Hayat who taught me that the pursuit of knowledge is the most noble of all professions.

Finally, this effort would not have been possible without the constant support and understanding of my wife Sana, who stood by my side while I worked hard at my PhD.

Acknowledgements

First and foremost, I want to thank my friends Dr. Mary E. Deane, Daniel Valdes-Amaro and Daniel Claus for their support and constant cheering to keep my spirits up. I would like to acknowledge the support provided to me by all the staff at Department of Computer Science, University of Warwick. I would like to particularly thank Dr. Roger Packwood, Richard Cunningham and Rod Moore for helping me on my way to achieving my goals. I would also like to make a special thanks to my colleagues Thomas Popham and Khalid Masud for their friendship and help. Most importantly I would like to thank Dr. Tim Nattkemper and Dr. Metin Gurcan for there guidance and support during my PhD. I am also deeply grateful to Prof. David B. Epstein for supporting my research work and including me in his Royal Society Grant.

This section would not be complete without thanking my friends from the Computing Society namely, Richard Warburton, Sadiq Jaffer, Tim Monk and Nick Ridgeway. All the gaming with these guys made my stay at Warwick much more interesting and pleasurable. Last but not the least, I am very grateful to my cousin Tabi Siddiqui and his wife Rehanna Siddiqui for supporting me during my stay in England and for being there when I needed them.

Abstract

Meningioma subtypes classification is a real world problem from the domain of histological image analysis that requires new methods for its resolution. Computerised histopathology presents a whole new set of problems and introduces new challenges in image classification. High intra-class variation and low inter-class differences in textures is often an issue in histological image analysis problems such as Meningioma subtypes classification. In this thesis, we present an adaptive wavelets based technique that adapts to the variation in the texture of meningioma samples and provides high classification accuracy results. The technique provides a mechanism for attaining an image representation consisting of various spatial frequency resolutions that represent the image and are referred to as subbands. Each subband provides different information pertaining to the texture in the image sample. Our novel method, the Adaptive Discriminant Wavelet Packet Transform (ADWPT), provides a means for selecting the most useful subbands and hence, achieves feature selection. It also provides a mechanism for ranking features based upon the discrimination power of a subband. The more discriminant a subband, the better it is for classification.

The results show that high classification accuracies are obtained by selecting subbands with high discrimination power. Moreover, subbands that are more stable i.e. have a higher probability of being selected provide better classification accuracies. Stability and discrimination power have been shown to have a direct relationship with classification accuracy. Hence, ADWPT acquires a subset of subbands that provide a highly discriminant and robust set of features for Meningioma subtype classification. Classification accuracies obtained are greater than 90% for most Meningioma subtypes.

Consequently, ADWPT is a robust and adaptive technique which enables it to overcome the issue of high intra-class variation by statistically selecting the most useful subbands for meningioma subtype classification. It overcomes the issue of low inter-class variation by adapting to texture samples and extracting the subbands that are best for differentiating between the various meningioma subtype textures.

Contents

List of Figures

Nomenclature

$\mathcal{B}$	Wavelet Bases
$\mathcal{B}^*$	Best Wavelet Basis
$\mathcal{D}$	Discriminant Function or Discrimination Power
Υ	Training data-set for Best basis computation
I	Image
I_g	Grey-scale Image
I_R	Image Red channel
I_G	Image Green channel
I_B	Image Blue channel
μ	Mean
$\mathcal{T}^a$	Texture template for class a
$\mathcal{P}$	Sum of pair-wise Discrimination Powers
h	Stability coefficient
$\mathcal{S}$	Stability measure or index
κ	Subband viability function
$\mathcal{R}$	Rand Index
$\mathcal{M}$	Distance function usefulness in terms of Stability and Accuracy
Ξ	Classification accuracy
$\mathcal{F}$	Feature set from Greedily selected subbands

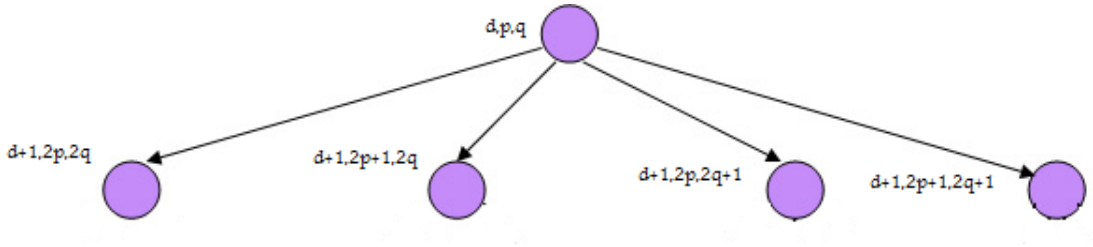


Figure 1: Nomenclature to denote nodes of a wavelet packet decomposition

Chapter 1

Introduction

Meningiomas are tumours of the brain and the nervous system [1]. They occur in the covering of the brain and the nervous system called the Meninges. Meningiomas account for 27% of all brain tumours and exist in three different grades of malignancy (WHO Grade I-III) [2], most being benign (over 80%) but some showing an increased propensity of recurrence with rare cases being malignant. Meningioma patients may live for up to five years with treatment [3]. A greater frequency of occurrence is seen in ageing populations with very rare cases reported amongst children. Most benign WHO Grade I Meningiomas belong to one of the four subtypes shown in Figure 1.1. The problem of Meningioma subtype classification essentially involves discriminating between four different subtypes of Meningiomas.

Fibroblastic, Meningiothelial and Transitional are the most common Meningiomas whereas Psammomatous occurs less frequently. There are five more Grade I Meningiomas which occur rarely. Each Meningioma subtype has certain characteristics that differentiate it from the other subtypes circled in Figure 1.1. They represent the various stages of tumour development. The first stage is the Meningothelial when the cells join together and become lobulated. A grouping of cells is formed called the synctium. The next stage is the Fibroblastic when the presence of stress in the tissue causes tearing. This in turn causes the formation of a matrix. Since the tissue is stretched, the cells become elongated and spindle shaped. The tears in the tissue results in spaces in the tissue which are filled by proteins causing formations called collagen. Transitional is the next stage where

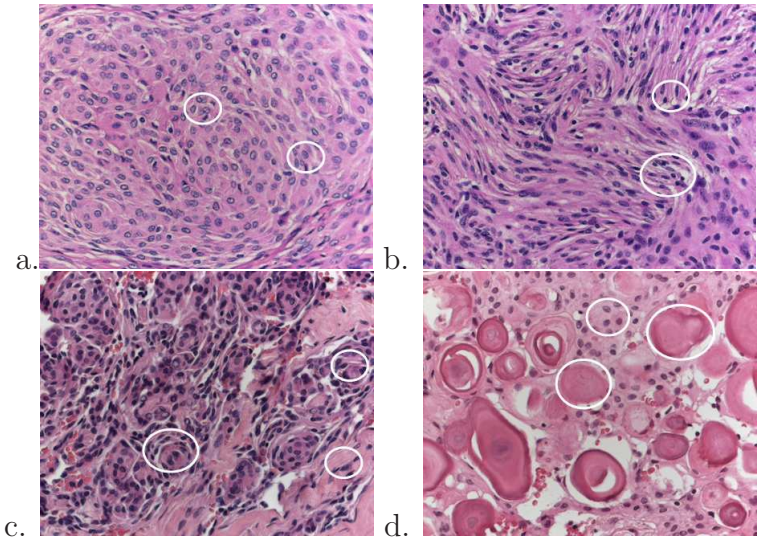


Figure 1.1: Meningioma images for each subtype a. Meningiothelial (cells form synctium), b. Fibroblastic (spindle shaped cells in collagen-rich matrix), c. Transitional (cells form whorls with Psammoma bodies present), d. Psammomatous (high number of Psammoma bodies)

whorls or circular formations can be found in the tissue. Tumours at this stage have both Fibroblastic and Psammomatous features present. The next stage is the Psammomatous where Psammoma bodies are formed. Psammoma bodies are circular objects that look like blobs. They form when the whorls from the previous stage surround themselves with proteins. At this stage, the tumour may appear as a protrusion from the surface of the brain. At this stage, the tumour may become lethal even if its benign because the protrusion would press against the skull and the brain causing bleeding and damage to brain cells. The other five Meningioma subtypes in WHO Grade I are not included in the study. Grade II & III are not included in the study.

As mentioned earlier, this study is dedicated to the issue of Meningioma subtype classification from the domain of histopathological image classification. The problem may be defined as follows:

Given a histology image of a tissue section of a Meningioma, assign a class label

to it from four different Meningioma subtypes.

Meningioma classification is a real world problem from the domain of medical image analysis that requires efficient pattern recognition. In the last five decades, a lot of research work has gone into developing a general purpose machine based pattern recognition system but the goal to this date remains elusive. Many novel and efficient pattern recognition algorithms have been invented with the data in most studies acquired using controlled conditions. For instance, in case of image classification Brodatz textures have been widely used. Recently the focus of research in image classification has shifted from synthetic textures or textures acquired using controlled conditions to real world textures. Real world problems, however, present a whole new set of challenges to the pattern recognition community. Many techniques that work exceptionally well in the lab, fail or do not perform as well in the real world where conditions are not perfect. In this thesis, we aim to present one real world problem from the domain of histological image processing and develop an algorithm to solve it using transform-based image processing and machine learning techniques. The aim of our study is to resolve the problem of computer-based Meningioma subtype (brain tumour) classification.

In this chapter we describe why Meningioma subtype classification is an important problem that requires solution. Section 1.1 describes the process of acquisition of Meningioma image data. Section 1.2 presents the reasoning behind the use of computational techniques in histological image classification. Section 1.3 presents a synopsis of the related work in the area of histological image analysis. Section 1.4 describes the main challenges in Meningioma subtype classification. Section 1.5 analyses the two main approaches of textural analysis for Meningioma subtype classification. The chapter ends with a section on main contributions of this thesis and another section on the thesis organisation.

1.1 Image Acquisition

Acquiring a Meningioma sample entails performing a brain biopsy which is fraught with risks and can be complicated. In many instances the samples are extracted

during surgery rather than taking a biopsy sample (as in the case of colon and prostate). Once the tissue samples are obtained the image acquisition tasks may be carried out.

The first stage of image acquisition is slide preparation. The tissue is placed on the slides using formalin-fixed paraffin-embedding and then cut into different sections and cross-sections. Next, Hematoxylin and Eosin (H & E) staining is carried out. The formalin-fixed paraffin-embedding and H & E staining are the most popular techniques for tissue slide preparation. The H & E staining renders the nuclei as blue whereas the rest of the tissue texture is turned pink. The next stage in slide preparation is the cutting of the tissue using a machine called the microtome. It is used to cut thin slices of the tissue for microscopic inspection.

The next stage is to analyze the tissue slides under a light microscope. A Zeiss Axioskop 2 plus microscope fitted with a Zeiss Archoplan 40×/0,65 lens is used for the purpose. Manual focussing and automated background correction is carried out. True colour RGB pictures are taken at standardised 3200 K light temperature. TIF format pictures are obtained using the Zeiss AxioVision 3.1 software and a Zeiss AxioCam HRc digital color camera (Carl Zeiss AG, Oberkochen, Germany). The microscope is integrated with a computer system on which the images are viewed and stored. This system allows for high throughput capture of histological image data.

1.2 Computer Assisted Diagnosis (CAD): Why?

Histopathological diagnosis of tumours of the brain and the spinal cord still requires decision making by human experts. Diagnosis and decision making is hampered by the fact that the reviewing of the histological slides is time consuming, prone to error, dependent upon the availability of the expert and more importantly there is considerable inter-rater variability amongst the experts. Although much effort has been expended to exactly define diagnostic criterion for all tumour entities within the World Health Organization (WHO) Classification of Tumours [2], the inter-rater variability still remains considerable (see e.g., [4]). Hence, a bias is introduced which influences further therapy regimens. Therefore, there is a need for an automated computer based technique to introduce more

objectivity in to the analysis. Moreover, computerised analysis may lead to identification of intrinsic textural features in tumour samples that are missed by the human eye and hence would lead to an improvement in definition of the tumour types as per their characteristics.

Most Meningiomas are benign [5] which means that neuropathologists are spending most of their time analysing and diagnosing benign tumours. Therefore, there is an urgent need to develop automated techniques to aid the neuropathologist. This would lead to reduction in costs and more objective and accurate diagnosis.

Due to the progress in digital image retrieval and analysis technologies, computer-assisted decision making can be used to support histopathologists by providing more objective diagnostic parameters (that may be used to define the tumour categories better) and allow high-throughput analysis. A first step, however, when developing new algorithms for image classification is to test whether an automated technique can reproduce human assignment of single tumour samples to diagnostic classes. In order to develop such a technique, we have focussed on Meningiomas. Correct histopathological diagnosis can be made in most cases by a trained human expert i.e. a neuropathologist. Therefore, this tumour is well suited for testing diagnostic properties of a computer assisted diagnostic system.

In the past decade or so, histopathology has advanced a great deal. The invention of new more advanced microscopes and advancement in digital imaging technology has opened new horizons and new possibilities for histopathological analysis. The great increase in computational power and invention of whole slide digital scanners has made the digitisation and the mass storage of histopathology slides possible. Therefore, the application of computerised image analysis and machine learning techniques for CAD has become possible. CAD algorithms are already in use to aid a radiologist and new CAD algorithms are being developed for disease detection, diagnosis and prognosis to help a pathologist [6]. Gurcan *et al.* [6] describe how different image analysis approaches have been applied to histological imaging combined with different machine learning techniques and feature selection approaches. In the next section, we describe the main works in the domain of histological image analysis.

1.3 Related Work in Histological Image Analysis

Since the advent of surgery, the naked eye examination of tissues and organs has come to be used for diagnosis of ailments. After the invention of non-invasive tools to obtain medical images, histopathology seems to have lost it predominant role in medical diagnosis. Although many advances have been made in non-invasive mechanisms such as Magnetic Resonance Imaging (MRI) etc., histopathology is still the most reliable tool for accurate diagnosis (as proved in various studies [7] [8]). The reason for this is that the non-invasive methods though are very convenient to use but are unable to provide the amount of information that may be obtained through histopathology. Non-invasive techniques such as MRI and Positron Emission Tomography (PET) provide limited resolution and poor spatial information.

Advances in histological image analysis have created new challenges as well. One of the most important benefits of the recent advances is that histopathological image data is now available at very high resolutions. This significantly improves our capability for better and more efficient diagnosis but increases the difficulty in diagnosis as huge volumes of image data are now obtained which consequently means that the task of analysing them is very time consuming and prone to error. Therefore, use of automated techniques for histopathological image analysis has become crucial. Moreover, the speed and quality of diagnosis has to be improved as well for better health care.

Apart from providing efficient diagnosis, there is also a need to extract complex information from images which is difficult to obtain merely by visual inspection. The analysis of histopathology data is crucial for treatment as it provides important information about the extent of malignancy and the developmental stage of the disease under study. Advances in histopathology and imaging technologies will enable us to be better able to define clinical courses and better treatment regimens. An interesting review of the developments in histological image analysis was carried out by Loukas and Linney [9].

Loukas and Linney contend in their paper that success of radiotherapy depends a great deal upon the biology of a tumor. This is another reason why a

detailed image analysis of the histological specimen is crucial. The proliferating activity of the tumour influences a great deal, the outcome of radiotherapy [8] [10] [11] [12] [13]. The extraction of potential prognostic information is also important [14]. The study of computational techniques applied to histopathology can be broadly categorised into the following areas.

1.3.1 Segmentation

There have been various techniques used for segmentation in histology. Some of these techniques are presented here.

1.3.1.1 Thresholding

The first studies in histopathological image analysis concentrated on thresholding. The principle of thresholding is based upon the premise that an object and its background can be differentiated by their grey-level values. Hence, the cells in a histopathological image specimen can be distinguished from the background intensities by carefully choosing a grey-level threshold. In a survey paper Sahoo *et al.* [15] have grouped thresholding into four categories; point dependent global thresholding, region-dependent global thresholding, local thresholding and multi-thresholding. In many instances thresholding is carried out using the histogram. The technique is simplistic but has found wide applicability in segmentation of meaningful structures and cell counting etc [16] [17] [18]. In certain techniques, thresholds could be set interactively [19] [20] [21]. In a study [22], a local thresholding mechanism was developed based upon an iterative selection method to segment histological images into nuclei and background. An interesting form of thresholding is carried out by diffusion with which an image is transformed into a family of threshold surfaces where choosing a threshold surface is analogous to applying a scale [23].

Automated thresholding techniques involve colour-based image analysis where coloured objects of interest are segmented from surrounding structures [24]. Sharipo [25] presented a technique for quantifying muscle content by simultaneously thresholding three monochrome images i.e. red, green and blue. A few other works of interest in this area were carried out by Dobrinski *et al.* [26], Lehr [27] and

Deverell [28]. Colour based analysis has also involved experimenting with other colour representation methods, namely hue, saturation and intensity (HSI) [29] which closely approximates the behaviour of the human eye [30]. Hue, saturation, luminance (HSL) [31] and hue, saturation and density (HSD) [32] are other models that have been studied. Pixel classification in the RGB space has also been researched extensively. Some approaches have chosen to use a combination of RGB or HSI components to achieve the best discrimination [33] [34]. Recently, novel imaging systems that take advantage of spectral features have also been proposed [35] [36].

1.3.1.2 Region Merging

Region merging or growing is another area which has received a lot of attention and a lot of research effort has gone into it. One such algorithm is a multistage one [37] for detection of membrane structures in kidney electron micrographs. Kate *et al.* [38] use a similar method for counting mitoses (represented by compact and dense objects) in stained breast cancer sections. Merging pixels and regions is based upon employing pixel proximity and grey-level uniformity [39] for the detection of cell nuclei in histopathological images acquired using a confocal laser scanning microscope. There are other interesting region-based segmentation techniques including seeded volume growing and 3D watershed algorithms [40].

1.3.1.3 Edge Detection

Edges are important features of images and are widely used in image classification and analysis. Some of the first filters used for edge detection were Prewitt, Roberts and Sobel [30], which approximate the first derivative of the image brightness function used for enhancing edges. Some studies that use edge detection for cell identification may be found in [41, 42]. First, the localisation of the intensity peak is obtained followed by radially directed edge detection in respect to the peak found. Laplacian and Laplacian of Gaussian operators have also been used for edge detection [43, 44].

1.3.1.4 Template Matching

A kind of template matching method is the Active Contour Model (ACM) [45] also known as snakes. An active contour is an energy minimising spline that searches for the local energy minimum on a potential surface. Snakes have been used for the detection of cell boundaries of nerve cells [46]. An elliptical Hough transform to locate interior points is followed by snakes to detect the interior and exterior cell boundaries. As the image complexity increases, snakes can converge to a wrong nuclear boundary with the result heavily dependent upon the choice of the initial snake. A similar effort was carried out by Bond *et al.* [23] who first performed segmentation using snakes and then used shape models for texture segmentation.

Statistical models have been used to overcome the drawbacks of snakes. Examples of such efforts can be found in [44, 47] where optimal cellular boundaries are found by exhaustive search and an optimal set of pixels are derived based upon a Bayesian metric dependent upon a statistical model. Dynamic matching methods such as the ones suggested by Yamada *et al.* [48] and Wu *et al.* [49] where polygons or parametric images are used to approximate the target image have also been used for histological image analysis.

1.3.2 Texture Analysis

Another approach to describing a texture is to quantify the content. Texture can be generally described as a repetition of patterns over an image region. Image processing carried out over texture can be broadly categorised into three types; statistical, structural and spectral [30]. Statistical analysis involves classifying images based upon the statistical features acquired from an image. Structural techniques deal with the arrangement of image primitives and texture geometry whereas spectral techniques analyse texture in transformed domains such as the Fourier, Gabor and Wavelet domains.

1.3.2.1 Statistical Approaches

Statistical approaches use statistical properties from gray level histograms (skewness and kurtosis etc.) and co-occurrence matrix features (energy, entropy, inertia

etc.) [50] to describe texture. A texture may be characterised as smooth, coarse or regular. Co-occurrence matrices have been used to characterise textures based upon spatial homogeneity [51]. A comparison of Gray Level Co-occurrence Matrices (GLCMs) for Meningioma classification with spectral approaches was carried out by Qureshi *et al.* [52]. The technique showed that spectral approach (Wavelet Packets) performed much better than statistical analysis using GLCMs.

Using computer software provided by microscope manufacturers or off-the-shelf software for statistical analysis of histological slides is widespread in histopathology. Mutter *et al.* [53] perform pre-cancer analysis of patient tissues for probability of it developing in to cancer using a statistical measurements software QProdit 6.1. Orbo *et al.* [54] used QProdit to perform morphometric and statistical analysis of endometrial hyperplasia histology samples. Inman *et al.* [55] use ImageJ for thresholding and statistical co-localisation in histological tonsil specimens for automated counting. Angelini *et al.* [56] perform computerised statistical analysis for breast tumour diagnosis.

1.3.2.2 Structural Approaches

Structural approaches have been used widely in medical image analysis for tasks such as locating diagnostically useful areas [57], discriminating between normal and cancerous tissues [58], classifying tumours of various grades [59] and quantifying their architectural organisation [60]. Such techniques have also been used in locating abnormal tissue areas in colorectal histology [57] [58]. Fractal geometry has also been used to study texture and perform texture classification [61] [62] [63]. Nuclear dimensions have also been analysed using Fractal geometry [64].

1.3.2.3 Spectral Approaches

Fourier domain texture analysis for image classification [65] and its comparison with standard texture measures [66] has been carried out in various studies. Lessman *et al.* [67] used wavelet transforms for analysing textures in Meningioma images. We have used wavelet packets with statistical analysis [68] [69] for Meningioma subtype classification. Recently some interesting work with wavelet

packets has been carried out by Katouzian *et. al.* [70] who employ wavelet packets for coronary artery segmentation. Jafari-Khouzani and Soltanian-Zadeh [71] use multi-wavelets for grading of histological prostate specimens.

1.3.3 Classification

Recently, a lot of research work in histological image understanding has been carried out. Many different types of image descriptors have been used ranging from geometric features (distance, perimeter, size and shape) to colour models (RGB, HSI etc.) and texture content-based features. The choice of classifier has also varied but linear discriminant analysis remains the most popular [72].

Like any other classification problem, the first stage in classification of histological images is the feature extraction process followed by statistical decision making [73]. A broad range of features may be extracted using color analysis, template matching, texture analysis, frequency domain techniques and surface modelling [74]. Esgiar *et al.* [58] have used pattern classification for colonic mucosa images where mutually exclusive training and testing data sets were used for first training a linear discriminant function and then subsequently testing it using the trained function. Similar studies have been carried out using this framework [75] [76]. Amaral *et al.* [77] perform pixel classification using posterior probabilities assigned using colour for scoring breast tumour microarrays.

Histopathological images are difficult to automatically characterise and segment because of their inherent complexity. Therefore, various mechanisms have been used for computer based histological image classification such as expert systems [78], morphological cellular features [79], machine vision techniques based upon segmentation of regions and diagnostic decision support systems [80], knowledge-based segmentation [81] [82] and fuzzy logic based techniques [83].

Artificial Neural Networks (ANN) have also found wide application in pathology as they have an added advantage that features may be computed by self-organisation which ultimately aid classification [84] [67]. Dytch and Wied [85] used neurocomputing for histological image understanding. Many other ANN-based efforts have been attempted for cell classification [86] [87] [88]. Lately, ANNs have also been used for segmentation though to a limited extent. One

such effort was carried out by Sjostrom *et al.* [89]. Amaral *et al.* [90] compare Gaussian process ordinal regression with a multi-layer perceptron for breast tumour scoring in breast tissue microarrays and conclude that perceptron performs better. A comparison of various neural networks with a neural network based feature reduction scheme for cervical cancer detection was presented by McKenna *et al.* [91] where high accuracy results were obtained.

1.3.4 Meningioma Classification

In [67], Meningioma subtype visualisation has been attempted by Lessman *et al.* who described how self organising maps can be combined with wavelet transforms for clustering of Meningioma images. Lessmann *et al.* in a subsequent work [92] describe how the various features extracted affect Meningioma clustering and a detailed analysis is presented. Work on Meningioma cell classification was carried out by Wirjadi *et. al.* [93] who uses a supervised learning method called Classification and Regression Trees (CART). The first approach employs textural analysis whereas the latter uses structural information derived after segmentation. However, the method proposed in [93] only differentiates between two types whereas Lessman [67] has produced good clustering results for four Meningioma subtypes.

1.4 Challenges in Meningioma Subtype Classification

Histological image analysis presents a new set of challenges to the scientific community. Histological images are real world data and are considerably different from synthetic textural data which is acquired or obtained using controlled procedures. Histological images have a uniquely complex texture which represents a new set of issues. The texture in histological images is more or less non-homogenous i.e. different areas in an image may have different textural properties which in turn may represent different patterns. Hence, textural analysis and subsequent recognition is not straightforward. Moreover, intra-class variation amongst the samples belonging to the same class may be high and to make

matters worse inter-class differences amongst the samples may be low. The problem is made more difficult by the fact that a very high degree of classification accuracy is required. An inaccurate analysis would have dire consequences for the patient.

Figures 1.2-1.5 show all the image samples acquired from various patients used in the study. These images are 1300×1100 in size and are later divided in to images of size 512×512. An image of 512×512 is cropped from the larger image and then the cropping window is moved 256 pixels forward in the horizontal direction. The same process is repeated in the vertical direction yielding 4×3 i.e. 12 images. The images acquired are overlapping which is analogous to how a neuropathologist would analyse a slide moving a small distance in each direction at a time. This though would introduce a bias making it easier to associate textures with each other but the training and test data is divided based upon patients which means that the entire image data belonging to a patient is left out in the training stage. Since there is considerable intra-class variation, the test data is most often very different from the training data keeping the test/training data mutually independent.

As stated earlier, there are a total of four Meningioma subtypes included in the study namely Meningiotheliamatous, Fibroblastic, Transitional and Psammomatous. The data for 5 patients per Meningioma subtype is available, with 48 images per patient (after dividing the original 4 images per patient), giving a total of 240 images per Meningioma subtype. Hence, for four Meningioma subtypes the total number of images available for analysis is 960. The gold truth or the labels to the data were assigned by a trained neuropathologist. The data was analysed for the relevance of Meningioma textural features to there assigned class by Lessmann *et al.* [92].

As can be seen from the Figures 1.2-1.5, the textural data in images of the patients belonging to the same Meningioma subtype can differ to a great extent. This can clearly be observed in the case of patients 3 and 4 for Fibroblastic samples shown in Figure 1.3. The tear in the tissue of the patient 3 is much advanced compared to other patients making the texture very different from that of other patients. The contrast and colour information in images of patient 4 is

different from other patients since red correction[1] has not worked properly for the data of other patients. These differences would ultimately make the data belonging to these patients more challenging to classify as the colour information is not dependable.

In the case of the Meningiotheliamatous data shown in Figure 1.2, there is the issue of red correction for some samples but more importantly some Meningiotheliamatous images can be grouped with Fibroblastic as they contain collagen and the formation of a matrix can be seen. This can be clearly seen in the data for patients 3, 4 and 5 where texture seems to be tearing apart with presence of collagen and stretched cells with a matrix present as well (a feature that is more closely associated with the Fibroblastic subtype). Hence, Meningiotheliamatous gets confused with Fibroblastic very often making it one of the more complicated Meningioma subtypes to classify.

Transitional again has issues with red correction as shown in Figure 1.4. Texture for patients 1 and 2 seems to be more broken with whorls, but there are some regions in images for patient 3 where whorls are not found and there is also the presence of collagen making the texture more like Fibroblastic. Hence, low intra-class texture variation is an issue for Transitional Meningioma subtype classification as well. Moreover, there are regions in the images that resemble the Fibroblastic subtype making the task of classification more difficult.

Psammomatous data shown in Figure 1.5 again shows a notable degree of intra-class variation. There is the issue of red correction working slightly in certain samples such as those of patient 3 and patient 4 whereas not working too well for other patients. Some patients have Psammoma bodies of large size dispersed, such as patient 1 and patient 5, while some have smaller Psammoma bodies, such as patient 2, making the texture in the samples very different on a per patient basis. It can also be seen that the texture in the images belonging to the same patient may differ as well, for example as can be seen in the samples for patient 1 and patient 5. Some images have smaller psammoma bodies present while the

[1]The histological specimens are H&E stained which causes the non-nuclei segment to become pink and nuclei becomes purple. Red correction is a feature in some microscopes that fixes the redness in the images. It basically processes the images to remove redness from the images which is an artifact acquired during biopsy slide preparation and image acquisition.

rest have large psammoma blobs. Psammomatous data is, however, considerably different from the image data belonging to other Meningioma subtypes and is therefore, easier to classify correctly.

There is also the issue of inter-class variation in the image samples being low in many instances. This can be clearly observed for the Meningiotheliamatous samples as data for patient 3 resembles the Fibroblastic patient 2 as presence of collagen with formation of a matrix can be seen. It can also be observed that the formation of whorls can be seen for Fibroblastic patient 4 which is essentially a property of Transitional samples as can be seen for the Transitional patient 5. On the other hand, looking at the texture within the same class a great amount of variation in the texture can be seen. The high intra-class variation can be clearly observed for the all Meningioma subtypes. The images in the rows show a great variation in texture. The variation in the colour means that the colour information is not dependable.

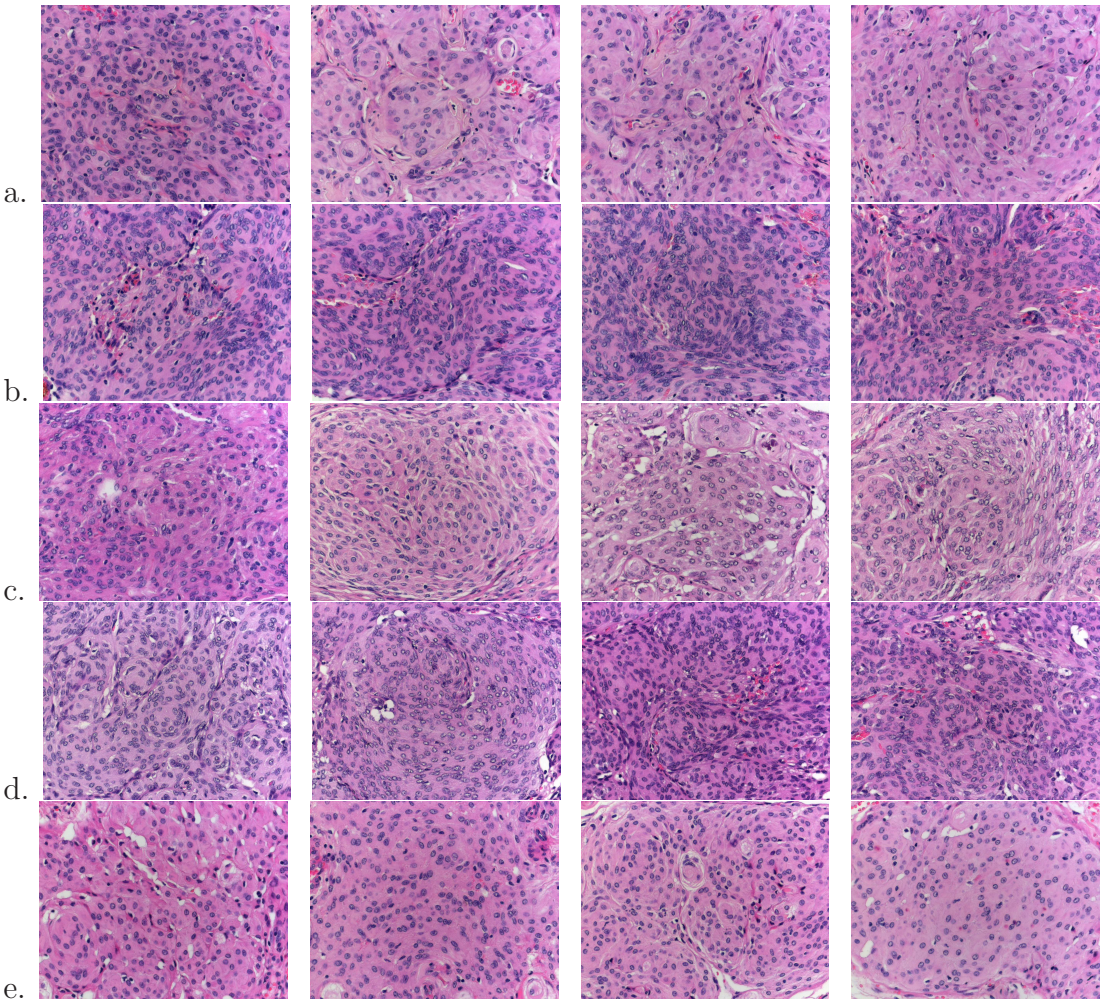


Figure 1.2: Meningiotheliamatous patient data a. Patient 1, b. Patient 2, c. Patient 3, d. Patient 4, e. Patient 5

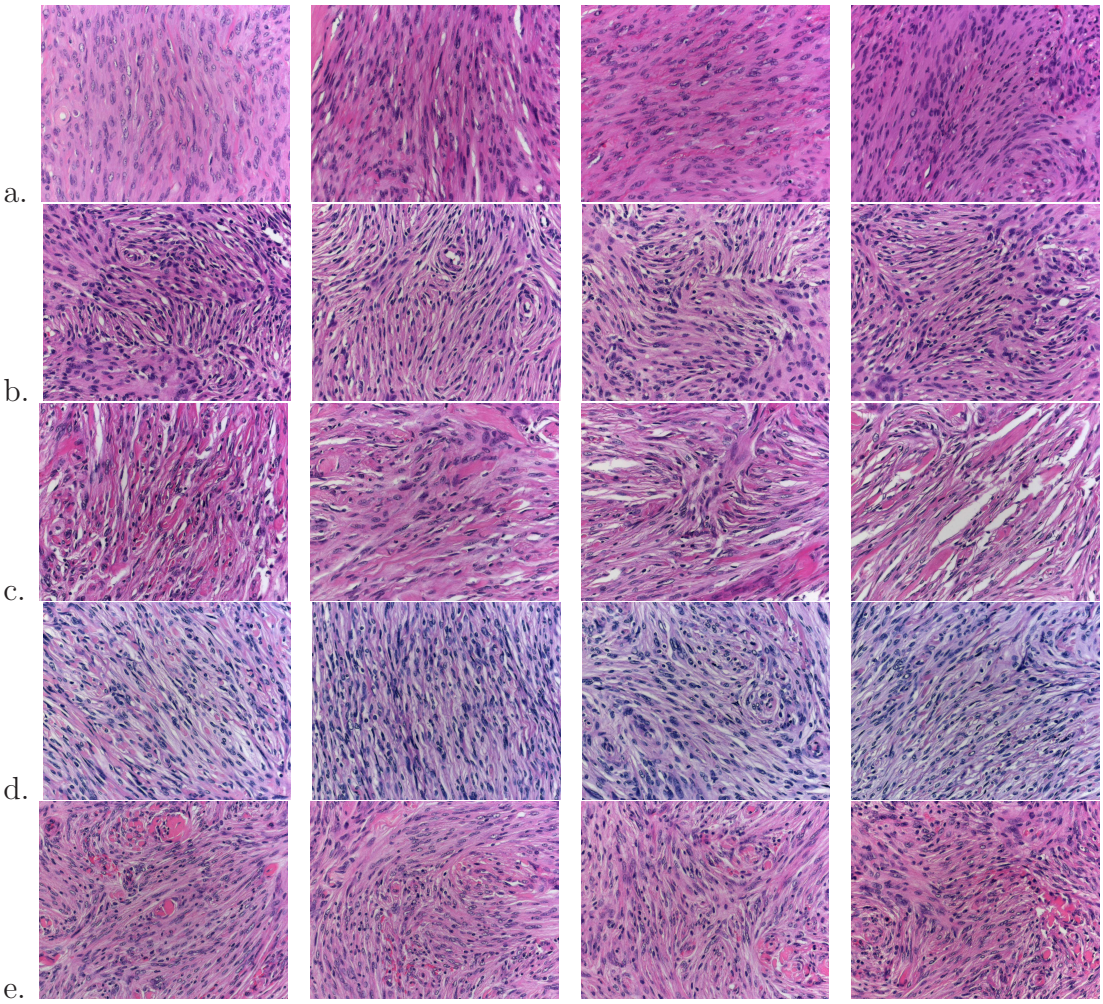


Figure 1.3: Fibroblastic patient data a. Patient 1, b. Patient 2, c. Patient 3, d. Patient 4, e. Patient 5

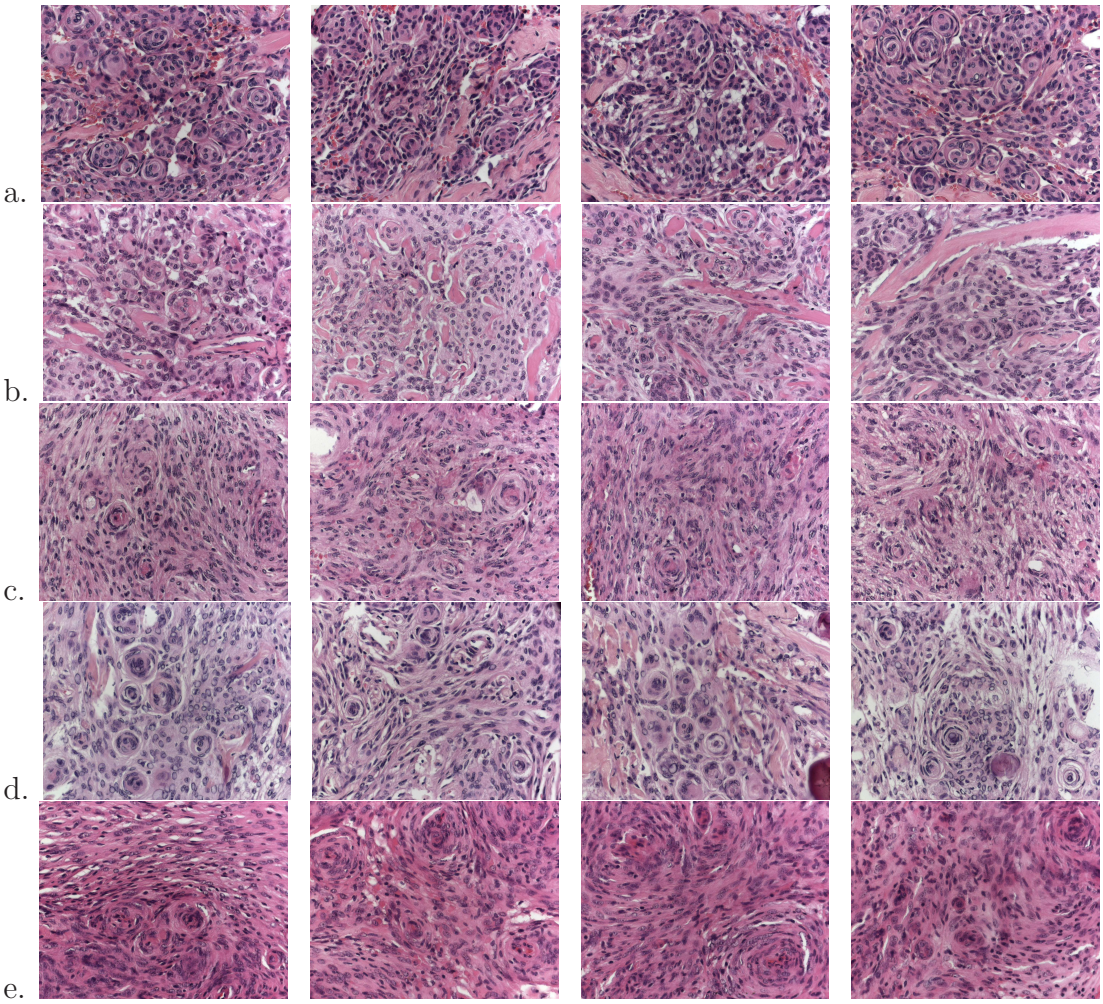


Figure 1.4: Transitional patient data a. Patient 1, b. Patient 2, c. Patient 3, d. Patient 4, e. Patient 5

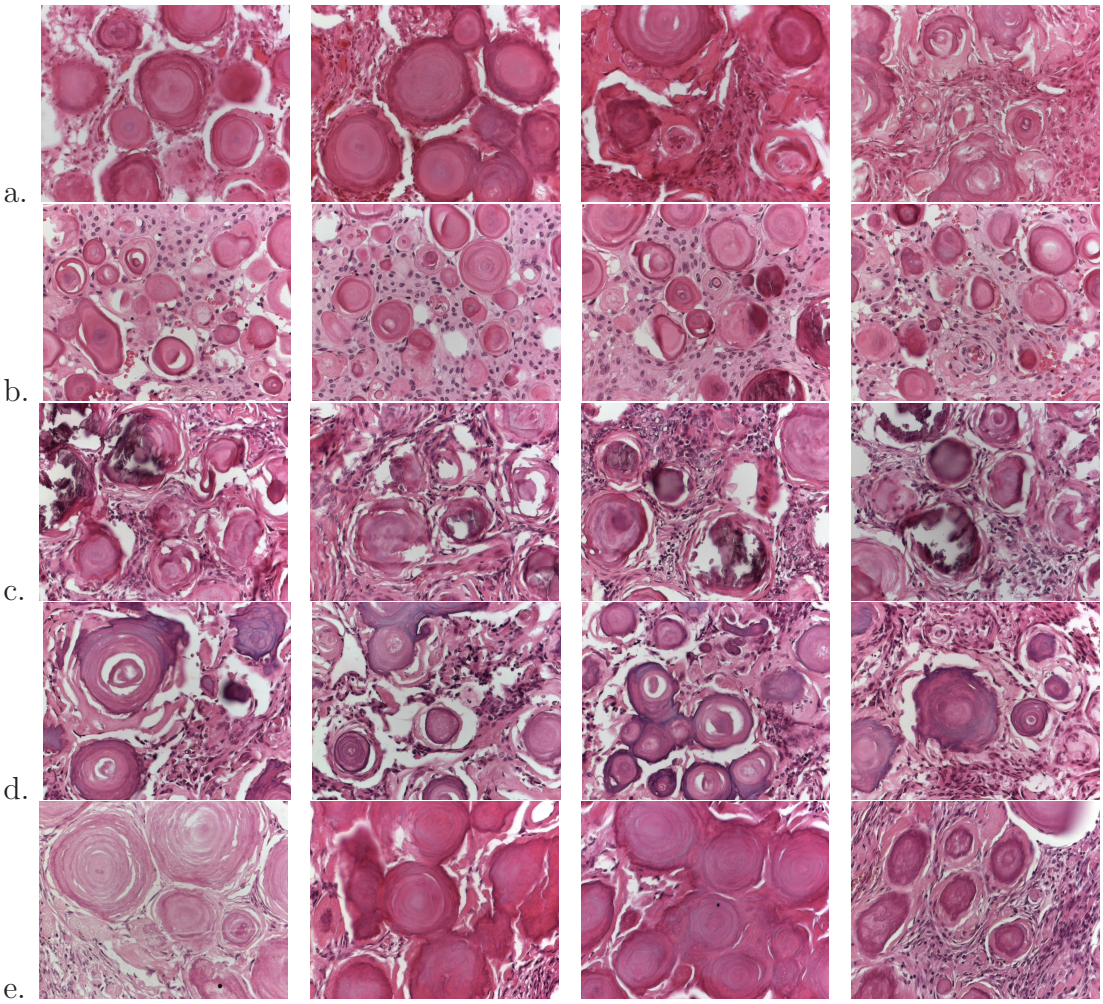


Figure 1.5: Psammomatous patient data a. Patient 1, b. Patient 2, c. Patient 3, d. Patient 4, e. Patient 5

1.5 Meningioma Subtype Classification: Segmentation Vs. Textural Analysis

Meningioma subtype classification is a relatively hard problem. There are two approaches that could have been used to solve the problem. Firstly, segmentation could have been used to extract structures in an image and then classification could have been carried out based upon the constituents of the image. The other approach that can be used is textural. Textural features can be acquired from each image and classification can be carried out based upon those features. Each approach has its pitfalls and strengths.

Segmentation does not suit our problem for many reasons. The first and foremost reason is the huge variation in the data. The tumour subtypes in our analysis exhibit high intra-class variability as can be seen in Figures 1.2-1.5. This variation entails the presence of structures that can be very different for images belonging to the same class. There is also the issue of the presence of structures representing different tumour subtypes in an image which otherwise may overwhelmingly represent one type of Meningiomas. Moreover, the type of texture present in the images of Fibroblastic, Meningiotheliamatous and Transitional subtypes makes the task of segmenting difficult. As can be seen from the Figures 1.2-1.5, its difficult to ensure that the structures obtained through segmentation would be able to accurately differentiate between textures belonging to these classes as there is low inter-class variation between textures in many instances.

Shape and structural analysis of constituents of images using segmentation, has found wide application in histology image analysis [94]. Numerous segmentation techniques such as thresholding [15, 21], adaptive thresholding [95, 96], watershed algorithms [40], fuzzy clustering and active contours [97] have been used for the segmentation of nuclei and glands in histology images but there are issues in using each of these techniques. Thresholding tends to work better for uniform images but produces inconsistent results if the variability within the image sets is high as stipulated by Gurcan *et al.* [6]. Watershed algorithms tend to suffer with the same problem [98]. Active contours deal better with the variability but multiple overlapping objects pose a difficult challenge for active contours as they may be enclosed to form one object [99]. Meningioma images suffer from

all the issues mentioned above. There is considerable variation within the texture samples and overlapping of structural components is frequent as can be seen for Meningiothelial and Psammomatous samples. Furthermore, high intra-class variation and low inter-class variation pose serious issues for any segmentation technique as the segmented constituents differ greatly for different image samples belonging to the same class. Hence, segmentation may not be the best approach for differentiating between the different subtypes of Meningiomas.

Textural analysis suffers from its inability to account for the structural characteristics of objects in an image. The strength of textural analysis lies in the fact that it examines the texture pattern as a whole and employs the differences amongst the patterns to classify images. There are strong distinct patterns that can be seen in the images of various Meningioma subtypes. Although there are variations in the patterns with subpatterns being present as well, the holistic approach of texture analysis may be useful in assigning a label to an image. Since, textural analysis of an image involves analysing an image as a whole rather than analysing its constituent parts, it should prove more useful (in our opinion) for classifying or differentiating between different Meningioma subtypes. Therefore, we have preferred using a textural analysis based approach which is novel and as we shall see in Chapters 4 and 5 produces high clustering and classification accuracies. Moreover, textural analysis has not been explored as much as segmentation for histological image analysis, hence there is a need to develop novel textural analysis algorithms for histological image classification.

1.6 Main Contributions

The main contributions of our work are:

1. A novel adaptive multiresolution wavelet transform based approach for Meningioma classification that allows for analysis of textures at various spatial and frequency resolutions.

2. A robust approach to overcome the challenge of high intra-class texture variation as seen in the case of Meningioma subtypes by statistically selecting subbands that have a high probability of occurrence.

3. Efficient and novel approach for feature selection that overcomes the problem of low variation in texture amongst the different classes often referred to as low inter-class variation by selecting the most discriminant subbands.

4. High classification accuracies are obtained.

5. The technique is suited to problems where more than two classes of textures need to be analysed and compared.

1.7 Thesis Organization

In this thesis, we present a novel textural analysis approach for Meningioma subtype classification. The next chapter provides a review of the various textural analysis techniques that use wavelet packets. The chapter also introduces our novel technique and some preliminary comparative results are presented. Chapter 3 describes the technique in detail with an analysis of the stability of the technique and analyses its robustness for Meningioma texture analysis and classification. In the subsequent chapter, clustering results are presented and various distance functions are analysed for the computation of the novel Adaptive Discriminant Wavelet Packet Transform (ADWPT). In Chapter 5, classification results are presented with a new approach for feature selection. The thesis ends with conclusions and future directions in Chapter 6.

Chapter 2

Towards a Multiresolution Analysis of Meningioma Subtype Textures

Multiresolution image representation has been one of the most important developments in the domain of image analysis [100–103]. The multiresolution approach was able to incorporate ideas from various disciplines which included subband coding from signal processing, quadrature mirror filtering from digital speech recognition and pyramid image processing. The strength of the technique lies in the representation of the signal at more than one resolution so that a feature which is undetected at one resolution is detected at another [30].

Multiresolution analysis acts like a virtual microscope that can be used to analyse complex signals and textures at various spatial resolutions. Moreover, the wavelet transform allows for signal analysis at various frequency resolutions as well. Hence, multiresolution analysis using wavelets can be used to analyse an image at various spatial-frequency resolutions so that some features may become more obvious at one resolution while other features can be analysed better at another resolution. The process is analogous to applying a microscope at different power levels for the analysis of histological slides.

In this chapter, we describe the development of transform based techniques as tools for multiresolution analysis of image textures such as Meningiomas. The pros and cons of each technique are discussed in Section 2.1. Wavelet transforms

and wavelet packets are described in Section 2.2 and Section 2.3 respectively. The chapter concludes with some comparative results on the discrimination ability of wavelet based methods and some other non-transform based methods.

2.1 Transform-based Analysis

In the 19th century, in his book "The Analytic Theory of Heat", Fourier described how an irregular and complex function can be represented as a sum of series of sines and cosines. These series were referred to as the Fourier series. The new theory proposed that all signals however complex, consist of simpler functions such as sines and cosines. Fourier series led to the development of the Fourier Transform. This was the beginning of a long tradition of transform based analysis. The development of the Fourier transform had a significant impact on the field of signal analysis [104–111], as it became possible to study complex phenomena such as heat dissipation and sound generation in a completely new way.

The Fourier transform of a continuous signal $f(t)$ can be computed as:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t}dt \tag{2.1}$$

where $e^{j\omega t} = cos\omega t + jsin\omega t$. The Fourier transform allows for the conversion of a signal from the time domain to the frequency domain. This means that individual frequencies in a signal can be analysed independently, hence breaking the problem of signal analysis into smaller more manageable parts using Fourier analysis. Although the Fourier transform is a powerful concept in the respect that it allows for frequency analysis of a complex signal, it is limited, as it assumes continuity of a signal in infinite time or space. This implies that the Fourier transform of a signal is possible only after assuming that the constituent frequencies would remain persistent throughout the duration of the signal and no new signals and frequencies would be added. A solution to the problem is taking the Short Time Fourier Transform or the Windowed Fourier Transform.

The Short Time or Windowed Fourier Transform, as the name implies, is the Fourier Transform of a signal at different points in time in a predefined time window [112–114]. This means that a signal is divided into constituent parts

at different points in time and then Fourier analysis of the signal in that time span is carried out. Gabor introduced the concept of a Gaussian function as a window in the synthesis phase. This resulted in what is now known as the Gabor transform [115]. The issue with STFT is that the time-frequency analysis is heavily dependent on the window chosen. The time-frequency resolution is fixed over the entire time-frequency plane, since the size of the window is fixed and the same window is used for the analysis of the whole signal. This introduces the 'time-frequency dilemma'.

The issue of time frequency resolution, as per the Heisenberg uncertainty principle [116], is that a signal cannot be localised precisely both in time and frequency at the same time. The problem occurs when the signal being analysed changes too quickly or too slowly i.e. for high frequencies, a shorter window is good for analysis but if the signal is low frequency, a wider window is more suitable. In the former case, the signal is well localised in time but may not be in frequency, as it represents a short segment of the signal, whereas in the latter case it would be well localised in frequency but not in time since the analysis window is too long.

This discrepancy in the spatial-frequency resolution would require examining a signal at various trade-off points. This implies that a signal may need to be analysed at various frequency and spatial resolutions with the analysing window contracted and dilated to represent various frequency and spatial resolutions. The need for a technique that would allow the analysis of a signal at various spatial and frequency resolutions led to the development of the wavelet transform. Unlike the Fourier transform, which severely limits the choice of the basis functions i.e. only sines and cosines can be used, wavelet transform offers possibly a very large number of choices in terms of basis functions. A basis function is an element of the function space that may be used to represent other more complex signals and functions as linear combinations of it. We shall see some examples of basis functions for wavelet transforms later in this chapter.

2.2 Wavelet Transforms

The wavelet transform is one of the most widely used tools in the area of signal and image processing. Its application area is diverse, ranging from image compression to texture analysis. The wavelet basis effectively represents the constituent frequencies that form the original signal at multiple frequency and time resolutions.

The ability that differentiates the wavelet transform from other techniques for signal analysis is that it allows the study of a signal both in time and frequency domain at the same time [117]. Recently, the wavelet transform (WT) has replaced the Fourier transform (FT) as the mainstay of transform-based image processing. The WT differs from the FT, as it is based upon small wave functions called wavelets, which differ from the sinusoid functions used in FT. Wavelets are functions of varying frequency and limited duration. The advantage the WT has over FT is that in the latter temporal information is hidden only to be available after inverse FT whereas WT not only retains the frequency information, but also the time or space information of the various components of the signal. This is achieved by decomposing a signal both in time and frequency domains at multiple resolutions.

Wavelet analysis is performed using the shifts and dilations of a prototype function $\psi(t)$ in the continuous domain. The function $\psi(t)$ satisfies

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{2.2}$$

and is also referred to as the mother wavelet. When the mother wavelet is dilated by a factor of a and translated by a scalar b, another wavelet denoted by $\psi_{ab}(t)$ is derived which is given by,

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \tag{2.3}$$

The wavelet transform $F_w(a,b)$ of a function $f(t)$ at a scale a and position b is computed by correlating function $f(t)$ with $\psi_{ab}(t)$

$$F_w(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} \psi(\frac{t-b}{a})f(t)dt \tag{2.4}$$

This transform is referred to as the Continuous Wavelet Transform (CWT). The CWT in theory is infinitely redundant but may be used to extract important characteristics of the signal being analysed as explained in [117]. The CWT involves the analysis of a signal at a very high number of frequencies using dilations of the mother wavelet leading to many possible bases of signal representation and hence, introduces redundancy and complexity. An effort to overcome the complexity of CWT implementation and reduce CWT redundancy, led to the development of the Discrete Wavelet Transform (DWT).

The main idea behind the Discrete Wavelet Transform (DWT) is basically the same as CWT. The time-scale representations are again obtained by translations and dilations of a wavelet function, the only difference being that in the case of DWT, the process is carried out using discrete values. The dilation is most often carried out by a power of 2 and is referred to as dyadic. DWT is essentially the sampled version of CWT.

The CWT is computed by changing the scale of the analysing function, shifting it in time, convolving it with the original signal and integrating over time. In DWT, filters with different cutoff frequencies are used to analyse a signal at different scales. An input signal is analysed with a series of low pass and high pass filters. The low pass filter is referred to as the scaling function while the high pass is called the wavelet function. Two dimensional (2D) signals such as images may also be analysed using DWT. Figure 2.1 shows the process of a 2D wavelet transform where $H(x)$ corresponds to the scaling function and $G(x)$ is the wavelet function and the results of a one level transform are shown as well. As can be seen from the Figure 2.2, the wavelet transform decomposes an image into scale-space representation referred to as subbands which represent different frequency and spatial information. This is a very important property of the wavelet transform as it decomposes complex textures into simpler components aiding in their analysis. Another important property of the wavelet transform is that it is invertible which means that the original signal can be recovered without loss by employing the inverse wavelet transform.

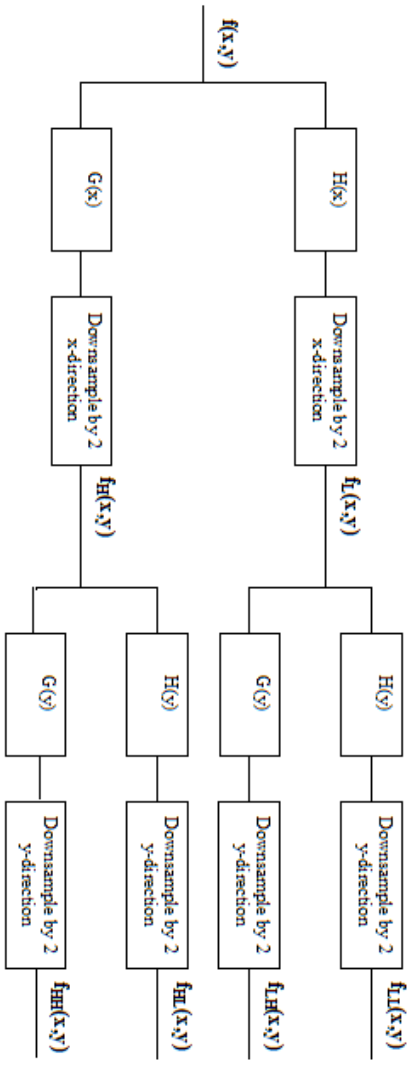


Figure 2.1: Process of obtaining a Wavelet transform of an image showing the filter bank arrangement to compute 1-Level DWT

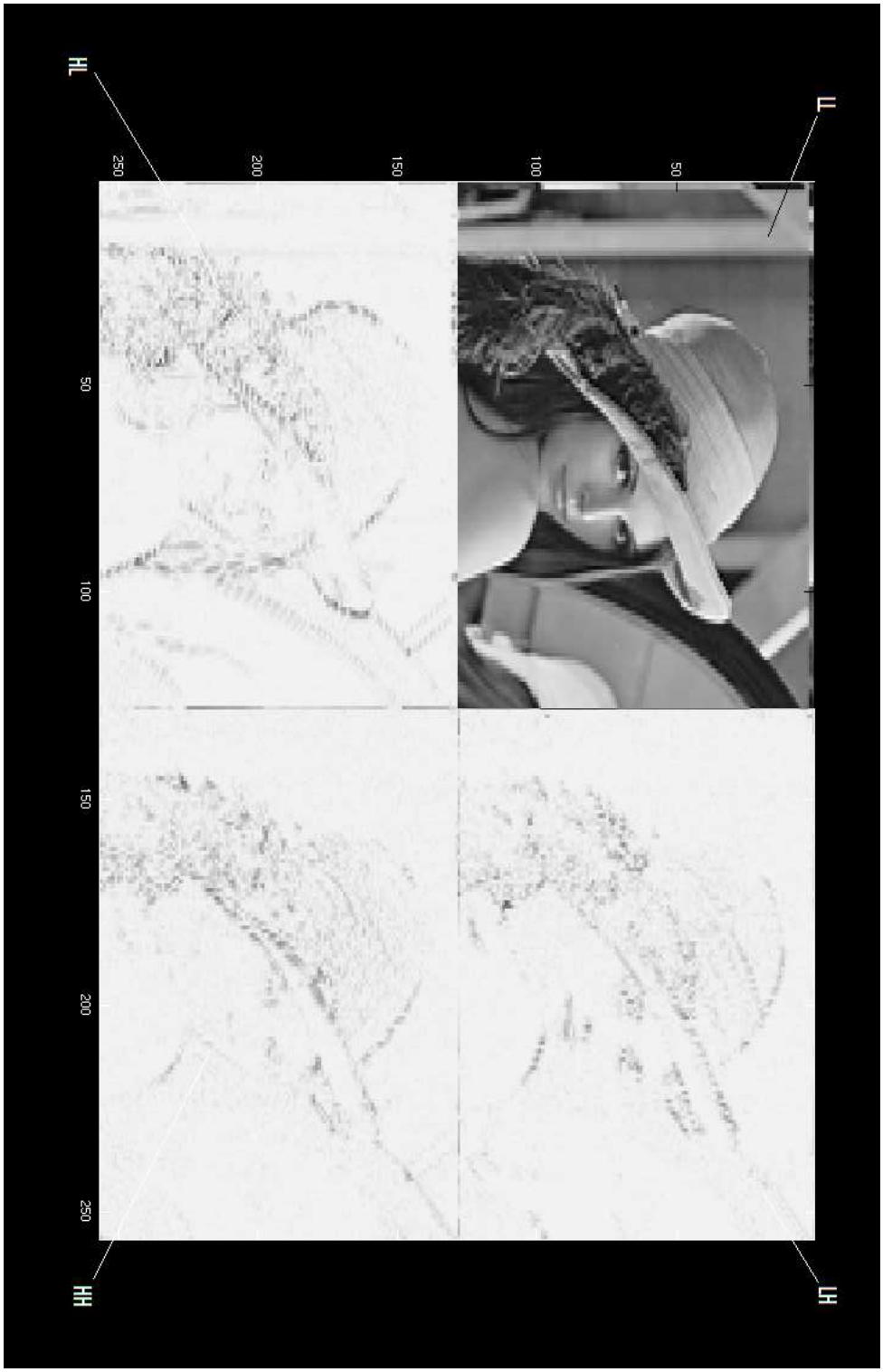
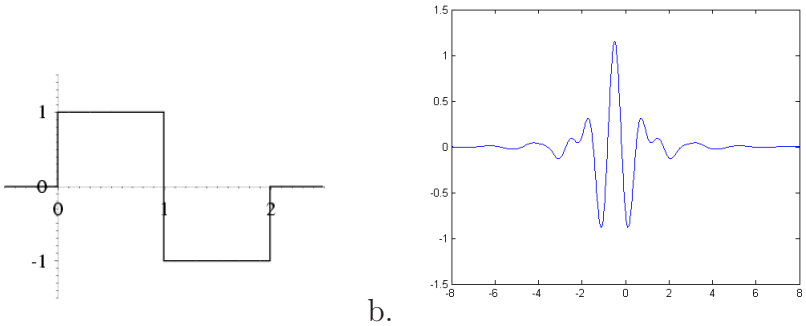


Figure 2.2: A wavelet transform of Lena showing 1st level Decomposed image



a. b.

Figure 2.3: Examples of Wavelets, a. Haar Wavelet, b. Meyer Wavelet (Source Matlab)

As an improvement to the Fourier analysis, the idea of wavelet analysis is to determine how an analysing function (wavelet) needs to be modified to approximate a signal. Various kinds of wavelet functions have been invented and used for signal analysis. Figure 2.3 shows some examples of a few well known wavelets. Wavelets have the capability to automatically adapt to different components of a signal by employing a small window to look at brief, high frequency components and a large window to look at long-lived and low frequency components. A wavelet function acts like a 'mathematical microscope' by compressing the wavelet to increase the magnification and dilating it to reduce it [117]. Hence, the signal can be analysed at various scales. The analysing function can be shifted over the signal for analysis at different times. Hence, its an ideal tool for constructing a virtual microscope for the analysis of meningioma subtypes.

Wavelet analysis may be carried out by a pair of functions $\phi(x) \in L^2(R)$ (scaling function) and $\psi(x) \in L^2(R)$ (wavelet function) which are equivalent to the functions $H(x)$ and $G(x)$ respectively shown in Figure 2.1. $L^2(R)$ denotes the vector space of measurable and square-integrable one dimensional functions [102]. The approximation of the signal $f(x)$ at the resolution 2^j, $A_{2^j} f(x)$, is characterised by

$$A_{2^j} f = (\langle f(u), \phi_{2^j}(u - 2^{-j} n) \rangle)_{n \in Z} \tag{2.5}$$

Once the approximation is obtained, the next stage is computation of the detail. Detail is defined as the difference of information between the approximations of a function $f(x)$ at resolutions 2^j and 2^{j+1}. $(2^{-j}\psi_{2^j}(x-2^{-j}n))_{(n,j)\in Z^2}$ is an orthonormal basis. Hence, $\psi(x)$ is called an orthogonal wavelet. The discrete detail signal at the resolution 2^j is given by,

$$D_{2^j}f = (\langle f(u), \psi_{2^j}(u-2^{-j}n)\rangle)_{n\in Z} \tag{2.6}$$

If the original signal has N samples, then the discrete signals $D_{2^j}f$ and $A_{2^j}f$ have 2^jN samples each. Hence the wavelet representation can be written as,

$$(A_{2^j}f, D_{2^j}f)_{-J\leq j\leq -1} \tag{2.7}$$

The wavelet representation would have the same total number of samples as the original signal, but divided amongst the approximations $A_{2^j}f$ and $D_{2^j}f$. This analogy can be extended to images and in this case the function space becomes $L^2(R^2)$, since we are dealing with 2D signals. The signal is now written as $f(x,y) \in L^2(R^2)$. The scaling function now becomes $\Phi(x,y)$, which is written as,

$$\Phi(x,y) = \phi(x)\phi(y) \tag{2.8}$$

The above equation emphasises the fact that the scaling is performed in both the horizontal and the vertical directions as seen in Figure **??**a. Consequently, the approximation of an image $f(x,y)$ at a resolution 2^j is characterised by,

$$A_{2^j}f = (\langle f(x,y), \Phi(x-2^{-j}m, y-2^{-j}n)\rangle)_{(m,n)\in Z^2} \tag{2.9}$$

Now, let $\psi(x)$ be the one dimensional wavelet associated with the scaling function $\phi(x)$. The three possible 2-D wavelet functions are then given by,

$$\Psi^1(x,y) = \phi(x)\psi(y), \Psi^2(x,y) = \psi(x)\phi(y), \Psi^3(x,y) = \psi(x)\psi(y) \tag{2.10}$$

It must be noted here that $\Psi^1(x, y)$ and $\Psi^2(x, y)$ perform wavelet filtering in one direction and scaling in another leading to horizontal and vertical details whereas $\Psi^3(x, y)$ leads to the diagonal detail with wavelet filtering in both directions. The three detail images are then given by,

$$D^1_{2^j} f = (\langle f(x, y), \Psi^1(x - 2^{-j}m, y - 2^{-j}n)\rangle)_{(m,n)\in Z^2} \tag{2.11}$$

$$D^2_{2^j} f = (\langle f(x, y), \Psi^2(x - 2^{-j}m, y - 2^{-j}n)\rangle)_{(m,n)\in Z^2} \tag{2.12}$$

$$D^3_{2^j} f = (\langle f(x, y), \Psi^3(x - 2^{-j}m, y - 2^{-j}n)\rangle)_{(m,n)\in Z^2} \tag{2.13}$$

In a classical wavelet transform, the filtering process defined is carried out for the approximation A_{2^j} iteratively to obtain further detail and approximations at various levels j. This process is also referred to as the discrete wavelet transform. The wavelet decompositions of a 2D signal can be viewed as nodes of a tree. The wavelet decompositions can be maintained in a quadtree structure, with the parent being the original subband or image and the children being the wavelet packet decompositions of the parent. The tree is referred to as the wavelet tree and the various nodes represent the different spatial-frequency information at various resolutions as shown in the Figure 2.4.

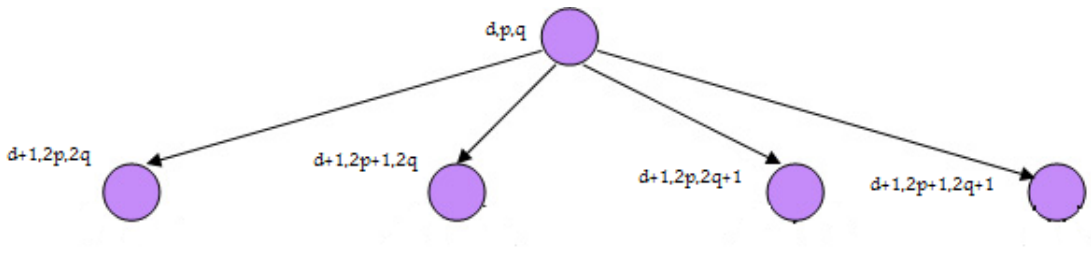


Figure 2.4: Wavelet tree showing the subband indices, depth (d) and frequency indices (p, q)

The nodes of the tree are expressed in terms of the indices d, p, q, where d is the depth of the tree (which is related to the spatial resolution) and p, q represent the frequency resolution indices. These nodes are referred to as subbands and can be obtained using the convolution operator as,

$$\mathcal{W}_{d+1,2p,2q} = \Phi * \mathcal{W}_{d,p,q} \tag{2.14}$$

$$\mathcal{W}_{d+1,2p+1,2q} = \Psi^1 * \mathcal{W}_{d,p,q} \tag{2.15}$$

$$\mathcal{W}_{d+1,2p,2q+1} = \Psi^2 * \mathcal{W}_{d,p,q} \tag{2.16}$$

$$\mathcal{W}_{d+1,2p+1,2q+1} = \Psi^3 * \mathcal{W}_{d,p,q} \tag{2.17}$$

where d represents the depth of the wavelet tree and p, q are the frequency indices of the subband. The subband $\mathcal{W}_{0,0,0} = I$ corresponds to the original image I being analysed. The three 2D wavelet functions are given by Ψ^i, where $i = 1, 2, 3$ and the scaling function is given by Φ. The above analogy will be used to represent the process of multiresolution analysis throughout this thesis.

The discrete wavelet transform (DWT) decomposes the input signal iteratively by filtering the approximation at various spatial and frequency resolutions. The detail subbands at each stage are not decomposed further. Hence, the subbands with frequency indices $p = 0$ and $q = 0$ are the only subbands that are decomposed at each level of the decomposition for DWT.

The detail subbands contain the high frequency information decomposed from the original image and the approximations. There is substantial information in detail components that can be further filtered to obtain a more elaborate representation of the original signal. The filtering of detail components forms the basis of the Wavelet Packet Transform.

2.3 Wavelet Packet Transform (WPT)

A generalisation of the discrete wavelet transform (DWT) led to the development of the wavelet packet transform (WPT). The difference between DWT and WPT is that in WPT both approximation and details are further filtered using the scaling and wavelet functions. Wavelet packet analysis of an image leads to a wavelet packet tree, the nodes of which may be indexed in terms of the depth d and subband indices p, q. Unlike DWT, in WPT subband indices with p, q values greater than 0 are also decomposed which are equivalent to the detail subbands.

Consequently, decomposing the approximation as well as the detail leads to an overcomplete wavelet packet basis representation, the terminal nodes of which

are referred to as the full wavelet packet transform. It must be noted that such a transform is overcomplete when all the nodes at all the various levels are included in the analysis.

2.3.1 Full Wavelet Packet Transform (FWPT)

In the Full Wavelet Packet Transform (FWPT), an image is decomposed into its respective subbands and then each subband is decomposed further until a predefined maximum depth is reached. Figure 2.5 shows the 2-level full wavelet packet tree. As Figure 2.5 shows, a wavelet packet transform tree is a poten-

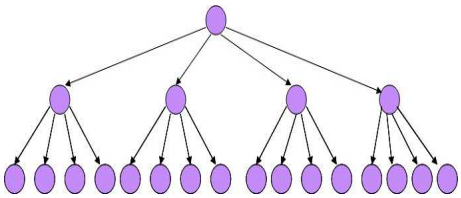


Figure 2.5: Full wavelet packet representation of an image

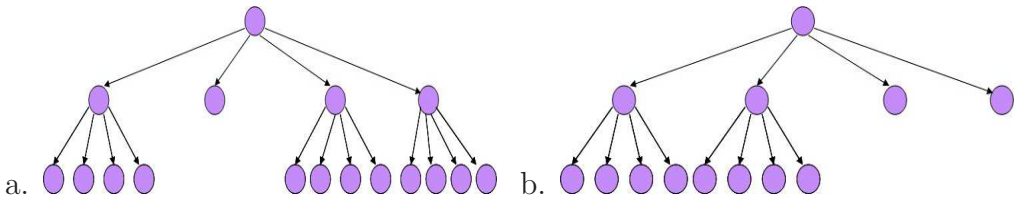


Figure 2.6: Different wavelet packet representations

tially redundant representation. This representation could be reduced to have different permutations of the terminal nodes but still retain all the information required to recover the original signal as can be seen in Figure 2.6. There are many possible plausible bases that could be acquired by keeping some nodes and truncating others as shown in Figure 2.6. The number of possible bases increase exponentially as the level of decomposition d increases and is given by,

$$N_0 = 1 \tag{2.18}$$

$$N_d = N_{d-1}^4 + 1 \tag{2.19}$$

Hence, the number of possible bases at a certain level is related to the number of bases at a previous level as deduced by Rajpoot in [118]. The important issue is what subbands to keep and which ones to get rid of. What spatial and frequency domain information is more useful for a given application? Hence, an algorithm for some sort of subband selection is very important for compactness and efficiency.

As stated earlier, subband selection remains an issue with there being a large number of possible wavelet bases for a given full wavelet packet tree. Coifman and Wickerhauser [119] combined dynamic programming with wavelet analysis to come up with a technique for obtaining the best wavelet packet basis. Thus the issue of redundancy in the overcomplete wavelet representation was resolved. In the next section, a novel technique is presented that performs subband selection by acquiring representations that help in differentiating between signals or textures under study. It can be used to obtain an optimal wavelet-based representation for meningioma subtype classification.

2.3.2 Adaptive Discriminant Wavelet Packet Transform (ADWPT)

Since the invention of best-basis selection algorithm by Coifman and Wickerhauser [119] many new algorithms have been proposed for the selection of a wavelet basis. Some of these will be discussed in the next section. Different optimal bases have been computed for various applications. In this thesis, we present a new algorithm for wavelet packet basis selection designed to obtain a wavelet packet basis that is optimal for differentiating between signals and textures being analysed.

Our novel algorithm, the Adaptive Discriminant Wavelet Packet Transform (ADWPT), seeks to obtain an optimal wavelet representation by optimising the discrimination power of the wavelet-bases. The ADWPT as a result obtains the optimal wavelet basis that best differentiates between signals under study. To define the ADWPT algorithmically, we define a discriminant function $\bar{\mathcal{D}}$ which may be represented as

$$\bar{\mathcal{D}}(\Upsilon, \mathcal{B}) \tag{2.20}$$

where Υ represents the training data-set which in this instance comprises of the four meningioma subtypes included in this study and $\mathcal{B}$ represents the wavelet packet bases. The objective of our computation would be to find a wavelet packet decomposition $\mathcal{B}^*$ that has the maximum discriminant function $\bar{\mathcal{D}}$ representing the overall discrimination power of a best basis $\mathcal{B}^*$.

$$\mathcal{B}^* = \underset{\mathcal{B}}{\operatorname{argmax}}\, \bar{\mathcal{D}}(\Upsilon, \mathcal{B}) \tag{2.21}$$

The above equation represents the computation function over a set of all possible wavelet packet trees to compute the wavelet packet representation that best differentiates between the different meningioma textures. The ADWPT thus obtained contains the subbands that are most discriminant. Hence, the aim is to obtain a set of subbands that collectively have the highest discrimination power. Various stages in the computation of the ADWPT are described in the next chapter. Next, we present an overview of the various wavelet packets based techniques used for texture analysis. The literature on meningioma classification and the use of wavelets in histological analysis was reviewed in Chapter 1.

2.4 Wavelet Packets for Texture Analysis

Over the last few decades, a large body of research work has been conducted in the area of texture analysis. Zhang and Tan [120] presented a review of various texture analysis methods. Various mechanisms and techniques have been employed ranging from statistical approaches to transforms and model based methods. Entirely new paradigms have been envisaged and efforts have been made to mimic the process of human perception in machines. Hence, the techniques employed for texture discrimination and classification are varied and bring together knowledge from various domains and fields of study. In this thesis, we are mainly concerned with techniques pertaining to the Wavelet Packet Transform (WPT).

In this section, the aim is to present a synopsis of the work done in texture analysis by employing wavelet packet transform and the various best bases selection techniques that have been used.

Many subband selection algorithms for wavelet packets have been presented. Some researchers perform subband selection whereas others have chosen to use all the subbands obtained in a wavelet packet decomposition. Many criteria have been used for subband selection, since the entropy based criterion proposed by Coifman and Wickerhauser [119]. Chang and Kuo [121] used energy as the criterion for selection of subbands by comparing subbands on the same scale and decomposing those subbands that have comparatively high energy. Pun and Lee [122] chose subbands that have high energy and achieve rotation and scale invariance using log polar wavelet signatures. Chen and Kundu [123] used wavelet packets and Hidden Markov Models (HMMs) for rotation and gray-scale invariant texture identification for Brodatz textures. The lowest level subbands in the wavelet packet tree were used to extract features. Laine and Fan [124] also performed Brodatz texture classification using wavelet packets but chose to select subbands arbitrarily by selecting all subbands from levels 1 to 3, only level 3, level 2 to 3 and only level 4, producing comparative results. Sengur *et al.* [125] combined a multi-layer perceptron classifier with wavelet packets for Brodatz texture classification. A 2-level wavelet packet decomposition was obtained for various textures in the Brodatz album and all subbands were used for the analysis.

The concept of wavelet frames was introduced by Unser [126] which was extended by Kim and Kang [127] to wavelet packets. Kim and Kang [127] used wavelet packet frames for Brodatz texture classification and segmentation. The frames with the greatest variance were selected producing high accuracies. Rajpoot [128] employed the local discriminant wavelet packet bases for subband selection which involved determining the discrimination power of a subband using various distance functions, namely Kullback-Leibler divergence, Jensen Shannon, Euclidean and Hellinger distance, for subband selection. The principle is that the higher the discrimination power the better the subband. A comparison is made between the parent and child nodes in the wavelet packet tree. Bhalerao and Rajpoot later extended the work in [129] to show that the discrimination power is an important measure that can indicate the usefulness of a subband in

differentiating between various textures. They add Bhattacharya distance to the set of distance functions and show that it performs better in most instances.

Some recent work on wavelet packet subband selection has been carried out by Huang and Aviyente [130, 131]. They propose a new mechanism for selection of subbands based upon mutual information amongst the various subbands. Two algorithms are presented, namely Mutual Information based Subband Selection (MISS) and Subband Grouping and Selection (SGS). Their algorithms are shown to perform better than the traditional subband selection paradigms described above. The techniques aim to compare all subbands at the same time based upon their energy values. The aim is to obtain a sparse representation whose coefficients are as independent as possible. This technique has been shown to produce promising results in comparison with other techniques for subband selection but their method compares low-pass subbands directly with high-pass subbands which may be flawed, since the frequency information they represent is very different. Moreover, the technique may result in selection of the parent subband with the children which would not have information exactly independent of the parents, a criterion which the technique aims to maximize i.e. mutual independence.

Some comparisons between wavelet packets and other techniques have also been carried out. An elaborate comparison of various filters namely Law masks, ring/wedge filters, Gabor filters, wavelet transforms, wavelet packets and wavelet frames, quadrature mirror filters, discrete cosine transform, linear predictors and finite impulse response filters was performed by Randen and Husoy [132]. The textural data used for the analysis was obtained from the Brodatz album, the MIT Vision Texture database and the MeasTex Image Texture Database. The results showed that no one filter performs better for all data sets and the study was important in the sense that the test and training data is completely separated with no overlap. This is very much the case in our study for meningioma subtype classification.

Arivazhagan *et al.* [133] uses Gabor wavelets for rotation invariant texture classification of Brodatz textures and compares the results with those of discrete wavelet transforms, indicating that Gabor wavelets perform better. Livens *et al.* [134] employed wavelet transform and wavelet packet transforms for corrosion

image classification. Wavelet packets perform better with no subband selection. Brodatz textures have been used in all the studies mentioned in this section.

As the literature review indicates, wavelet packets have been used quite widely for texture analysis but they have not been employed as widely in medical texture analysis, specifically in histology. Although some work has been carried out for wavelet packets [70] they are not as popular as other techniques such as Gabor or the Discrete Wavelet Transform. The reason for this remains the complexity of wavelet packets due to the exponential rise in number of subbands as the depth of the transform increases. There is also the issue of subband selection. Many studies indicate that the wavelet packet transform is better for texture classification in comparison to the wavelet transform, provided a viable mechanism for subband selection is used. Gabor transform and Gabor wavelets produce comparable results to wavelet packets as indicated in some studies [132].

In the next section, we provide a comparison between three transform based techniques, namely the wavelet transform, the full wavelet packet transform, the Gabor transform and the two spatial domain texture analysis techniques namely Local Binary Patterns (LBPs) and Gray Level Co-occurrence Matrix (GLCMs) for meningioma subtype clustering. We finally compare all these results with the Adaptive Discriminant Wavelet Packet Transform (ADWPT) and show that ADWPT performs better.

2.5 Meningioma Clustering: A Preliminary Study

Many transform and non-transform based techniques have been used in texture analysis for feature extraction. In this section, we compare the results for meningioma clustering obtained using some non-transform techniques namely GLCMs and LBPs and some transform based techniques such as Gabor and types of wavelet transforms with ADWPT. In the results shown all images in our data-set were used.

2.5.1 Non-transform based Textural Analysis Techniques

Non-transform techniques have found wide application in histological image classification as indicated in Chapter 1. We discuss two and present their results.

2.5.1.1 Gray Level Cooccurrence Matrix (GLCM)

Since their invention [135], GLCMs have been used widely for texture analysis. A comparison between ADWPT and GLCM was presented in [52]. Figure 2.7 shows the projections on the first three principal components obtained for the GLCM Energy feature using Principal Component Analysis (PCA) [136]. In this GLCM analysis, 4 directions namely 0^o, 45^o, 90^o and 135^o were used with 5 distances i.e. 1 to 5. 20 GLCM matrices are obtained and the GLCM-based energy feature is used to construct the feature set.

The four different colours represent the four meningioma subtypes. Blue corresponds to Fibroblastic, red corresponds to Meningiotheliamatous, green corresponds to Psammomatous and yellow represents the Transitional. Figure 2.7 clearly shows that there is no clear separation between any of the meningioma subtypes with Meningiotheliamatous, Psammomatous and Transitional forming a cluster at the centre while Fibroblastic is dispersed all around the centre with some samples in the middle.

2.5.1.2 Local Binary Pattern (LBP)

Local binary patterns (LBP) have also been widely used for texture analysis. Some comparative results between LBP and ADWPT were presented in [137]. Figure 2.8 shows the projections on the first three principal components obtained for the LBP features using PCA. A radius of 1 was chosen and 8 neighbourhood pixels are used in the analysis. Other radii and number of pixels were also used with no apparent improvement in results. A more detailed description of the technique can be found in the paper by Ojala *et. al* [138].

The results show that Meningiotheliamatous forms a relatively good cluster at the centre but is intermixed with Transitional samples. Fibroblastic samples are again dispersed while Psammomatous is projected vertically. The technique seems to be forming clusters that look better than the results for GLCM.

2.5.2 Transforms based Textural Analysis

An advantage of transform based multiresolution analysis is that apart from allowing analysis of a signal at different scales, it also breaks down the input signal into various constituent frequencies. This implies that at each scale of decomposition, each constituent band represents frequency content which may be classified as low frequency (approximation) or high frequency (detail). Hence, if η represents the frequency content in 2D signal $f(x, y)$ with the highest frequency represented by η_+ and the lowest frequency by η_-, and if a represents the frequency content of approximation, ω^1 the horizontal detail, ω^2 the vertical and ω^3 the diagonal detail, then the following holds,

$$a_- < a < a_+ \tag{2.22}$$

$$\omega_-^1 < \omega^1 < \omega_+^1 \tag{2.23}$$

$$\omega_-^2 < \omega^2 < \omega_+^2 \tag{2.24}$$

$$\omega_-^3 < \omega^3 < \omega_+^3 \tag{2.25}$$

$$\eta_- < \omega_-^1 \leq \omega_-^2 < a_- \tag{2.26}$$

$$a_+ < \omega_+^1 \leq \omega_+^2 < \eta_+ \tag{2.27}$$

This frequency resolution will help us in acquiring features. Since the frequency content has been resolved into low frequency and high frequency aspects with high frequency further divided into three levels, therefore, the process of extracting features is aided. This frequency resolution in essence distributes the inherent features of the image into various subbands representing periodic texture, horizontal edges, vertical edges and diagonal edges at various scales. This in turn helps in obtaining a feature-set that would be a good representation of the textures under study.

2.5.2.1 Gabor Analysis

Gabor analysis of the textures was carried out as proposed by Ma and Manjunath [139]. Four scales and six orientations were used to provide texture representations at various scales and orientations. Energy feature is used to construct the feature set so that a comparative analysis may be made between the various techniques discussed in this section. The mean and variance as suggested by Ma and Manjunath was also used but no improvement was seen in the results.

Figure 2.9 shows the results obtained for Gabor Wavelet Analysis. No separation is seen amongst the Fibroblastic, Meningiotheliamatous and Transitional subtypes. Psammomatous is much better differentiated with some points intermixed with the cluster at the centre.

2.5.2.2 Discrete Wavelet Transform (DWT)

A four level discrete wavelet transform was obtained as discussed previously in this chapter. Energy of each subband was used as a feature. Figure 2.10 shows the results for the discrete wavelet transform. The projection results on the first three principal components are comparable with those of Gabor Wavelets. Psammomatous is again much better separated while the other three meningioma subtypes are all clustered in the middle.

2.5.2.3 Full Wavelet Packet Transform (FWPT)

The Full Wavelet Packet Transform (FWPT) has been discussed in detail in previous sections. A four level FWPT is obtained and the energy feature is acquired from each subband. Figure 2.11 shows that the results do not improve much. Psammomatous is still the best differentiated with Transitional samples clustered southwards while Fibroblastic and Meningiotheliamatous inter-mixed.

2.5.2.4 Adaptive Discriminant Wavelet Packet Transform (ADWPT)

The Adaptive Discriminant Wavelet Packet Transform (ADWPT) was also carried out to the fourth level. The subband selection for the most discriminant

decomposition was obtained using the Fisher Discriminant. A detailed discussion on ADWPT will be presented in the next chapter.

Figure 2.12 shows the PCA analysis results (projections on the first three principal components) obtained using the energy feature from each subband. The results are quite encouraging. ADWPT forms the most defined cluster for Psammomatous differentiating it very well from all the other subtypes. Transitional is also seen to form clusters around the edge of the cluster intermixed with Fibroblastic and Meningiotheliamatous samples. There is no clear distinction between the Fibroblastic and Meningiotheliamatous samples but the results produced are the best so far as Psammomatous is very clearly separated while Transitional has also formed a distinct cluster.

The above results have compared some non-transform and transform based textural analysis techniques. The results show that ADWPT performs better amongst the transform based techniques while LBP does better in comparison to GLCM in non-transform based techniques. However, the preliminary PCA results indicate that ADWPT results are encouraging as it overall performs much better than LBP in differentiating between the meningioma subtypes. This is due to the fact that the multiresolution analysis using ADWPT allows for both spatial and frequency resolution whereas LBPs strength lies only in spatial analysis. The comparative results here show that ADWPT is clearly a good technique for meningioma subtype classification.

2.6 Summary and Conclusion

This chapter highlighted the importance of multiresolution analysis for texture analysis. We show that transform based analysis has numerous advantages over non-transform based techniques. Wavelet packet based methods are shown to produce better clustering results in comparison to other multiresolution techniques.

We present some comparative results for meningioma clustering with some transform and non-transform based techniques. The ADWPT provides the best results for clustering meningioma subtypes. In the next chapter, we develop the ADWPT further with a detailed discussion on the choice of the discriminant

function and the construction of the multiresolution texture approximation and how they impact the ADWPT best basis (subband selection).

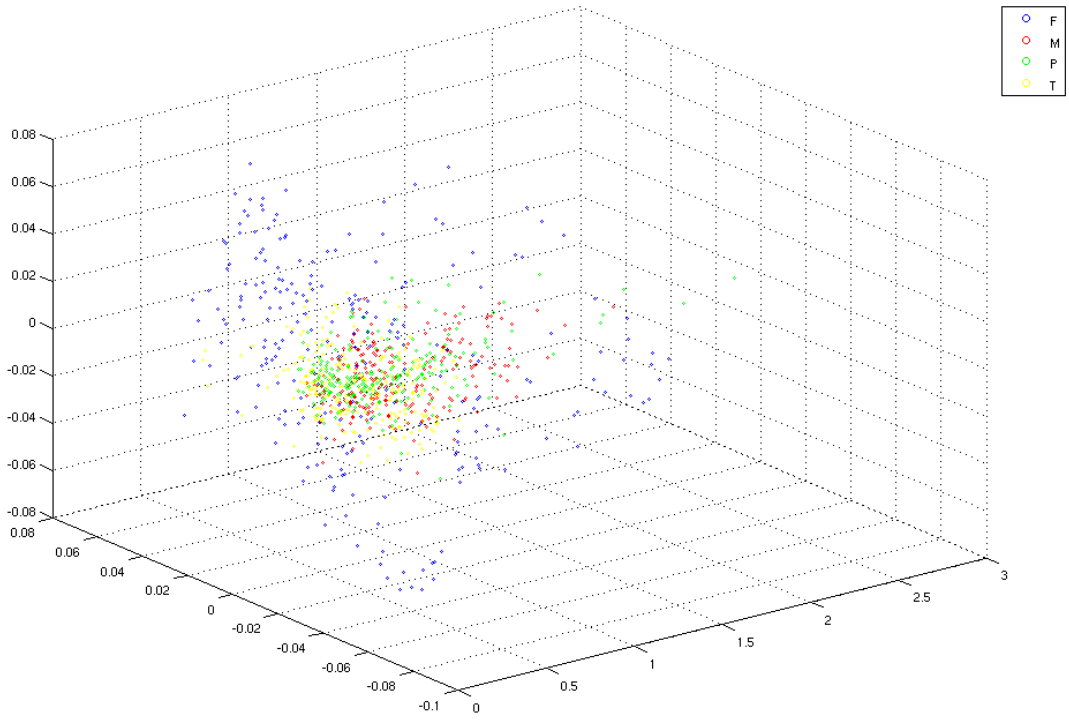


Figure 2.7: Projections on the first 3 principal components obtained using the PCA analysis of the GLCM based Energy feature-set (Fibroblastic (F), Meningiotheliamatous (M), Psammomatous (P) and Transitional (T))

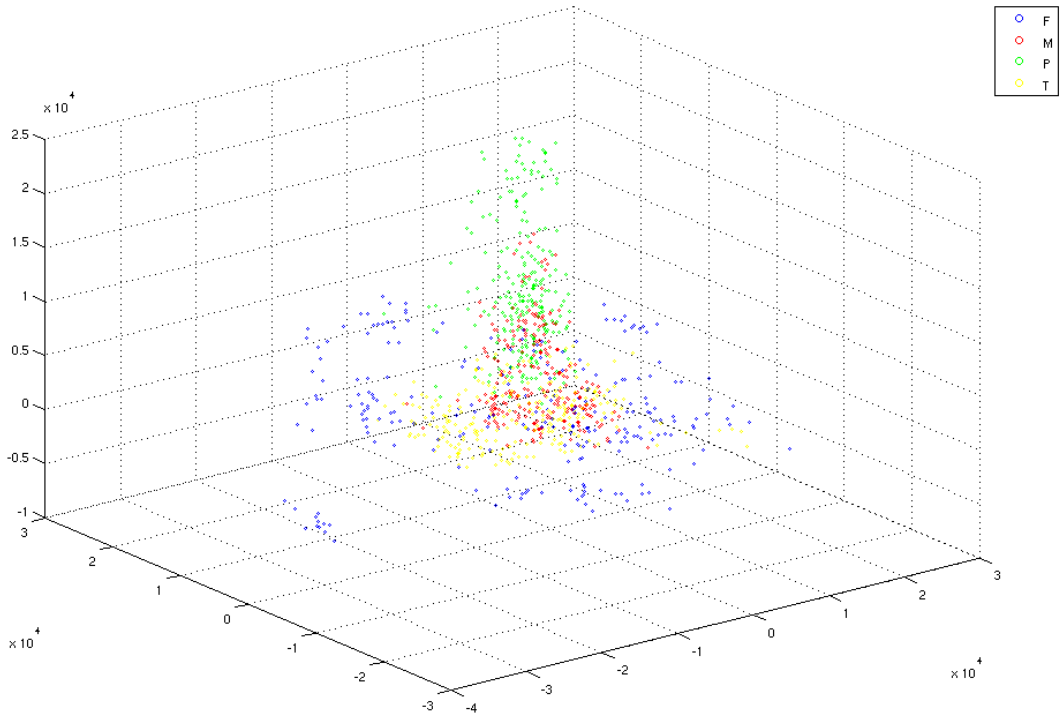


Figure 2.8: Projections on the first 3 principal components obtained using the PCA analysis of the Local Binary Patterns based feature-set (Fibroblastic (F), Meningiotheliamatous (M), Psammomatous (P) and Transitional (T))

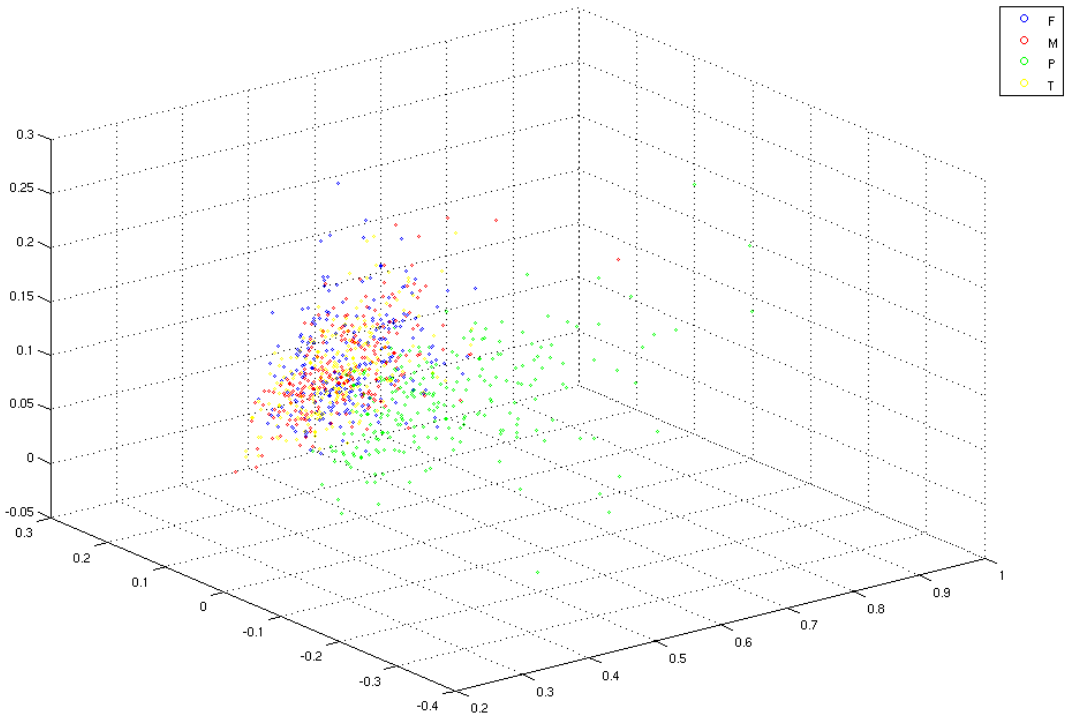


Figure 2.9: Projections on the first 3 principal components obtained using the PCA analysis of the Gabor Transform based Energy feature-set (Fibroblastic (F), Meningiotheliamatous (M), Psammomatous (P) and Transitional (T))

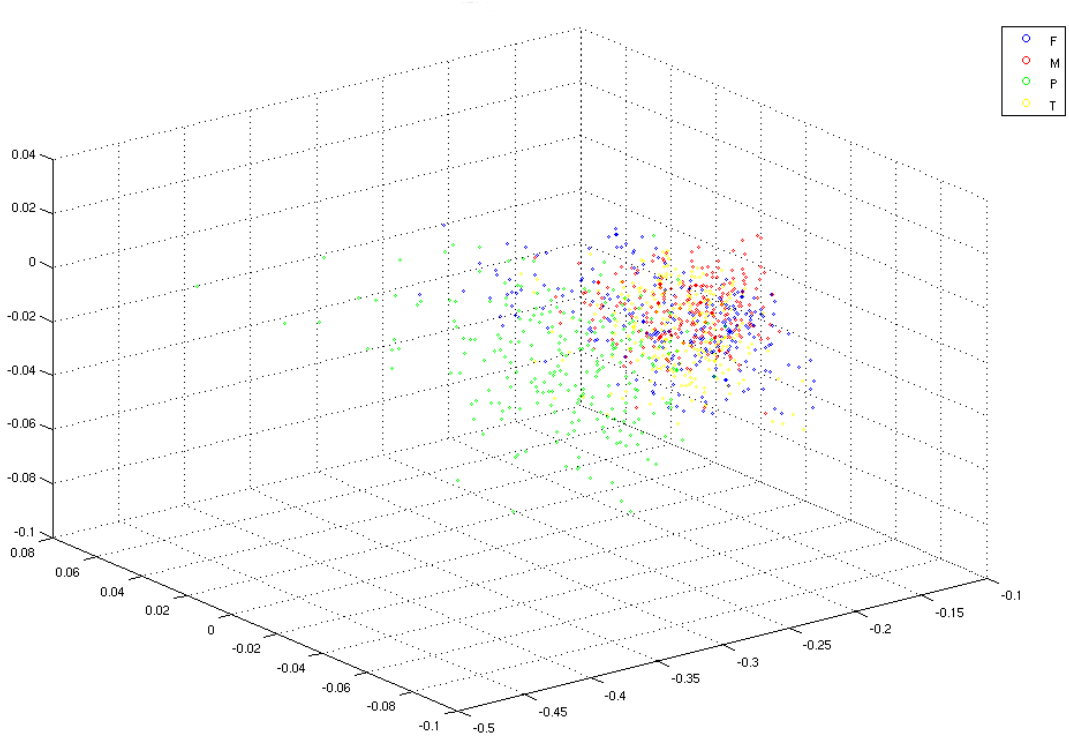


Figure 2.10: Projections on the first 3 principal components obtained using the PCA analysis of the Discrete Wavelet Transform based Energy feature-set (Fibroblastic (F), Meningiotheliamatous (M), Psammomatous (P) and Transitional (T))

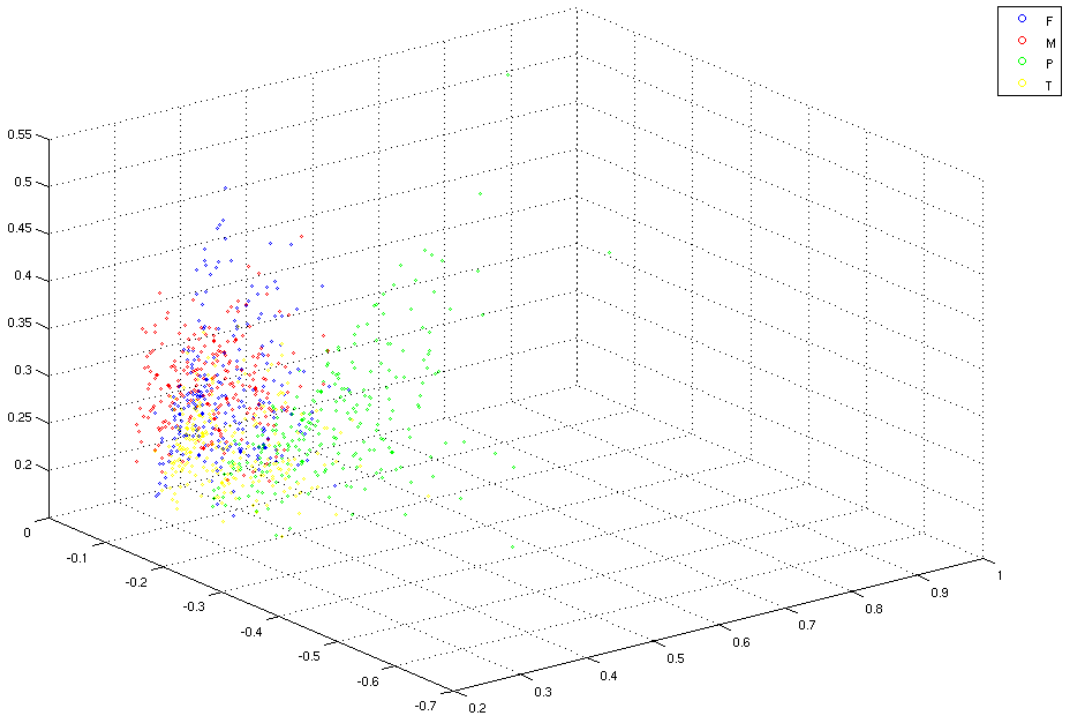


Figure 2.11: Projections on the first 3 principal components obtained using the PCA analysis of the Full Wavelet Packet Transform based Energy feature-set (Fibroblastic (F), Meningiotheliamatous (M), Psammomatous (P) and Transitional (T))

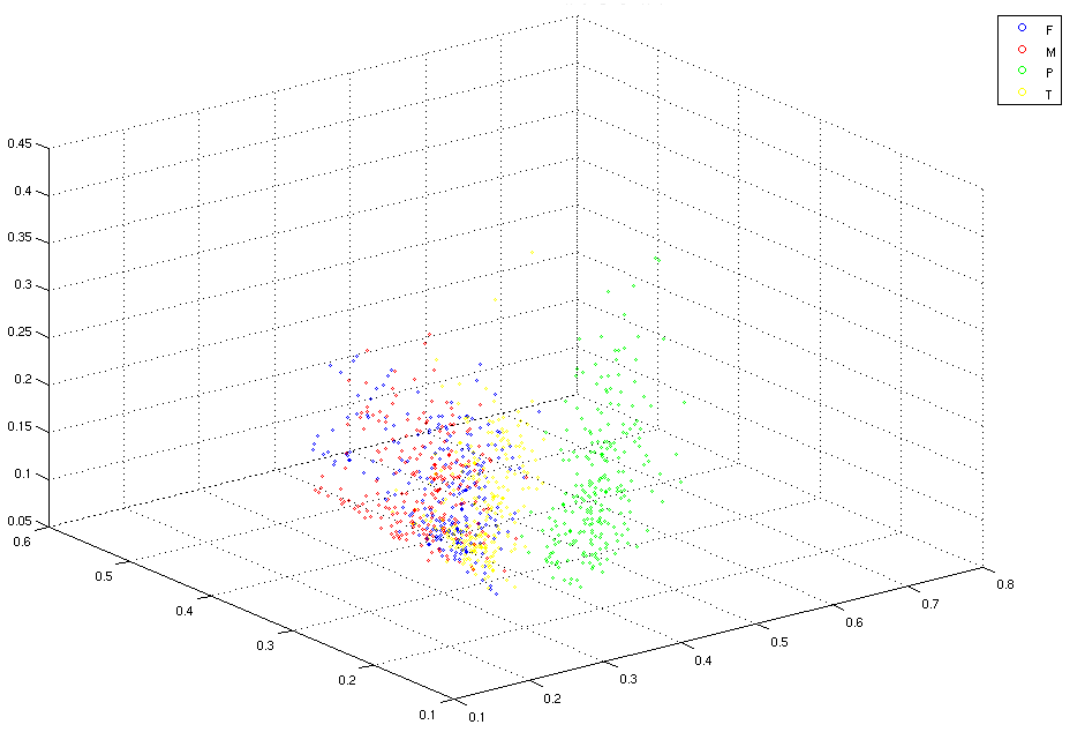


Figure 2.12: Projection on the first 3 principal components obtained using the PCA analysis of the Adaptive Discriminant Wavelet Packet Transform (Fisher Distance) based Energy feature-set (Fibroblastic (F), Meningiotheliamatous (M), Psammomatous (P) and Transitional (T))

Chapter 3

ADWPT and Its Stability

In this chapter, a detailed description of the various stages of the Adaptive Discriminant Wavelet Packet Transform (ADWPT) is presented. The most important aspect of ADWPT is the multiresolution comparison of textures. Texture class comparison involves two main steps namely texture template construction and computation of subband discrimination power which are described in detail in this chapter. An introduction to the various distance functions that have been used in the study for computing the discrimination power of a subband is also presented. The chapter discusses the issue of variability in ADWPT decomposition due to the variation in Meningioma textures used. An analysis on what type of texture template construction methods and distance functions are better for ADWPT computation is also presented.

As the name suggests ADWPT is an adaptive approach. This implies that it adapts to the problem semantics and incorporates problem domain information in order to resolve it. This enables ADWPT to be a viable approach for problems such as Meningioma subtype classification which have high texture variability. ADWPT is able to acquire textural features from various scale-space representations for variable and complex textural data by adapting to the type of texture data under study. However, this introduces the issue of ADWPT decomposition variability due to the variation in the data set. This would be referred to as ADWPT stability. Stability of ADWPT decomposition is important for extraction of a consistent set of features for a texture classification problem such as Meningioma subtype classification. If there is no stability then no consistent set

of features may be acquired. Section 3.1 describes the process of ADWPT computation with a detailed discussion on template construction and computation of subband discrimination power while section 3.2 describes the issue of stability of an ADWPT representation in detail. The various stages of ADWPT are described next.

3.1 Computation of ADWPT

The various stages of ADWPT are presented below.

1. Pre-processing (color to greyscale conversion) of each image.
2. Full Wavelet Packet Transform (FWPT) of each image.
3. Texture template construction for each texture class using the image FWPT trees computed earlier.
4. Computation of the discrimination power of each subband.
5. Selection of discriminant subbands so that ADWPT decomposition is obtained.

3.1.1 Preprocessing

Before the main stages of ADWPT, some preprocessing is applied to the images. First and foremost, the images are converted to gray scale. The following expression is used for conversion.

$$I_g = 0.299I_R + 0.587I_G + 0.114I_B \tag{3.1}$$

where I_g represents the image I in gray scale whereas I_R represents the red channel, I_G is the green channel and I_B is the blue channel. This allows the reduction of image data by incorporating info from all channels in to a gray scale image representation. Subsequently, each image is zero-meaned which essentially involves subtracting the image mean from all the pixels and is given by,

$$\bar{I}_g(m,n) = I_g(m,n) - I_g^{\mu} \tag{3.2}$$

where $\bar{I}_g(m,n)$ represents the zero-meaned grey scale image while $m = 0$ to $M_I - 1$ and $n = 0$ to $N_I - 1$. $M_I \times N_I$ is the resolution of the image. I_g^{μ} is the mean of the pixel values in the image. This is essentially the zero-meaning the data so that the image brightness does not effect the computation of the wavelet transform. Next, the FWPT of the grey scale image is obtained.

3.1.2 Full Wavelet Packet Transform (FWPT)

Daubechies 8-tap filter is used to compute the FWPT of up to 4 levels. An increase in the number of transform levels causes the FWPT computation complexity to rise exponentially. For analysing Meningioma subtype images of size 512×512 each, 4 level transform was found to be adequate. A total of 340 subbands are found in the FWPT decomposition tree of up to 4 levels. A 5 level FWPT causes the number of subbands to rise to 1364 which is a substantial increase in the number of features. This causes the complexity to increase many folds and does not provide substantial benefit. A 3 level transform was not used because the resolution of the terminal subbands was quite high at 64×64. A 4 level transform leads to subbands 32×32 in size and keeps the number of subbands at 340 which offers a good compromise as to the complexity of subband selection (it will be described later) and the 32×32 size of the subband allows for effective comparison of textures. Therefore, all analysis in this thesis is carried out with a transform of up to 4 levels.

Meningioma subtype classification with several wavelet filters was carried out in [69] and the various wavelet properties useful for textural analysis were investigated. Daubechies 8-tap filter was found to be one of the best wavelet filters for the analysis of Meningioma images. Hence, Daubechies 8-tap filter was chosen for all analysis in this thesis. The next stage is the computation of the multiresolution texture class representations so that the texture classes can be compared.

3.1.3 Texture Class Representation

An important step in ADWPT is the computation of representative texture templates for each textural class (Meningioma subtype texture). Hence, the aim is to come up with a standard texture template $\mathcal{T}^a$ representing each texture type a.

$$\mathcal{T}^a = \mathcal{F}\{a_i\} \tag{3.3}$$

where a_i represents textures belonging to each class and $i = 1...N_s$ while N_s is the number of texture samples. The simplest mechanism for obtaining a template is to choose the image that is most representative of each texture type and use the FWPT of that image for comparing the textures. In our first publication [140], a sample image that best represented the texture associated with a Meningioma subtype was chosen as a template for ADWPT computation. However, this was soon found to be inadequate for our problem domain because of the inherent complexity and intra-class variability in the texture classes being compared as discussed in chapter 1. Therefore, a novel solution to the problem was found which involved the computation of multiresolution texture approximations as described next.

Multiresolution Adaptive Wavelet Texture Class Templates (MAWTT)

The ADWPT algorithm aims to compare the various Meningioma subtype textures so that the most important and useful features that can be used to differentiate between textures are extracted. Texture images have components that can adequately represent textures referred to as textons in the literature [141]. Textons are defined as atomic features of a texture. Multiresolution analysis allows for the decomposition of the texture so that elements such as textons may be identified at various frequency and spatial resolutions. In a multiresolution wavelet-based decomposition, the texture is decomposed into simpler components in the shape of sparse subbands each representing the texture at a different spatial and frequency resolution. These subbands represent texton-like information as they capture intrinsic textural properties in various spatial frequency domains and hence can be used to compare textures in a meaningful way.

Multiresolution Adaptive Wavelet Texture Class Templates (MAWTT) as the name suggests are multiresolution wavelet based representations that are constructed from the available image data and hence are referred to as adaptive. The MAWTTs represent textural characteristics at various spatial frequency resolutions represented by the subbands in the MAWTT tree. The computation of MAWTTs has many stages with the first stage being the computation of the probability functions (pfs) of all the subbands in the FWPT decomposition.

Probability Function of Subbands

A probability function (pf) for each subband is computed by obtaining the normalised energy of the subband coefficients. A pf is computed by dividing the energy of a coefficient by the total energy in a subband as follows,

$$s_{d,p,q}(m,n) = x^2_{d,p,q}(m,n) / \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} x^2_{d,p,q}(i,j) \tag{3.4}$$

where $x_{d,p,q}(m,n)$ is the coefficient at location (m,n) in the subband $\mathcal{W}_{d,p,q}$ of size $M \times N$ at depth d and frequency indices p, q. This is done to acquire a distribution of energy in the subbands. This energy distribution is referred to as a probability function of the subband because it follows the properties associated with a probability function (i.e. the sum of all the coefficients is equal to 1). A probability function estimates the energy in a subband. The subband coefficients energy provides an estimate of how much textural information is present in the scale-space resolution represented by the subband.

The FWPT produces sparse representations with textural data localized both in space and frequency in each subband. Hence, the textural data is de-correlated i.e. filtered in to various domains with low frequency and high frequency content found separately in different subbands. Moreover, the subbands also represent the texture at various spatial scales. Hence, the Meningioma texture at various spatial and frequency resolutions is analysed just as a microscope analyses Meningioma slides.

As mentioned earlier, a pf estimates the energy distribution in a subband so that the presence of useful information in the subband can be ascertained. It

also represents the probability of occurrence of texture in a subband at the spatial frequency resolution with each coefficient's energy acting as an estimate of presence of texture. The idea is to estimate the probability of the existence of useful textural characteristics in a subband in the multiresolution wavelet-based decomposition for a texture class. The process is similar to maximum likelihood estimation and the estimate obtained would be referred to as the pseudo probability estimate (ppe) as it is not equivalent to a maximum likelihood estimate. The probability estimate for a subband coefficient is obtained over the entire sample space. We would refer to it as a pseudo probability estimate (ppe) as averaging of normalised energies is performed to derive a ppe.

As stated earlier, the averaging is done to derive an estimate of the energies in each subband for a textural class, so that subsequently when the textures are to be compared (since most distance functions compare energies), the distance functions prove to be effective. Using single images as textural templates is though the simplest mechanism but is not the best approach in this instance as there is considerable intra-class textural variability. No single image, hence, could serve as an effective basis for comparing these textures. Another method could have been comparison of subbands based upon features extracted such as energy of the subband. This would imply averaging the energy of each and every corresponding subband for all images belonging to a textural class. An advantage of such an approach would be that rotation invariance may be acquired. The disadvantage would be that when comparing textures only one feature per subband would be compared rather than a pixel wise comparison or the comparison of statistical properties of subband textures. This may be a viable approach and may become a subject of a future study.

The averaging of the pfs of corresponding subbands from images representing a textural class, such as a Meningioma subtype, estimates the underlying textural properties captured by the spatial-frequency resolution represented by the subband. However, there are two averaging approaches that may be applied i.e. pairwise (using two pfs) or averaging all the pfs over the entire sample space at the same time referred to as pseudo and standard averaging respectively. Both approaches are used to compute the subband probability estimates of the subband pfs so that MAWTTS representing the texture classes can be constructed.

The MAWTTs obtained using standard and pseudo-averaging are described next.

Standard Averaging of the Probability Functions

Standard averaging of the pfs is the averaging of the normalised energies in a subband over the entire texture class, i.e. all images available representing a texture class. This implies that all the samples belonging to a particular Meningioma subtype are decomposed in to a FWPT representation and then the coefficients in the corresponding subbands are averaged. The averaging process is given by,

$$\mathcal{A}^{a}_{m,n} = \sum_{i=1}^{N_a} s^{a_i}_{d,p,q}(m,n)/N_a \tag{3.5}$$

where $s^{a_i}_{d,p,q}(m,n)$ is the (m,n)th coefficient of the pf s of a subband at depth d and frequency indices p,q for the training image a_i belonging to class a while N_a is the number of samples belonging to class a. $\mathcal{A}_{d,p,q}$ is referred to as a pseudo probability estimate of the subband $\mathcal{W}_{d,p,q}$.

Pseudo-averaging of the Probability Functions

The pseudo average of the pfs is computed by iteratively computing the pairwise averages of the corresponding subbands in the wavelet-based multiresolution representation of the images in a texture class. The process can be described as given below,

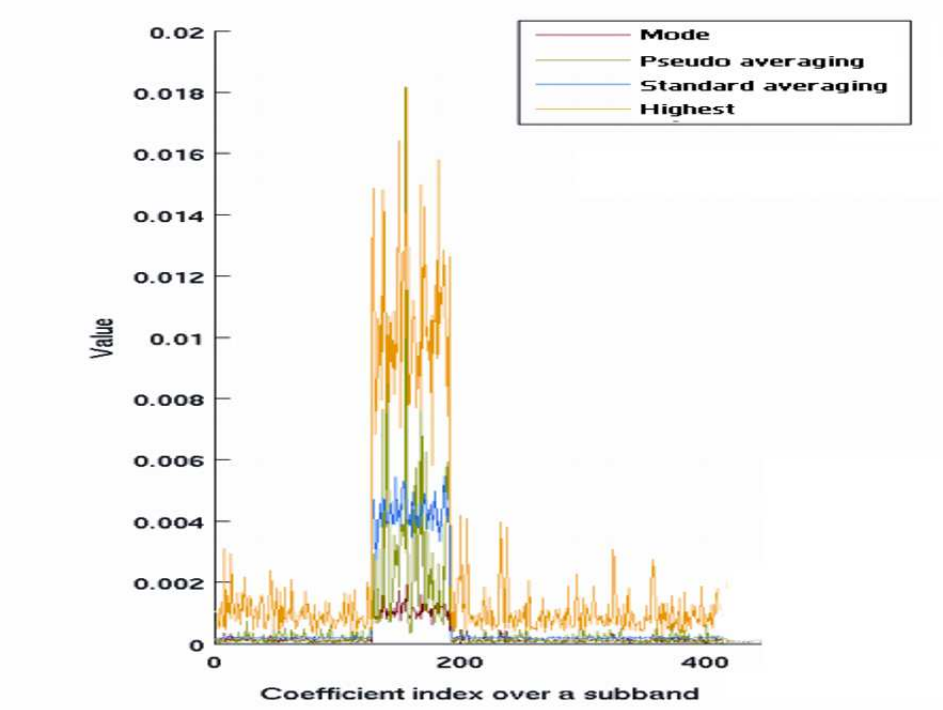
$$\mathcal{A}^{a_2}_{d,p,q}(m,n) = \frac{s^{a_1}_{d,p,q}(m,n) + s^{a_2}_{d,p,q}(m,n)}{2}, \tag{3.6}$$

$$\mathcal{A}^{a_i}_{d,p,q}(m,n) = \frac{\mathcal{A}^{a_{i-1}}_{d,p,q}(m,n) + s^{a_i}_{d,p,q}(m,n))}{2}, \qquad i = 3 \ldots N_a. \tag{3.7}$$

where $s^{a_i}_{d,p,q}(m,n)$ is the (m,n)th value of the pf of a subband $\mathcal{W}_{d,p,q}$ for the image a_i belonging to class a and N_a is the number of training samples available for the texture class a. $\mathcal{A}^{a}_{d,p,q}(m,n)$ corresponds to the pseudo probability estimate (ppe) value for the class a at the m,nth position in the pf. It is important to note that an average of the normalized energies of a subband for two images is computed per iteration. The objective is to acquire a basic model of energy distribution in a subband for each class so that the difference between the classes may be estimated. This averaging is to be referred as pseudo-averaging as it is

different from standard averaging. This is done to account for any sudden rise or fall in the pfs. As can be seen in Figures 3.1 and 3.2, the pseudo-averaging is better able to follow the trends in subband coefficients. It can be seen from the figures that pseudo-averaging approximates the discontinuities (that is, jumps or high values) as well as the most repeated coefficient values (mode) better than standard averaging. Hence, pseudo-averaging provides better probability estimates in comparison to standard averaging.

a.



b.

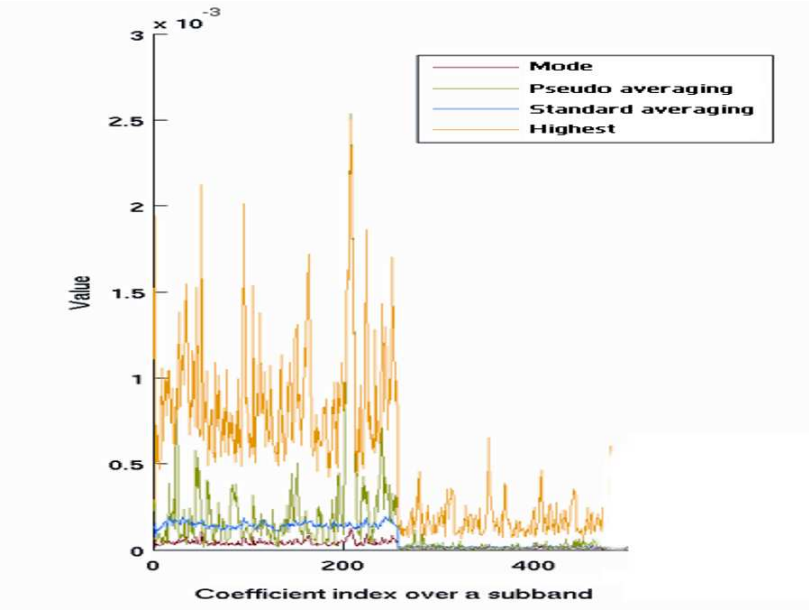
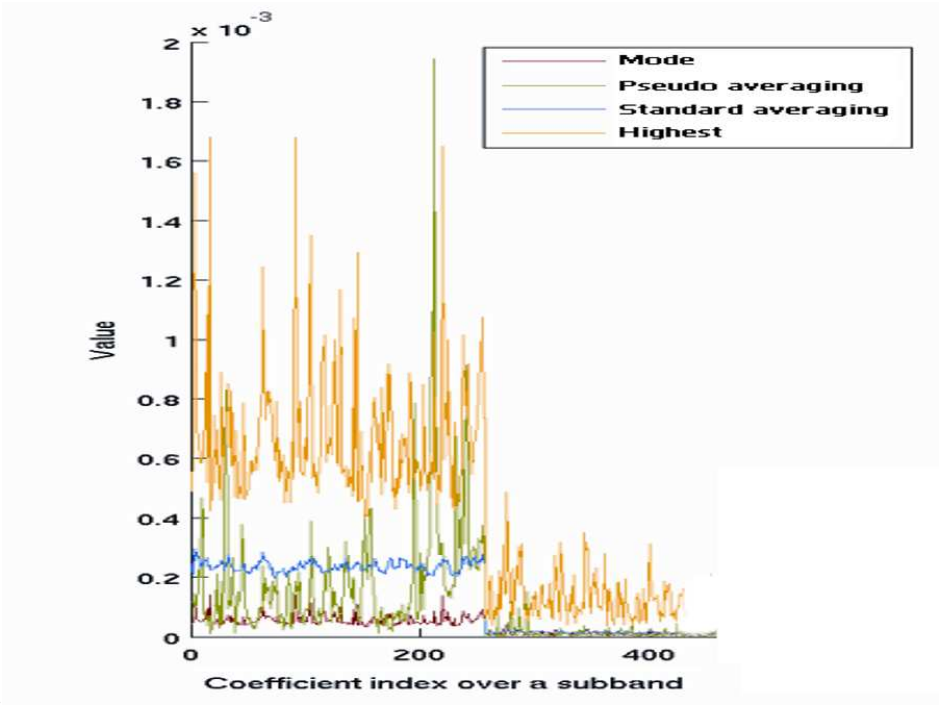


Figure 3.1: Plots of how closely standard averaging and pseudo-averaging of the coefficients per subband approximate the mode (most repeated coefficient) and the highest value recorded for that coefficient over the entire sample space a. Fibroblastic (subband (3,3,0)) b. Meningiotheliamatous (subband (1,0,1))

a.



b.

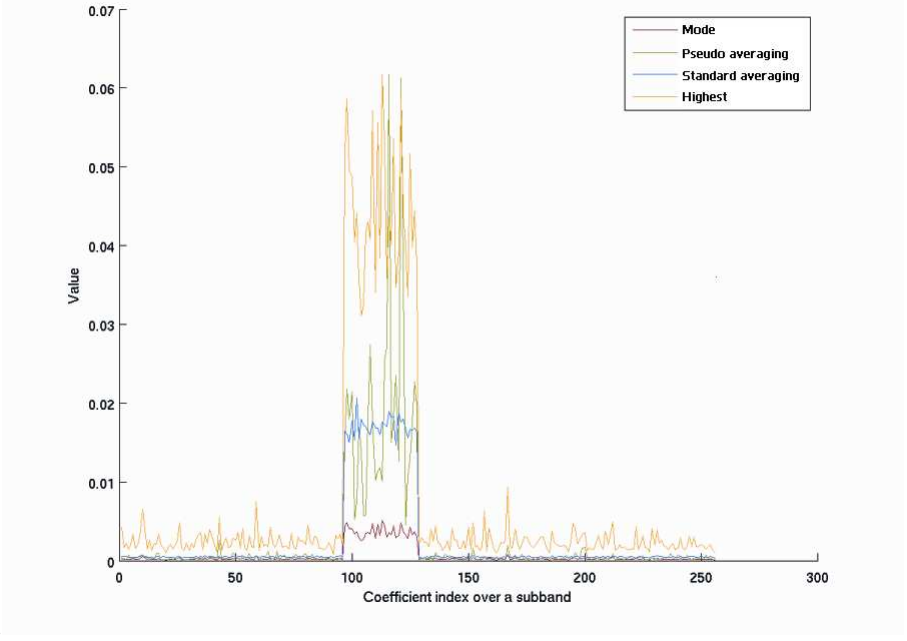


Figure 3.2: Plots of how closely standard averaging and pseudo-averaging of the coefficients per subband approximate the mode (most repeated coefficient) and the highest value recorded for that coefficient over the entire sample space a. Psammomatous (subband (1,0,1)) b. Transitional (subband (4,3,2))

Pseudo-averaging vs. Standard Averaging: An Analysis

The aim of the methods described above is to come up with a multiresolution wavelets decomposition that represents the salient characteristics of the texture classes being compared. The result is a pseudo probability estimate (ppe) which is representative of the textural properties of a texture class at each subband level. The ppes of the pfs of each subband together form the Multiresolution Adaptive Wavelet Texture Templates (MAWTTs). The idea is to obtain MAWTTs that would closely correspond to the textures in the original texture samples. Figures 3.3-3.6 show the MAWTT subbands constructed using pseudo and standard averaging. It can be seen from the figures that the intrinsic texture characteristics such as formation of matrix in Fibroblastic images, lobulated cells in case of Meningiothelial, presence of blobs for Psammomatous and formation of whorls in case of Transitional are captured well with pseudo-averaging. Different subbands in the MAWTTs represent different features as they represent different spatial-frequency information. The MAWTT subbands capture different textural characteristics in the various subbands. We need to select the subbands which are best for discriminating between the textures belonging to the various Meningioma subtypes.

Each value in the ppe represents the textural characteristic at the m, nth coefficient. Hence, we would expect to have a value that is closest to the most repeated value i.e. the mode of the values at that coefficient index and also represent any irregularities (high values). The process is similar to computing probability distribution of an event where if in an experiment an outcome has greater probability of occurrence, the overall probability is updated accordingly.

We are interested in most probable coefficients and most important discontinuities which are better captured by pseudo-averaging in most cases. The discontinuities or irregularities not present in most samples should not overly influence the templates. We see, however, there are instances when pseudo-averaging may not work too well. An obvious issue in pseudo-averaging is that it is weighted averaging i.e. the later probability values in the pseudo-averaging process have a higher impact on the overall pseudo probability estimates. This is resolved by performing the pairwise averaging in both the forward and backward direction and obtaining an average of the two overall pseudo probability estimates. This

would still leave the issue of the weight assigned to the probability densities in the middle of the pairwise computation. The averaging of the forward and backward pairwise averaging solves the problem somewhat but not completely. An obvious solution is to associate equal importance to all the probability values which is achieved with standard averaging. Hence, the MAWTTs using both standard and pseudo-averaging are computed and the results are compared.

However, any other form of averaging would require determining the size of a suitable averaging window (Parzen window for probability estimation) which would be application and data dependent. For instance, in the case of Meningioma subtype classification different averaging windows may be found suitable for different subsets of data. This is due to the inherent intra-class variability. The search for an adequate window can be a subject of another study. We would further discuss this issue in the concluding chapter of this thesis. However, it should be noted that our focus is not on determining the optimal window for comparison but to obtain viable ppes that could be used to compare the textural classes under study. As we see in the Figures 3.3-3.6, good ppes or MAWTTs can be obtained using pseudo-averaging.

It can be seen from graphs in Figures 3.1 and 3.2 that the pseudo-averaging curve tends to oscillate between the mode and the highest frequency for different coefficients but the standard averaging tends to be more stable and does not represent the higher frequencies in the pfs as shown in the figures. This could be seen more clearly from the decompositions in Figures 3.3-3.6 which show that pseudo-averaging represents Meningioma subtype textures (included in the study) much better than standard averaging in all cases. The high frequencies represent edges which are better captured by pseudo-averaging than standard averaging. Hence, better texture templates ($\mathfrak{T}^a$) are acquired using pseudo-averaging. Figures 3.3-3.6 show subbands of the texture templates $\mathfrak{T}_P^a$ and $\mathfrak{T}_S^a$ obtained using pseudo and standard averaging respectively while a denotes the texture class that the template represents.

Since pseudo-averaging is better able to capture the high frequency content in the subbands, therefore, the templates obtained using pseudo-averaging approach are better able to capture the salient features of the texture. On the other hand, the subband templates obtained using standard averaging seem like noise and do

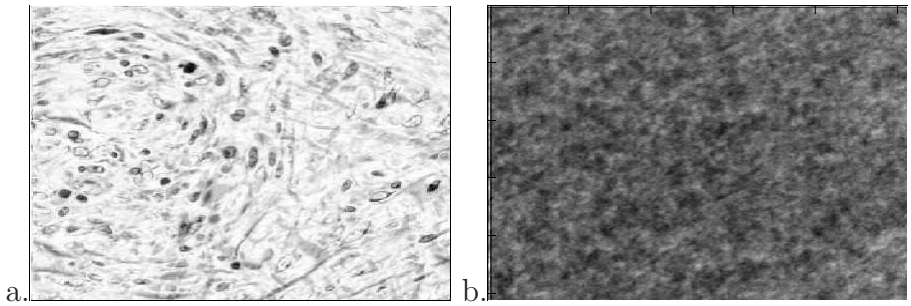


Figure 3.3: Texture templates obtained using pseudo-averaging (a) and standard averaging (b) of Fibroblastic Meningioma subtype at level 1

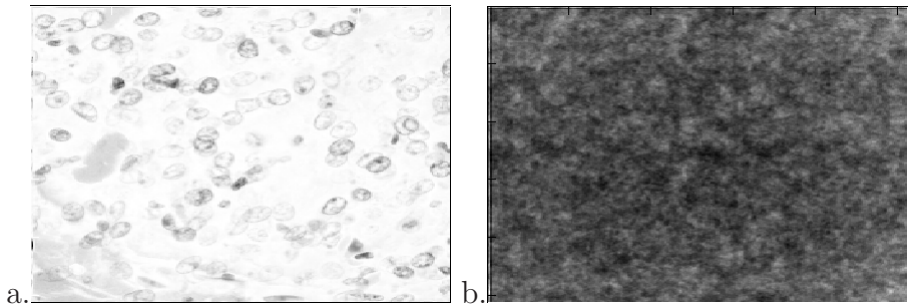


Figure 3.4: Texture templates obtained using pseudo-averaging (a) and standard averaging (b) of Meningiotheliamatous Meningioma subtype at level 1

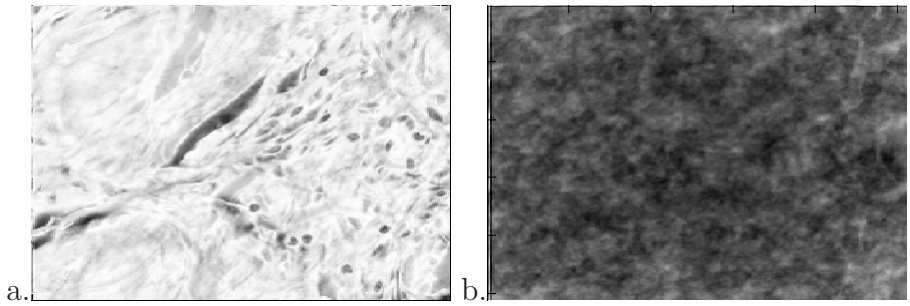


Figure 3.5: Texture templates obtained using pseudo-averaging (a) and standard averaging (b) of Psammomatous Meningioma subtype at level 1

not capture the edges present in the textural classes being analysed. This was expected since our graphs in Figures 3.1 and 3.2 show that standard averaging

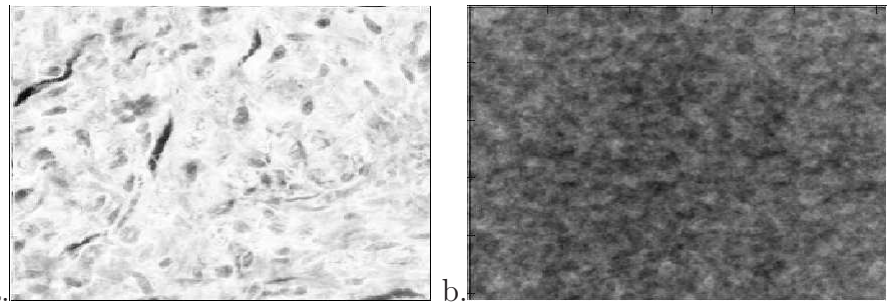


Figure 3.6: Texture templates obtained using pseudo-averaging (a) and standard averaging (b) of Transitional Meningioma subtype at level 1

does not respond well to the high values in the pfs. These high values correspond to edges present in the texture.

The next stage is the calculation of the discrimination power $\mathcal{D}_{d,p,q}$ associated with each subband. There are various methods to do this. We have explored a few well-known distance metrics for this purpose. These include Hellinger distance, Kullback-Leibler divergence, Fishers Linear discriminant, Jensen-Shannon distance, Bhattacharyya distance, Mahalanobis distance, Energy and the novel Relative Energy distance. The description of each distance function is given in Appendix B.1. Other distance functions may also be used for the purpose but we have chosen a few from the various types of distance functions for comparison.

3.1.4 Multiple Classes and Distance Computation

The distance functions used to compute $\mathcal{D}$ in most instances (except in the case of Multi-class Mahalanobis distance) are pairwise i.e. the distance is computed between a pair of texture classes. Since the distance is calculated pairwise, for a four class problem six pairwise distances are computed. Subsequently, the calculation of the overall discriminatory power $\mathcal{P}$ of a subband is given by,

$$\mathcal{P}_{d,p,q} = \sum_{(a,b)\in O} \mathcal{D}^{a,b}_{d,p,q} \tag{3.8}$$

where O is the set containing all the pairwise class combinations of the 4 different classes ($|O| = 6$) for subband $\mathcal{W}_{d,p,q}$. The process is repeated for all the subbands

at various levels in the MAWTT decomposition. This procedure is not required in the case of Multi-class Mahalanobis distance which by definition obtains the discrimination measure between all classes simultaneously. The next stage is the best bases selection.

3.1.5 Best Basis Selection: An Algorithm

As stated in Chapter 2, a wavelet packet decomposition may have multiple bases of representation. We aim to obtain a representation that is best for a specific purpose such as data compression or texture comparison. A best basis is a wavelet packet representation that is obtained for a specific purpose maximising upon a criterion. In this section, we present the algorithm that we employ for the selection of the best basis. This algorithm selects a best basis by maximising the discrimination power of the subbands of a wavelet packet decomposition.

1. Compute the J-level FWPT as described in Section 2.1 for each texture class.
2. Compute the MAWTTs for all subbands and for all classes using the different averaging schemes.
3. Calculate the discrimination power $\mathcal{P}_{d,p,q}\ \forall\ d,p,q$ using Eq.3.8.
4. Initialize $d = J - 1$.
5. For all $0 \leq p < 2^d$, $0 \leq q < 2^d$, do the following:
   - 5a. If $\mathcal{P}_{d,p,q} < max[\mathcal{P}_{d+1,2p,2q}, \mathcal{P}_{d+1,2p,2q+1}, \mathcal{P}_{d+1,2p+1,2q}, \mathcal{P}_{d+1,2p+1,2q+1}]$ Keep the four child subbands at depth $d+1$ where $\mathcal{P}_{d,p,q}$ represents the discrimination power of a node at position p, q and depth d.
   - 5b. Otherwise keep the parent at depth d and remove the child subbands.
6. Decrement d by 1.
7. If $d < 0$, then stop, otherwise goto Step 5.

It can be seen from the algorithm that a comparison is made between the parent and child nodes at every level. This would lead to the selection of the most discriminant subbands yielding a best basis of representation referred to as ADWPT basis for discriminating between Meningioma subtypes. At each instance, a comparison is made between subbands from one branch of the ADWPT tree, i.e. subbands in the approximation domain are not compared with subbands in the detail domains. This would ensure that relevant information is compared. We will compare the high frequency subbands with the low frequency subbands in the final feature selection phase in Chapter 5.

Subband selection for different Meningioma data samples results in selection of different ADWPT decompositions. This is due to the high intra-class texture variation inherent in Meningioma subtype samples. In the next section, we discuss the issue of stability in ADWPT computations. A metric for measuring stability is proposed and the distance functions that produce more stable decompositions are identified.

3.2 ADWPT Stability

There are three factors that affect ADWPT stability namely, the texture in the Meningioma samples used, the type of MAWTT construction mechanism employed and the distance function used for computing the subband discrimination power. In the subsequent section, a discussion on the variation in Meningioma textural data is presented with an analysis on how it impacts the ADWPT decomposition produced for a given distance function.

3.2.1 Meningioma Data Variation

As described in Chapter 1, there can be high intra-class and little inter-class variation in Meningioma samples. Meningioma subtype classification is a non-trivial problem to solve since inter-class variation is low in many instances. Moreover, the texture in the data samples is quite complex in comparison to synthetic textures such as Brodatz etc. It is important to discuss the texture found in the Meningioma samples belonging to different patients as that would ultimately

have an impact on the subband selection and stability of the ADWPT decomposition. Stability of ADWPT decomposition is important since it represents the propensity of the subbands selected to remain consistent when the data for analysis changes. If the ADWPT decomposition changes by too much for the given problem, then no consistent set of features can be extracted for classifying Meningioma subtypes.

Intra-class variation in textures raises the issue that if a patient's image data has texture which is highly variant and is left out from the template construction and the ADWPT computation phase, then the information pertaining to its textural characteristics would not be adequately represented in the ADWPT best bases decomposition. Moreover, the high intra-class variation would also make the task of Meningioma subtypes classification more difficult. We shall see that our technique responds to this challenge in Meningioma subtypes classification very effectively.

3.2.2 Stability of ADWPT Decompositions

As stated earlier, the geometry of the ADWPT decomposition is dependent upon three factors: the distance measure, the type of MAWTT construction mechanism used and the training data. It has been observed that as the data changes the decomposition obtained also changes. Therefore, it is important to investigate the effect of data on the structure of the ADWPT decomposition.

3.2.2.1 Stability Analysis using Meningioma Image Data

As stated in Chapter 1, the data available to us is composed of images obtained from 5 patients per Meningioma subtype for four Meningioma subtypes. We select 4 patients from each subtype leaving one out at each iteration which is analogous to the leave-one-out algorithm used frequently in pattern classification. This is done to investigate the stability of our decompositions in a real world scenario. Hence, 4 patients in total (1 for each Meningioma subtype) are left out at each iteration for testing. This gives a total number of 625 possible combinations of test and training data. Most often the decompositions obtained using the above methodology are similar but sometimes they differ to a great extent. This is

due to the fact that there is high variation in the texture of Meningioma images belonging to different patients.

We perform an analysis for each decision metric and find out which are the most stable or most frequent decompositions. These would eventually be used to evaluate which distance measures are more suitable for our application. The quality of a distance function would be evaluated based upon the stability of the decompositions produced and the classification accuracies obtained with it (which are presented in Chapter 4). The decomposition stability of a distance function is based upon the criterion of how often a few decompositions are repeated i.e. if a few different types of decompositions are produced and they are repeated for most of the data combinations, then the respective distance function would be termed as stable. We carry out an in-depth evaluation of each distance measure and choose a subset of most suitable ones for our application i.e. classification of Meningioma subtypes. The analysis is done for both the pseudo-averaging and standard averaging of the subband pfs.

Applying the ADWPT algorithm with Hellinger distance, a total of 119 different decompositions are obtained in the case of standard averaging. Most of the decompositions occur only once or twice. A few are repeated more frequently. All the decompositions acquired are combined and shown in a single decomposition by including all the subbands found in the various decompositions and indicating how frequently a subband is found in the various decompositions. A measure of stability is indicated by displaying how often a certain subband is found. The overall subband decomposition thus obtained with stability information is shown in Figure 3.7. The colour of each subband indicates its discrimination power $\mathcal{P}_{d,p,q}$, with white being the most discriminant and black being the least discriminant subband.

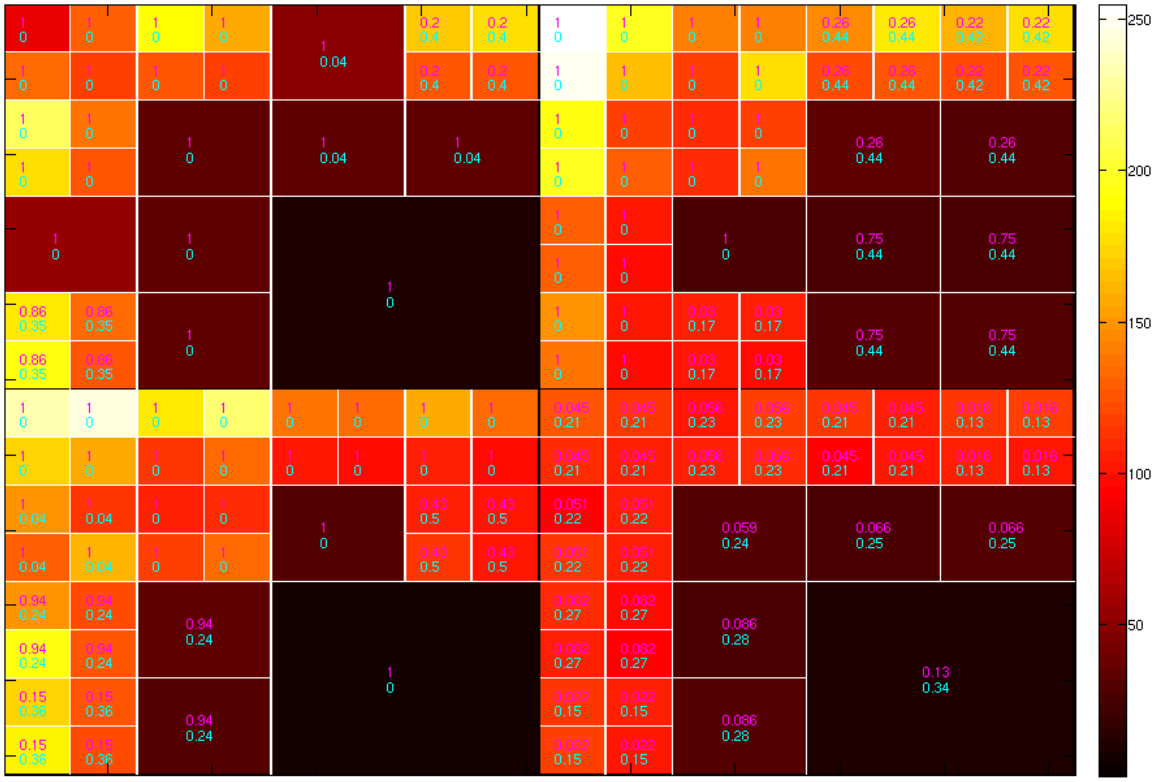


Figure 3.7: Stability of the decomposition obtained using the Hellinger distance function over standard averaging based MAWTTs

The Figure 3.7 shows the probability of occurrence of a subband which is displayed in magenta whereas the number in cyan denotes the standard deviation in the frequency of occurrence of the subband over the different ADWPT decompositions obtained. The stability of a subband is computed as,

$$h_{d,p,q} = \sum_{i=1}^{N_T} N_{d,p,q}(i)/N_T \tag{3.9}$$

where $N_{d,p,q}(i)$ is the total number of times the subband $\mathcal{W}_{d,p,q}$ is found in the decomposition produced for the ith trial run and N_T is the total number of trial runs which is equal to 625. The standard deviation σ^S for a subband at depth d and location p, q is given by,

$$\sigma^S_{d,p,q} = \sqrt{\frac{\sum_{i=1}^{N_T}(N_{d,p,q}(i) - h_{d,p,q})^2}{N_T - 1}} \tag{3.10}$$

Figure 3.8 shows the stability of the decompositions obtained using the Hellinger distance over pseudo-averaging based MAWTTs. Fewer subbands are obtained in comparison to the decomposition obtained using the same distance measure over standard averaging shown in Figure 3.7.

A high value of $h_{d,p,q}$ means that the subband is consistently regarded by the bases selection part of the algorithm (Section 3.1.6) as being discriminant and also robust, as its selection is less dependent on the variation in the data and texture samples. The decompositions showing stability values $h_{d,p,q}$ for each subband are given in Figures 3.7 and 3.8.

A $h_{d,p,q}$ value of 1 denotes that the subband is found in all the decompositions all the time. A standard deviation in the stability values ($\sigma^S_{d,p,q}$) of greater than 0 indicates that a subband may not be found in all the decompositions. These values reveal interesting facts about the stability of the decompositions computed using a distance function. The Table 3.1 shows the average of the stability of the subbands for various distance functions computed using the following expression,

$$\overline{h} = \sum_{p=0}^{N_p-1} \sum_{q=0}^{N_q-1} h_{d,p,q}/N_S \tag{3.11}$$

Figure 3.8: Stability of the decompositions obtained using the Hellinger Distance over Pseudo-averaging based MAWTTs

where N_p and N_q indicates the maximum number of the subband indices and N_S is the total number of subbands.

Table 3.1 indicates that standard averaging based decompositions most often produce a higher value of subband stability. The only exceptions are the Bhattacharyya distance, the multi-class Mahalanobis distance and the Relative Energy distance which are more stable in the case of pseudo-averaging. Interestingly the Energy distance produces a stability value of 1 which is due to the fact that decompositions obtained using it are equivalent to the FWPT and no variability is seen in the ADWPT decompositions produced. The next most stable distance function is the Fisher discriminant for both forms of averaging. The numbers in Table 3.1 reveal interesting facts about the stability of ADWPT decompositions obtained using different distance functions. The difference between the stability values $\overline{h}$ for standard and pseudo averaging remains small in all instances except

	$\overline{h}$(SA)	$\overline{h}$(PA)
HLD	0.37	0.36
K-LD	0.41	0.37
FD	0.92	0.78
J-SD	0.41	0.37
BD	0.71	0.74
MD	0.31	0.15
MCMD	0.23	0.24
EneD	1	1
REneD	0.66	0.70

Table 3.1: Stability analysis of subbands acquired using the various distance functions (HLD=Hellinger Distance, K-LD=Kullback-Leibler Distance, FD=Fisher Linear Discriminant, J-SD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance) and averaging methods (namely SA=Standard Averaging and PA=Pseudo-Averaging)

for Fishers discriminant. These estimates are, however, severely dependent upon the number of subbands found in an ADWPT decomposition. Later in this chapter, we will compare the stability of standard averaging and pseudo-averaging based decompositions using other more objective measures.

Standard averaging vs. Pseudo-averaging Stability

In the above sections, we have seen that the standard averaging is overall more stable than pseudo-averaging. However, the templates obtained using standard averaging contain texture that can be categorised as nothing but noise and have no inherent resemblance to the texture classes that they represent as shown earlier in this chapter. On the other hand, we see that pseudo-averaging tends to create templates that represent the inherent characteristics of the textures under study quite well and captures the salient features of the texture classes being studied. Moreover, the decompositions produced are relatively stable with stability values comparable to standard averaging in most instances. Hence, pseudo-averaging is

a better approach for template construction.

Table 3.2 shows the number of unique decompositions obtained using the various distance functions for the two modes of template construction. $N_{\mathcal{D}}$ indicates the number of unique decompositions produced. A unique decomposition is an ADWPT decomposition which is different from all other decompositions.

	$N_{\mathcal{D}}$(SA)	$N_{\mathcal{D}}$(PA)
HLD	119	16
K-LD	114	16
FD	316	16
J-SD	118	16
BD	557	16
MD	618	43
MCMD	619	81
EneD	1	1
REneD	340	16

Table 3.2: Stability analysis of the decompositions acquired using various distance functions (HLD=Hellinger Distance, K-LD=Kullback-Leibler Distance, FD=Fisher Linear Discriminant, J-SD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance) and averaging methods (namely SA=Standard Averaging and PA=Pseudo-averaging)

The table shows that there is not much stability in the decompositions obtained using standard averaging. This can be seen in the case of all distance metrics and particularly for the two modes of Mahalanobis distance, Bhattacharya distance and Fisher's linear discriminant where more than 300 hundred different decompositions are obtained for a total possible texture sample combinations of 625. This implies that no decomposition is reproduced more than twice whereas in the case of pseudo-averaging we see much more stability. No distance function produces more than 16 different decompositions. Therefore, in the light of the

above observations we conclude that in terms of ADWPT decomposition variability pseudo-averaging of coefficients for MAWTT construction is superior to standard averaging. However, most of these distance functions show high stability in terms of subbands which means that most subbands remain consistent and only a very few subbands change over the various ADWPT decompositions.

Mahalanobis distance is one of the least stable in both averaging modes. This may be due to the fact that the distance depends upon the difference in variance in the probability estimates of the subbands for deriving a measure of the discrimination power whereas the other measures are either based upon coefficient values or the first order statistics of the coefficients. Mahalanobis distance proves to be not suitable for our application as it intends to derive a distance measure based upon the spread of the ppe coefficients covariance and mean of the probability estimates. The use of covariance for distance estimation is unique to Mahalanobis distance and has been found to be not a very suitable approach for ADWPT computation.

In the instance of the Energy distance only 1 unique decomposition is obtained. This decomposition is equivalent to a FWPT which is undesirable since it represents all the subbands that can be decomposed and our aim is to obtain a meaningful reduction of features that would enhance our capability to differentiate between various Meningioma subtypes under study. We will see in the later chapters that the decompositions obtained using the Energy distance fail to produce a high clustering accuracy.

To analyse the stability of each distance function objectively, we propose a stability measure that would indicate how stable a certain distance function is in terms of the decompositions produced. The stability measure could be computed using two methods. First, an estimate of the number of subbands which are decomposed in most iterations would provide us with an estimate of how stable a certain distance function is. If a higher percentage of subbands are decomposed more often then the stability would be high and fewer unique decompositions would be produced. The second source of stability estimation is the number of unique decompositions produced. The two stability measures are inherently linked. Based on this principle, we provide two stability measures $\mathcal{S}_1$ and $\mathcal{S}_2$, with the first estimating stability based upon the frequency of occurrence of the

individual subbands and the other on the total number of unique decompositions obtained. These would be used to objectively compare the stability of distance functions in standard averaging and pseudo-averaging. The first stability measure is hence given by the formula,

$$\mathcal{S}_1 = \frac{\aleph_{hf}}{\aleph_t} \tag{3.12}$$

where $\aleph_{hf}$ indicates the number of subbands which are highly frequent (i.e. subbands with mean frequency of occurrence of 0.75 or higher may be considered more frequent) and $\aleph_t$ indicates the total number of subbands decomposed i.e. all subbands found in the decomposition. In this instance, a sum of most frequent subbands $\aleph_{hf}$ would be given by,

$$\aleph_{hf} = \sum_{i=0}^{t-1} \aleph_i \quad \forall \ \aleph_i = \begin{cases} 1 & \text{if } \aleph_i \geq T_\aleph \\ 0 & \text{otherwise} \end{cases} \tag{3.13}$$

where $T_\aleph = 0.75$. The selection of threshold $T_\aleph$ is arbitrary and a suitable value may be selected based upon the application. In our analysis, we found 0.75 to be a suitable threshold for determining highly frequent subbands. Table 3.3 shows the stability values for $\mathcal{S}_1$ obtained for the various distance functions over the two different forms of averaging. The second measure based upon the number of unique decompositions is given by,

$$\mathcal{S}_2 = 1 - \frac{N_\mathcal{D}}{N_T} \tag{3.14}$$

where $N_\mathcal{D}$ is the number of unique decompositions obtained for a certain distance function $\mathcal{D}$ and N_T represents the total number of possible decompositions. Table 3.4 shows the stability values for $\mathcal{S}_2$ obtained for the various distance functions over the two different forms of averaging. The overall stability measure $\mathcal{S}$ would be an average of $\mathcal{S}_1$ and $\mathcal{S}_2$, which is given by,

$$\mathcal{S} = \frac{\mathcal{S}_1 + \mathcal{S}_2}{2} \tag{3.15}$$

$\mathcal{S}$ would provide us with a measure of stability that ranges between 0 and 1 with 0 representing the least stable and 1 representing the most stable. The results are given in Tables 3.3-3.5.

	$\aleph_{hf}$(SA)	$\aleph_t$(SA)	$\mathfrak{S}_1^{SA}$	$\aleph_{hf}$(PA)	$\aleph_t$(PA)	$\mathfrak{S}_1^{PA}$
HLD	85	145	0.59	72	133	0.54
K-LD	90	151	0.60	69	133	0.52
FD	226	250	0.90	184	244	0.75
J-SD	90	151	0.60	72	142	0.51
BD	138	232	0.59	160	256	0.63
MD	24	253	0.09	8	175	0.05
MCMD	0	256	0	8	253	0.03
EneD	256	256	1	256	256	1
REneD	143	205	0.70	141	253	0.56

Table 3.3: Stability analysis of decompositions acquired using various distance functions (HLD=Hellinger Distance, K-LD=Kullback-Leibler Distance, FD=Fisher Linear Discriminant, J-SD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance) and averaging methods (namely SA=Standard Averaging and PA=Pseudo-Averaging)

	$N_{\mathcal{D}}$(SA)	$N_{\mathcal{D}}$(PA)	$\mathbb{S}_2^{SA}$	$\mathbb{S}_2^{PA}$
HLD	119	16	0.81	0.97
K-LD	114	16	0.82	0.97
FD	316	16	0.49	0.97
J-SD	118	16	0.81	0.97
BD	557	16	0.11	0.97
MD	618	43	0.01	0.93
MCMD	619	81	0.01	0.87
EneD	1	1	1	1
REneD	340	16	0.46	0.97

Table 3.4: Stability analysis of the decompositions acquired using various distance functions (HLD=Hellinger Distance, K-LD=Kullback-Leibler Distance, FD=Fisher Linear Discriminant, J-SD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance) and averaging methods (namely SA=Standard Averaging and PA=Pseudo-averaging)

	$\mathbb{S}_1^{SA}$	$\mathbb{S}_1^{PA}$	$\mathbb{S}_2^{SA}$	$\mathbb{S}_2^{PA}$	$\mathbb{S}_{SA}$	$\mathbb{S}_{PA}$
HLD	0.59	0.54	0.81	0.97	0.7	0.76
K-LD	0.60	0.52	0.82	0.97	0.71	0.75
FD	0.90	0.75	0.49	0.97	0.70	0.86
J-SD	0.60	0.51	0.81	0.97	0.71	0.74
BD	0.59	0.63	0.11	0.97	0.35	0.80
MD	0.09	0.05	0.01	0.93	0.05	0.49
MCMD	0	0.03	0.01	0.87	0.005	0.45
EneD	1	1	1	1	1	1
REneD	0.70	0.56	0.46	0.97	0.58	0.77

Table 3.5: Overall stability measure of the decompositions acquired using the various distance functions (HLD=Hellinger Distance, K-LD=Kullback-Leibler Distance, FD=Fisher Linear Discriminant, J-SD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance) and averaging methods (namely SA=Standard Averaging and PA=Pseudo-averaging)

It can be seen from the Table 3.3- 3.5 that the decompositions obtained using pseudo-averaging are more stable than standard averaging since the difference in the stability measure is high especially in the case of Bhattacharya distance and Mahalanobis distance. All distance functions are more stable in the pseudo-averaging mode compared to the standard averaging mode in terms of the overall stability index $\mathcal{S}$. Energy distance (EneD) is equally stable in both averaging modes. However, there are instances where standard averaging is more stable compared to pseudo-averaging for the first stability index $\mathcal{S}_1$. All distance functions except for the Bhattacharyya distance are more stable in the standard averaging mode as far as $\mathcal{S}_1$ is concerned. A higher stability index $\mathcal{S}_1$ and low stability index $\mathcal{S}_2$ indicates that a small number of subbands occur with high variation causing the overall stability index $\mathcal{S}$ to suffer.

Although the difference in the values is small in most instances except for Fisher discriminant and Relative Energy distance. Overall, as indicated in Table 3.5, pseudo-averaging is overall more stable. An analysis of the stability using our novel methodology has been concluded here. In the next section, we use graph matching to describe the issue of stability from another perspective.

3.2.2.2 Decomposition Stability Analysis using Graph Matching

Decomposition stability $\mathcal{S}$ is a very important factor in determining the utility of a distance function. In the previous section, we carried out an analysis of the stability of the decompositions obtained using a distance function. This is an important factor in the selection of a distance function for further analysis. In this section, we would present a detailed analysis of the decomposition stability in terms of Graph Edit Distances and Maximal Common Subgraphs. This would provide a graphic view of the stability of the decompositions in the figures by showing by how much the decompositions differ. Results for two distance function namely Hellinger and Mahalanobis are shown in this section. This is because Hellinger is one of the most stable whereas Mahalanobis is the least stable distance function. The results for the remaining distance functions can be found in the Appendix B.5.

Graph Matching

Graph matching and finding distances between graphs is one of the most frequently used concepts in pattern matching. It has been used in a variety of areas such as character recognition [142], image registration [143] and 3-D object recognition [144].

Although the legacy algorithms for graph matching are graph and subgraph isomorphisms [145] due to the fact that they are not error tolerant they were replaced by more developed algorithms such as graph edit distances [146]. The graph edit distances represent the similarity of two graphs in terms of the cost of the shortest sequence of edit operations that would transform one graph to another. This method suffers with the weakness that a cost must be associated with each edit operation namely insertion, deletion or substitution of nodes and edges. Maximal common subgraph is a more recent method for error-tolerant graph matching [147, 148]. Bunke has described the relationship between graph edit distances and maximal common subgraphs (MCSs) [149] and then proposed a distance metric based upon MCSs for graph matching [150]. Although any method for graph-matching would be adequate, we use Bunke's method since its error-tolerant and acts as a metric for graph-matching.

Maximum Common Subgraph

A graph g may be defined as,

$$g = (V, E) \tag{3.16}$$

where V is a finite set of vertices or nodes and E is the set of edges linking the vertices. A maximum common subgraph can be defined in terms of isomorphisms of graphs $g_1 = (V_1, E_1)$ and $g_2 = (V_2, E_2)$. A common subgraph of g_1 and g_2 is a graph $g_c = (V_c, E_c)$ such that there exists a subgraph isomorphism from g_c to $g1$ and g_c to $g2$. The subgraph g_c would be referred to as a maximum common subgraph of g_1 and g_2 if it has more nodes than any other common subgraph of g_1 and g_2. A set of isomorphisms G of graph g may be defined as,

$$G = g_i | i = 1...N_g, \sharp|g_i| \leq \sharp|g|, \bigcup_{i=1}^{N_g} g_i = g \tag{3.17}$$

where $\sharp|g_i|$ represents the number of nodes in the graph g_i and $\bigcup_{i=1}^{N_g}$ represents the union operation over all the graphs g_1 to g_{N_g}.

$$g_c = mcs(g_1, g_2) \Rightarrow g_c \epsilon G_1, g_c \epsilon G_2, \sharp|g_c| > \sharp|g_i^1|, \sharp|g_c| > \sharp|g_i^2|; i = 1...N_g \tag{3.18}$$

where $\sharp|g_c|$ represents the number of nodes or vertices in g_c and g_i^1 and g_i^2 denote the various isomorphisms of graphs g_1 and g_2.

In the next section, we use the metric proposed in [150] for measuring graph distances to show how the decompositions change when the input data changes. The variability of the decomposition trees would be computed with respect to the union of the decomposition trees. A union of the decompositions contains all the subbands that are ever composed in any of the data configurations using the leave-one-out procedure.

Distance Functions and Stability Graphs

The distance metric measures the difference amongst two graphs in terms of the Maximum Common Subgraph (MCS). The distance between two graphs g_1 and g_2 is given by the formulae,

$$d(g_1, g_2) = 1 - \frac{\sharp|mcs(g_1, g_2)|}{max(\sharp|g_1|, \sharp|g_2|)} \tag{3.19}$$

$$\Rightarrow d(g_1, g_2) = 1 - \frac{\sharp|g_c|}{max(\sharp|g_1|, \sharp|g_2|)} \tag{3.20}$$

where $\sharp|g_1|$ and $\sharp|g_2|$ represents the number of nodes or vertices in graphs g_1 and g_2 respectively. This distance would be computed for each decomposition tree in the 625 different decompositions. Each tree g_t represents a decomposition $\mathcal{B}_t^*$ which would be compared against the union of all trees $\mathcal{B}_{\bigcup}^*$.

$$\mathcal{B}_{\bigcup}^* = \bigcup_{i=1}^{N_t} \mathcal{B}_i^* = \mathcal{B}_1^* \bigcup \mathcal{B}_2^* \bigcup \dots \mathcal{B}_{N_t}^* \tag{3.21}$$

The union of the decomposition bases $\mathcal{B}_i^*$ where $i = 1, 2, 3....N_t$ are used for analysis of decomposition stability. The graphs obtained for Hellinger and Mahalanobis distance functions are provided here while the remaining graphs can be found in the Appendix B.5. Hellinger is one of the most stable in terms of ADWPT decomposition stability whereas Mahalanobis is one of the least stable.

Hellinger Distance

This is one of the relatively stable distance functions. The graph for this is shown in Figures 3.9 & 3.10. Hellinger distance over pseudo-averaging is much

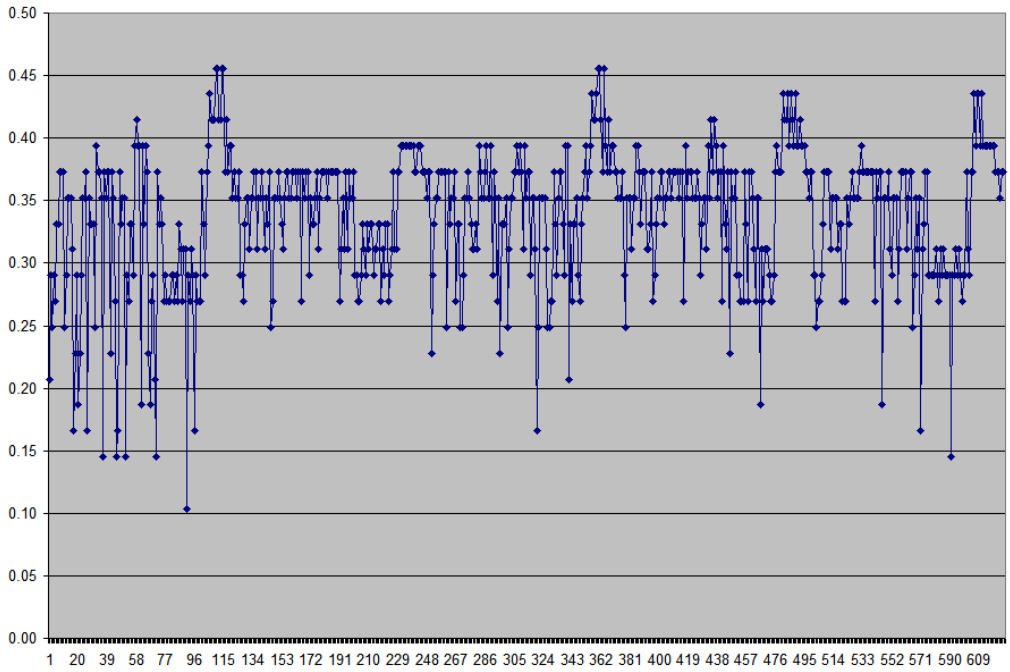


Figure 3.9: Graph showing distances between various decompositions from the Union of all decompositions obtained using Hellinger Distance over Standard Averaging

more stable. This can be seen from the graphs in Figures 3.9 & 3.10. It can be seen from the graph that a few decompositions are obtained again and again, since the same distance values are repeated. On the other hand, standard averaging curve seems to be erratic and shows that there is a great variety in the

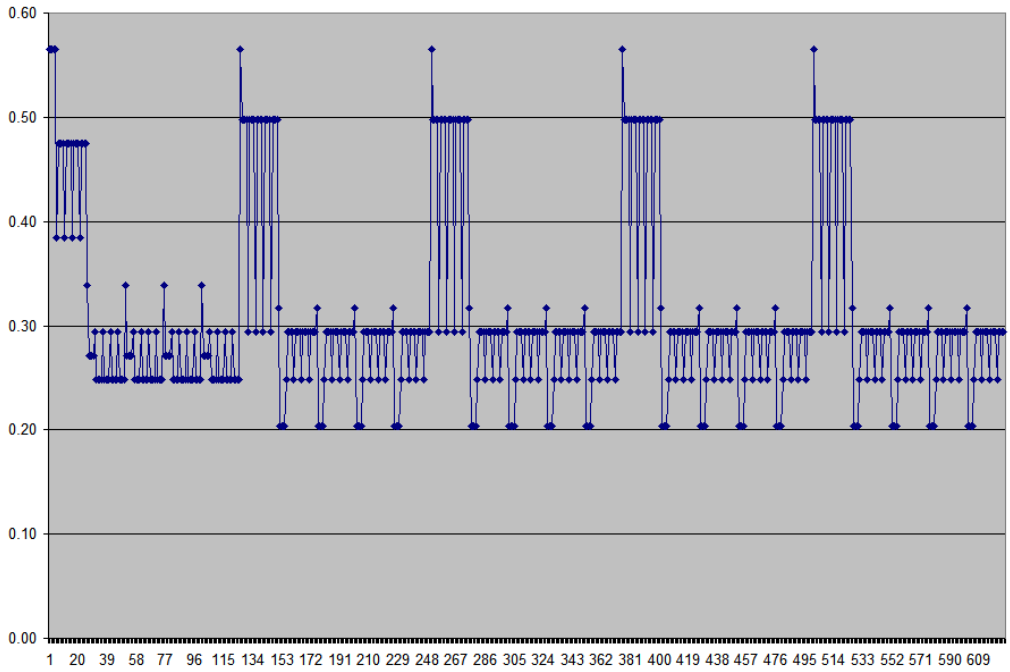
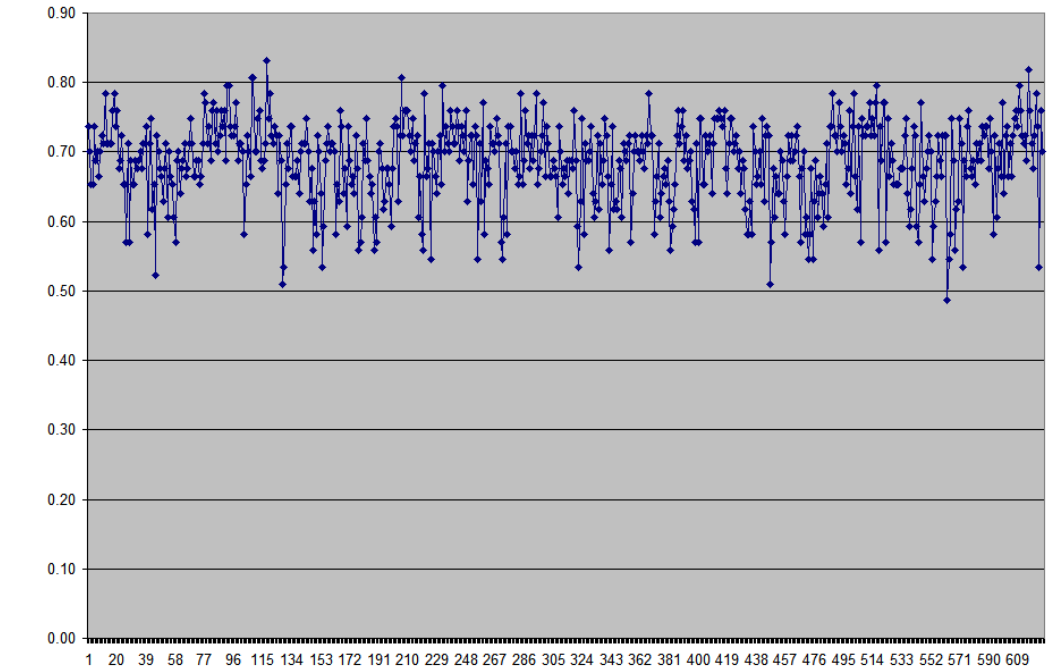


Figure 3.10: Graph showing distances between various decompositions from the Union of all decompositions obtained using Hellinger Distance over Pseudo-averaging

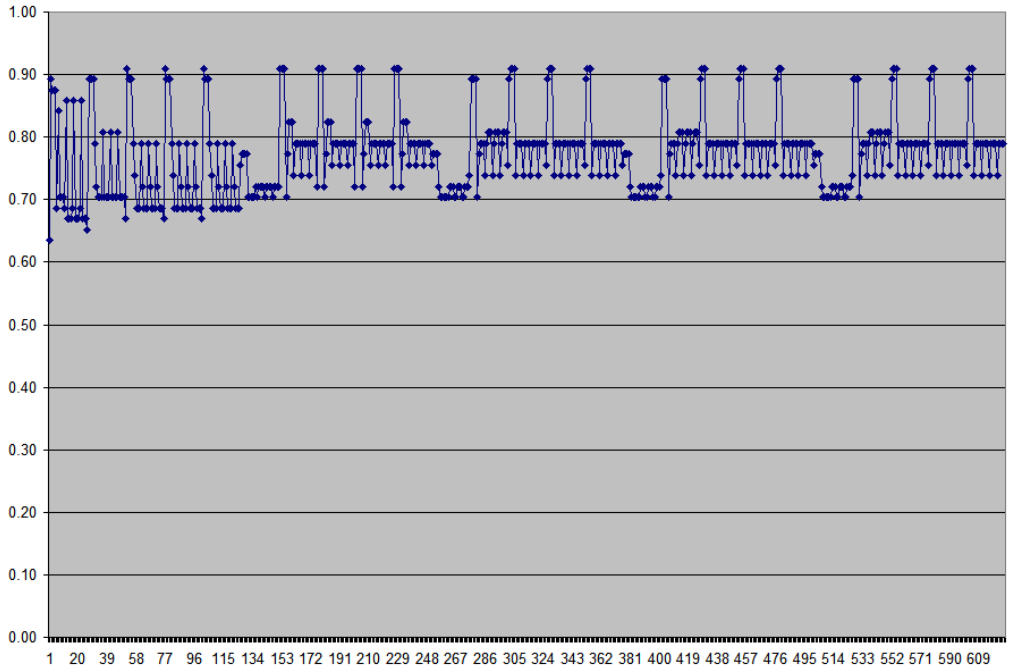
decompositions obtained with Hellinger distance over standard averaging.

Mahalanobis Distance

The decompositions produced by the two cases of Mahalanobis distance are quite interesting as the stability is low. Although the various decompositions do not differ much in terms of distance in the case of multiple classes for Mahalanobis distance but the number of unique decompositions produced are greater than the two class Mahalanobis distance. This could be seen from the graphs in the Figure 3.11 and 3.12. There are still a few decompositions that are repeated more frequently than others but there is no systematic repetition of decompositions. It can also be seen that the decompositions obtained using standard averaging show greater variety and the curve obtained shows erratic behaviour.

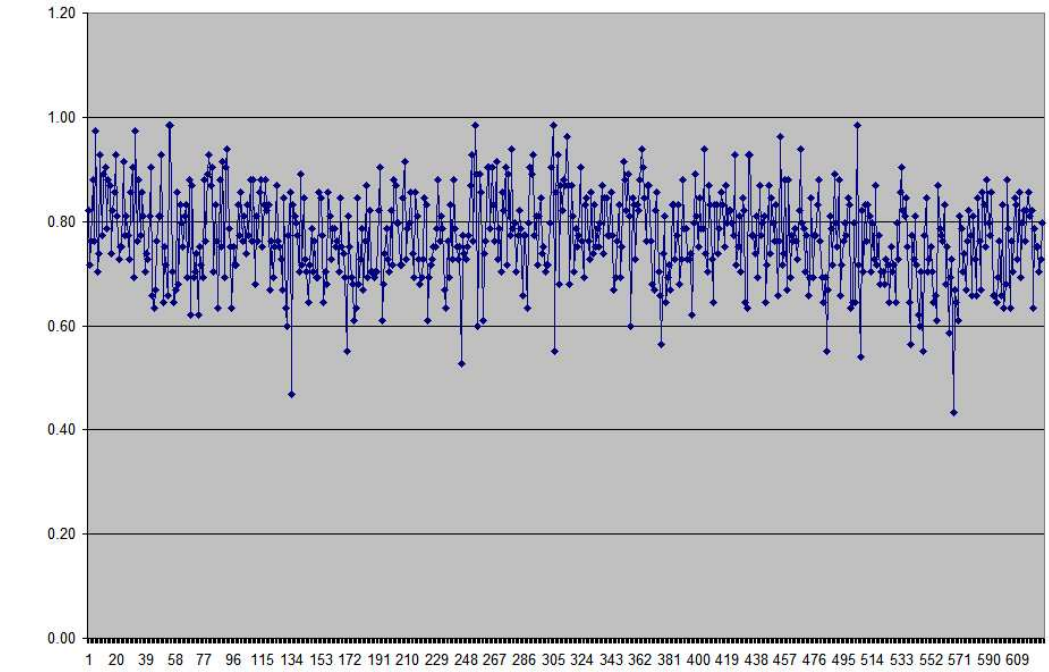


a.

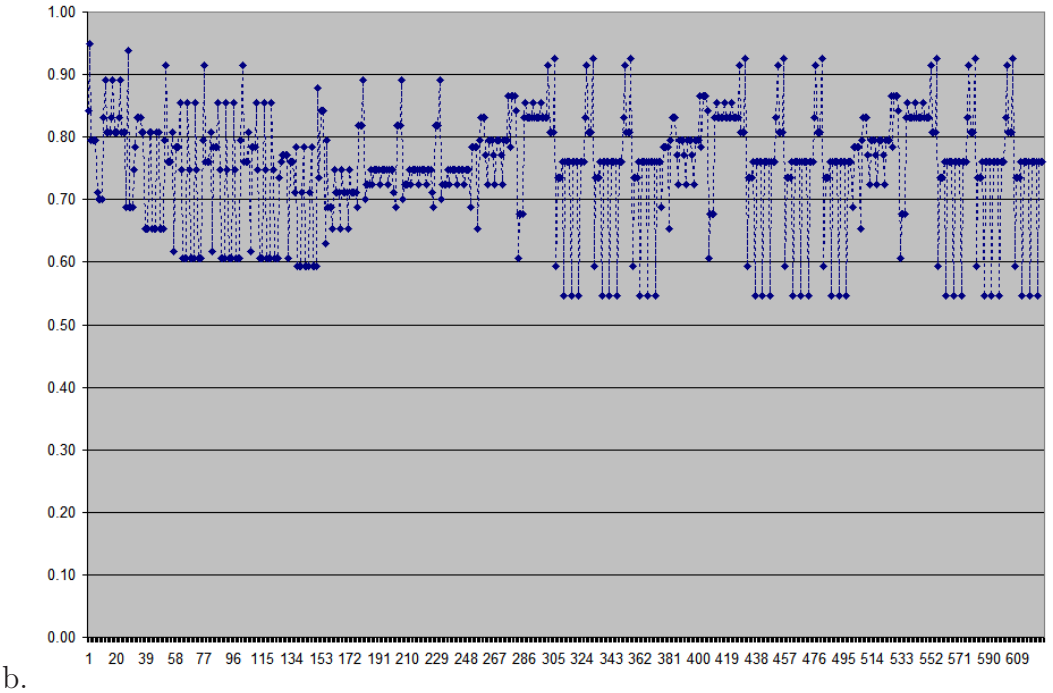


b.

Figure 3.11: Graph showing distances between various decompositions from the Union of all decompositions obtained using Mahalanobis Distance over a. Standard Averaging and b. Pseudo-averaging



a.



b.

Figure 3.12: Graph showing distances between various decompositions from the Union of all decompositions obtained using Multi-class Mahalanobis Distance over a. Standard Averaging and b. Pseudo-averaging

The results shown in the figures are very interesting in the respect that they describe how the decompositions change when the data changes for a specific function. Most of the distance functions, exhibit a repeatable pattern in the case of pseudo-averaging. This is a very desirable trait as it implies that the decompositions would only change when the change in data is substantial. Though, this is true only for the case of pseudo-averaging. Standard averaging has been shown to produce very variable decompositions.

There are however some functions that are more unstable. In case of pseudo-averaging, we can see that the multiple class Mahalanobis distance is relatively unstable when it comes to decomposition stability. This distance function is unique as it allows us to measure the variability in various texture classes at the same time. On the other hand, the variability in decomposition trees in case of standard averaging may be attributed to poor representation of textures by the templates obtained using standard averaging.

Another important factor apart from decomposition stability that influences our choice of a distance function is the clustering or classification performance. In the next chapter, we perform clustering using the various distance functions and identify which distance functions are better for Meningioma subtype classification. An analysis in terms of both stability and clustering accuracy is presented, concluding with the selection of a set of the best distance functions for Meningioma subtype classification.

3.3 Summary and Conclusions

In this chapter, we described in detail the computation of ADWPT with a focus on the two core processes of template construction and computation of the subband discrimination power. Two techniques of template construction namely standard averaging and pseudo-averaging of the probability functions are compared. It can be seen that better texture templates are obtained for pseudo-averaging.

As stated earlier, ADWPT obtains a best basis for texture discrimination of textures such as Meningioma subtypes. There are various ways in which the discrimination power of a subband may be estimated and some distance functions for the purpose have been presented. ADWPT is adaptive as indicated in this

chapter. This raises the issue of ADWPT best bases variability when the data used for analysis changes. This is referred to as the issue of stability and it occurs due to the intra-class texture variation in Meningioma subtypes. In this chapter, we also investigated the issue of stability of the ADWPT best bases indicating the distance functions and the MAWTT construction mode that is better for acquiring more stable ADWPT decompositions. Mahalanobis distance was found to be the least stable distance function. In the next chapter, some further analysis of the various distance functions would be presented indicating which distance functions are better considering the stability of ADWPT best bases and the clustering accuracy of the Meningioma subtypes.

Chapter 4

Towards a Robust Discriminant Basis

As stated earlier, in this thesis we propose a wavelet-based multiresolution texture analysis approach (referred to as ADWPT) for Meningioma subtypes classification. The most important aspect of wavelet packets analysis is the selection of subbands based upon discrimination power. In the previous chapter, the details of the technique pertaining the selection of subbands using the various distance functions is presented. In this chapter, we evaluate each distance function in terms of the stability of the ADWPT decomposition and the clustering accuracy acquired. We also show that more stable subbands produce better clustering results.

Section 4.1 presents all the steps followed for clustering of Meningioma subtype features acquired using ADWPT decompositions. Multiple decompositions are acquired using various distance functions and their results compared. Finally, a subset of distance functions which are better for Meningioma subtype classification are selected.

Due to the intra-class variation, stability of the ADWPT decomposition is an issue. Changes in the data-set lead to different ADWPT decompositions. The aim is to select the most representative and optimal ADWPT decomposition. Our solution is the selection of subbands based upon their stability index, i.e. the probability of occurrence as described in Chapter 3. We employ two schemes. The first involves using all the subbands that are decomposed in the various

different decompositions obtained for a particular distance function. The second involves selecting only those subbands that are found in all the decompositions produced using the distance function. These would be referred to as the union and intersection decompositions, respectively. We will show that the subbands found in the intersection of decompositions are better for classification accuracy in comparison to the subbands found in the union of decompositions as there is a substantial reduction of features acquired using intersection while there is no reduction in classification accuracy.

4.1 Union and Intersection of ADWPT Decompositions

In Chapter 3, we have used the ADWPT algorithm to obtain various wavelet packet decompositions using different distance functions. Since, more than one unique decomposition is obtained, we need to derive a decomposition that is representative of the variety of subbands selected by ADWPT. The decomposition computed must represent salient features for the textures being compared. For this purpose, we obtain union and intersection of all the decompositions obtained for a distance function using the ADWPT. It is important to understand here that intersection is inherently the selection of subbands who have a probability of occurrence of 1 whereas a union is essentially selection of all the terminal subbands that are found in any of the best basis obtained. Subsequently, in Chapter 5 we show that selection of subbands may also be carried out by thresholding i.e. different thresholds may be used and different number of subbands selected. The selection is dependent upon statistical significance of a subband which is derived based upon its frequency of occurrence in all the different best basis acquired for the various test-trial runs. This measure would be updated for every new image acquired and classified using the system. Hence, the best basis would adapt statistically to new data. Any noise would be filtered based upon its statistical insignificance. The way to acquire union and intersection is described next.

4.1.1 Union of ADWPT Decomposition

As stated earlier, in order to evaluate the effectiveness of a distance function, we need to obtain an overall ADWPT representation for a particular distance function. To that end we obtain a union of all the different ADWPT representations that contains all the subbands that have been decomposed for a particular distance function for all the trial runs. The union decomposition $\mathcal{B}^*_{\bigcup}$ is defined as,

$$\mathcal{B}^*_{\bigcup} = \bigcup_{i=1}^{N_t} \mathcal{B}^*_i = \mathcal{B}^*_1 \bigcup \mathcal{B}^*_2 \bigcup \dots \mathcal{B}^*_{N_t} \tag{4.1}$$

where B^*_i represents the best bases obtained using the ith dataset, $i = 1, 2, 3 \dots N_t$ and N_t is the total number of trial runs. Union decompositions for some distance functions are given in Figure 4.1 while the rest are given in the Appendix B.2.

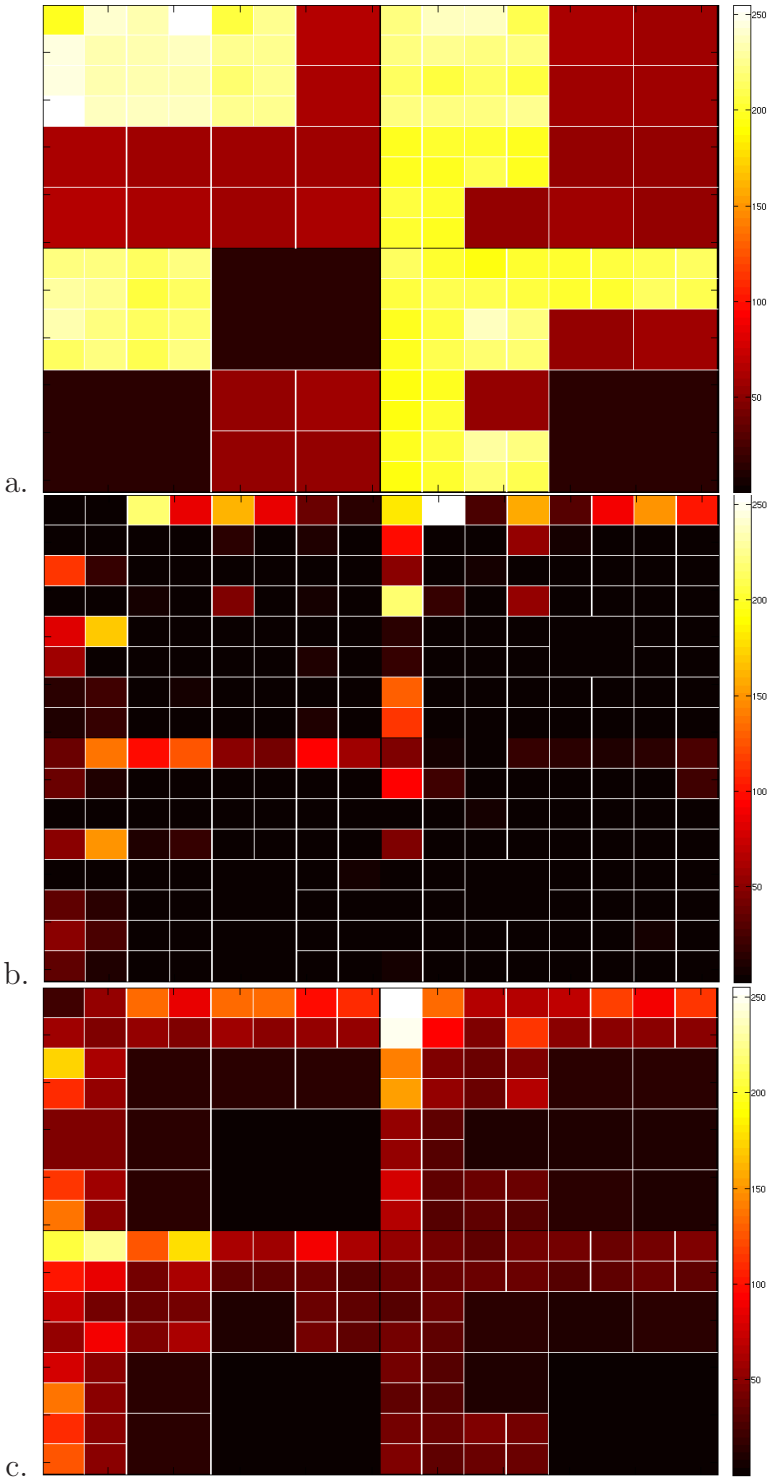


Figure 4.1: Union of ADWPT decompositions obtained using the a. Hellinger Distance (Pseudo-averaging), b. Fishers Discriminant (Pseudo-averaging) and c. Kullback-Leibler Distance (Standard averaging)

The color of a subband indicates its discrimination power. The darker the subband, the less discriminant it is and vice versa.

4.1.2 Intersection of ADWPT Decompositions

Similar to the selection of decomposed subbands using the union paradigm, we also select subbands based upon the intersection of decompositions. In this case, only those subbands are selected which are found in all the decompositions. This implies that any subband which is not decomposed in all of the trial runs is not represented in the ADWPT intersection decomposition. The intersection decomposition $\mathcal{B}^*_{\bigcap}$ is defined as.

$$\mathcal{B}^*_{\bigcap} = \bigcap_{i=1}^{N_t} \mathcal{B}^*_i = \mathcal{B}^*_1 \bigcap \mathcal{B}^*_2 \bigcap \dots \mathcal{B}^*_{N_t} \tag{4.2}$$

where $\mathcal{B}^*_i$ represents the best wavelet bases obtained using the ith dataset and $i = 1, 2, 3 \dots N_t$. The intersection operation would result in a decomposition $\mathcal{B}^*_{\bigcap}$ containing subbands with a decomposition frequency of occurrence or stability index of 1. The intersection decompositions for some distance functions are given in Figure 4.2 while the rest can be found in Appendix B.3.

The union and intersection decompositions for each distance function would subsequently be used to evaluate how good a distance function is for differentiating between the tumour subtypes under study. The subbands in $\mathcal{B}^*_{\bigcup}$ and $\mathcal{B}^*_{\bigcap}$ are then used to extract feature sets representing each image. These feature sets would subsequently be used for clustering in order to evaluate how good a distance function is for Meningioma subtype classification.

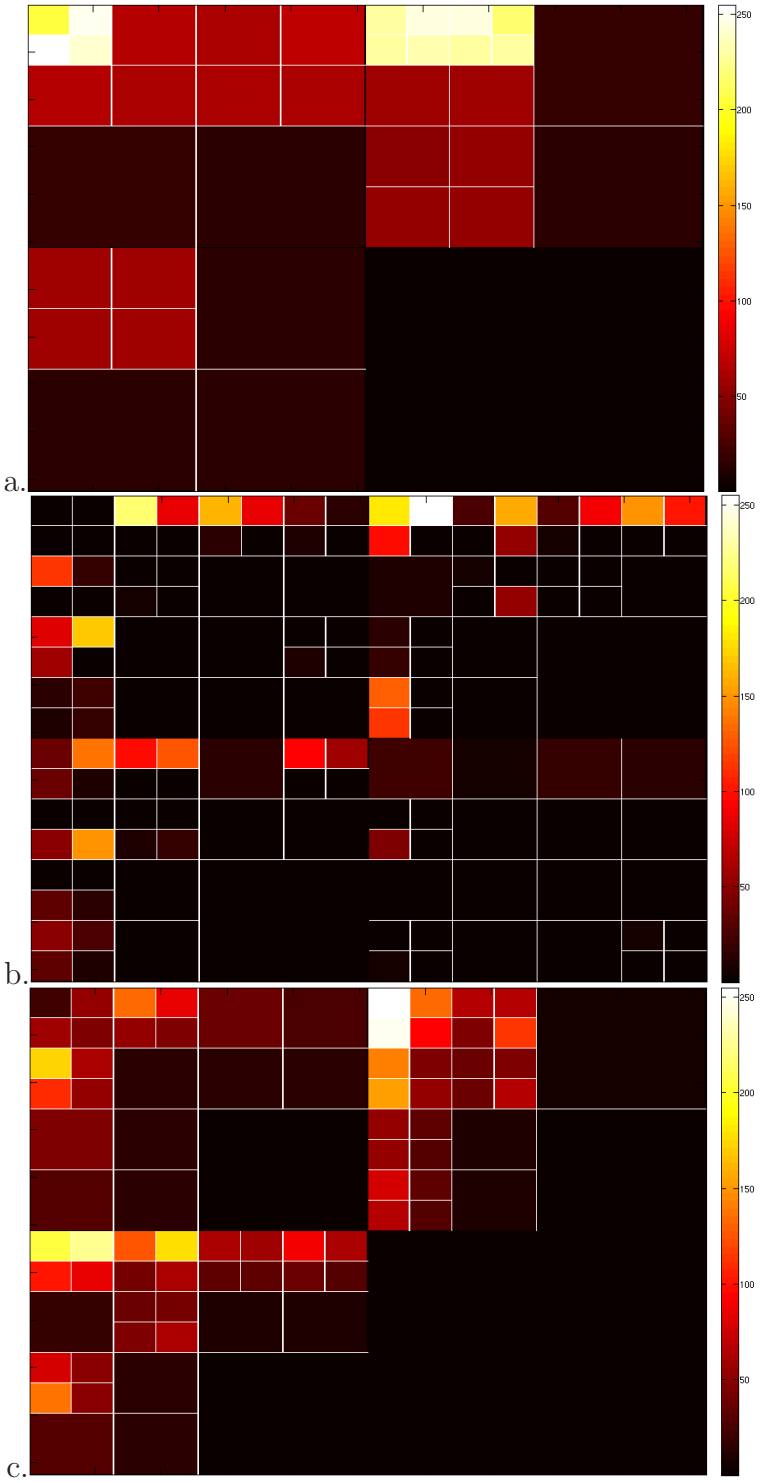


Figure 4.2: Intersection of ADWPT decompositions obtained using the a. Hellinger Distance (Pseudo-averaging), b. Fishers Discriminant (Pseudo-averaging) and c. Kullback-Leibler Distance (Standard averaging)

4.2 Feature Extraction

In the first step, the union and intersection decompositions obtained using ADWPT are computed for each image in the data set. Next, we acquire all the terminal subbands representing a decomposition for all the images as shown in Figures 4.1 and 4.2 and then statistical features are computed for each subband. These statistical features are subsequently used for clustering and classification of Meningioma subtypes. These features represent characteristics of a subband and are highly effective in acquiring textural characteristics as we shall see in the results section.

Many statistical features were obtained and evaluated for their clustering and classification efficiency. The statistical features used were first-order histogram based features, co-occurrence matrix based features and Grey-Tone Difference Matrix (GTDM) features [151]. None of the GTDM features perform as well as GLCM features. Some of the features that performed better were selected and are described below.

Mean

If the coefficients of a subband $\mathcal{W}_{d,p,q}$ are represented by $x^{a_i}_{d,p,q}$, then the mean $\mu^{a_i}_{d,p,q}$ may be computed as

$$\mu^{a_i}_{d,p,q} = \left[\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} x^{a_i}_{d,p,q}(m,n)\right] / M \times N \tag{4.3}$$

where $x^{a_i}_{d,p,q}(m,n)$ is the (m,n)th coefficient of the subband in a decomposition for an image a_i belonging to class a with $M \times N$ as the size of the subband.

Standard Deviation

The standard deviation $\sigma_{d,p,q}$ of the ppde of the subband $\mathcal{W}_{d,p,q}$ is given by

$$\sigma^{a_i}_{d,p,q} = sqrt\frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}[x^{a_i}_{d,p,q}(m,n) - \mu^{a_i}_{d,p,q}]^2}{M \times N - 1} \tag{4.4}$$

Skewness

The skewness $\nu_{d,p,q}^{a_i}$ of the ppde of the subband $\mathcal{W}_{d,p,q}$ is given by,

$$\nu_{d,p,q}^{a_i} = \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}[x_{d,p,q}^{a_i}(m,n) - \mu_{d,p,q}^{a_i}]^3}{(M \times N - 1)(\sigma_{d,p,q}^{a_i})^3} \tag{4.5}$$

Kurtosis

The kurtosis $\Gamma_{d,p,q}^{a_i}$ of the ppde of the subband $\mathcal{W}_{d,p,q}$ is derived as,

$$\Gamma_{d,p,q}^{a_i} = \left[\frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}(x_{d,p,q}^{a_i}(m,n) - \mu_{d,p,q}^{a_i})^4}{(M \times N - 1)(\sigma_{d,p,q}^{a_i})^4}\right] - 3 \tag{4.6}$$

Energy

The normalised energy $\xi_{d,p,q}^{a_i}$ of the subband $\mathcal{W}_{d,p,q}$ is given by,

$$\xi_{d,p,q}^{a_i} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}(x_{d,p,q}^{a_i}(m,n))^2 \tag{4.7}$$

Entropy

The entropy $\varepsilon_{d,p,q}^{a_i}$ of a subband $\mathcal{W}_{d,p,q}$ is computed by,

$$\varepsilon_{d,p,q}^{a_i} = -\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}(x_{d,p,q}^{a_i}(m,n)) * log(x_{d,p,q}^{a_i}(m,n)) \tag{4.8}$$

Grey Level Co-occurrence Matrix (GLCM) Features

GLCM features were shown as a viable approach for texture analysis by Haralick *et al.* [135]. The GLCM is computed over each subband. The aim is to acquire intrinsic features representing each subband. The Gray Level Co-occurrence Matrix C over a subband $\mathcal{W}_{d,p,q}$, parameterised by an offset $(\triangle x, \triangle y)$ is computed,

$$C_{d,p,q}^{\triangle x,\triangle y}(i,j) = \begin{cases} C_{d,p,q}^{\triangle x,\triangle y}(i,j) + 1 & \text{if } \mathcal{W}_{d,p,q}(m,n) = i \text{ and } \mathcal{W}_{d,p,q}(m+\triangle x, n+\triangle y) = j \\ C_{d,p,q}^{\triangle x,\triangle y}(i,j) & \text{otherwise} \end{cases}$$

where $\mathcal{W}_{d,p,q}$ is the p,qth subband at depth d. The $(\triangle x, \triangle y)$ defines the offset in the horizontal and vertical direction and m,n represents the coefficients index. The offsets $(\triangle x, \triangle y)$ are used to define the distance and the angle of the direction in which co-occurrence of pixel values is being searched. A value of $(1, 1)$ would mean an angle of 45^o and a distance of 1 whereas $(-2, 0)$ would mean an angle of -180^o and a distance of 2. GLCM analysis at various number of bins was carried out and 32 bins were found to be the best for Meningioma subtypes classification. The GLCM matrix is computed for four directions i.e. 0^o, 45^o, 90^o, 135^o and the distance remains equal to 1. Other distance measures such as 2, 3 and 4, using the above directions, were also used for clustering and classification but results produced are not better, hence we present results for distance 1. The features are computed for a normalised GLCM matrix which is given by:

$$Q_{d,p,q}^{\triangle x,\triangle y} = C_{d,p,q}^{\triangle x,\triangle y}(i,j) / \sum_{i,j=0}^{N_C-1} C_{d,p,q}^{\triangle x,\triangle y}(i,j) \tag{4.9}$$

where $N_C \times N_C$ is the size of the GLCM and N_C corresponds to the number of bins. The GLCM features used in this study are given below.

GLCM Contrast

The contrast feature for the normalised GLCM matrix $Q_{d,p,q}^{\triangle x,\triangle y}$ may be computed as:

$$\mathcal{G}_{\varsigma,d,p,q} = \sum_{i,j=0}^{N-1} (i-j)^2 Q_{d,p,q}^{\triangle x,\triangle y}(i,j) \tag{4.10}$$

Contrast measures the difference or the gradient in coefficient values over a subband.

GLCM Correlation

The correlation feature for a normalised GLCM matrix $Q_{d,p,q}^{\triangle x,\triangle y}$ is defined by the equation:

$$\mathcal{G}_{\upsilon,d,p,q} = \sum_{i,j=0}^{N-1} \frac{(i-\mu_{d,p,q})(j-\mu_{d,p,q})}{(\sigma_{d,p,q})^2} Q_{d,p,q}^{\triangle x,\triangle y}(i,j) \tag{4.11}$$

where $\mu_{d,p,q}$ and $\sigma_{d,p,q}$ represent the mean and standard deviation of $\mathcal{W}_{d,p,q}$. Correlation determines the relationship in terms of how the coefficient values vary over a subband.

GLCM Energy

Energy is a measure of the number of non zero coefficients or the presence of texture in a subband. The Energy feature may be computed as:

$$\mathcal{G}_{\varepsilon} = \sum_{i,j=0}^{N-1} (Q_{d,p,q}^{\triangle x, \triangle y}(i,j))^2 \tag{4.12}$$

GLCM Homogeneity

Homogeneity reflects the degree of similarity of coefficient values over a subband. The homogeneity feature for a normalised GLCM matrix $Q_{d,p,q}^{\triangle x, \triangle y}$ is given by:

$$\mathcal{G}_{\tau} = \sum_{i,j=0}^{N-1} \frac{Q_{d,p,q}^{\triangle x, \triangle y}(i,j)}{1+(i-j)^2} \tag{4.13}$$

Each of these features capture a textural property of the image and will be used in subsequent sections for k-means clustering of the texture samples. The next stage is clustering based upon the features presented above.

4.3 Clustering

Over the years many algorithms have been developed for clustering data. k-means clustering [152] is one algorithm that has been used widely for computing clusters [153] [154] [155]. Rand indices is one of the mechanisms used to ascertain the quality of a cluster in terms of how accurately it represents the data characteristics [156] [157]. Both these techniques are used in our clustering analysis.

4.3.1 k-Means Clustering

k-Means clustering is one of the simplest algorithms for unsupervised classification. It aims to produce k clusters by obtaining centroids c_j where j=1 to k. The process of computing the centroids involves the selection of certain initial points

in the data as centres and then finding all the data points that are nearer to these centres and grouping them into respective clusters. Next the centroids are re-calculated with the previously clustered data and new centroids are estimated. The algorithm is repeated until the centroids no longer shift. At each iteration the algorithm seeks to minimise the Euclidean distance between the data points x_{ij} and the centroid c_j which is given by,

$$\Delta E = \sqrt{\Sigma_{j=1}^{k} \Sigma_{i=1}^{N_f} (x_{ij} - c_j)^2} \tag{4.14}$$

Since the clusters obtained are sensitive to the selection of initial points of reference, the clustering procedure is repeated 5 times and the result with the best clusters is selected as the final result. The features discussed in the previous section form the input feature set for k-Means clustering. A feature value is computed for each subband separately representing an image and a feature set is obtained with length N_f which is equal to the number of subbands in the decomposition. It must be noted here that this is done for the union and intersection decompositions separately.

4.3.2 Rand Indices

Rand index [157] provides a measure of similarity between clusters in two clustering results. It ranges from 0 to 1 where 0 corresponds to the instance where the clusterings are completely dissimilar whereas 1 is the value when the clusters are absolutely the same. Say we need to compare two clusterings namely C_1 and C_2 where C_1 is the ground truth i.e. the true value of the cluster to which a data point belongs to while C_2 is the clustering obtained by an unsupervised clustering method such as k-Means clustering. If G_i^1 and G_j^2 are clusters found in the clusterings C_1 and C_2 respectively, then the Rand index may be computed as follows

$$\mathcal{R} = \frac{C_2^n + \sum_{i=1}^{G_i^1} \sum_{j=1}^{G_j^2} n_{ij}^2 - \frac{1}{2} \sum_{i=1}^{G_i^1} (\sum_{j=1}^{G_j^2} n_{ij})^2 - \frac{1}{2} \sum_{j=1}^{G_j^2} (\sum_{i=1}^{G_i^1} n_{ij})^2}{C_2^n} \tag{4.15}$$

where C_2^n represents the various combinations that may be obtained for n observations taken 2 at a time. In our case, n is the number of data points. However,

n_{ij} is the number of observations that are in group i in the case of C_1 and in group j in the case of C_2. The Rand indices for different distance functions and features for the union and intersection decompositions are presented in the next section.

4.4 Evaluation of Distance Functions

In order to select the best distance function, we need to evaluate the effectiveness of each in differentiating between the various Meningioma subtypes. The stability of the decomposition is also one of the factors that determines the viability of a distance metric. In order to evaluate each distance metric, we acquire features from the subbands in each decomposition and use k-Means clustering to determine how good a certain distance function is. The first stage is the acquisition of statistical features per subband and then k-Means clustering of the features is carried out. Rand indices are provided to indicate the viability of a distance function for Meningioma subtypes classification.

The results of Rand index analysis are given in Tables 4.1-4.4. As indicated, the Rand values for each feature were obtained separately and the clustering was carried out using k-Means clustering. Although k-Means is a very basic and simple clustering method, it nevertheless produces very good results for our data. As can be seen from Tables 4.1- 4.4, the intersection of decompositions performs slightly better than union for both standard averaging and pseudo-averaging. Especially in the case of pseudo-averaging, it can be seen that the difference in Rand indices for union and intersection is relatively high. The steps followed for acquiring the results are:

1. Intersection and union of ADWPT decompositions is obtained for each image in the data set (Note: Different decompositions are obtained for different distance functions).

2. Statistical features for the intersection and union ADWPT decompositions are extracted.

3. k-Means clustering of the feature sets is carried out.

4. Subsequently Rand indices are computed for each feature set and for each distance function separately.

Tables 4.1- 4.4 show the Rand indices indicating the accuracy of the clusters produced using the decompositions obtained with various distance functions over standard averaging mode of texture template construction. The Highest Rand Index (HRI) produced for each distance function is shown in bold. The Average Rand Indices (ARIs) are shown under $\mathcal{R}^{\mu}$ which indicates the overall clustering efficiency for the various types of features.

4.4.1 Results: Standard Averaging

Rand indices computed for standard averaging based ADWPT decompositions are shown in the Tables 4.1 and 4.2. The results are quite encouraging as relatively high values for Rand indices are obtained.

Table 4.1: Rand Indices for various features of **Intersection** decomposition obtained using **Standard averaging** and various distance functions (HLD=Hellinger Distance, K-LD=Kullback-Leibler Distance, FD=Fishers Linear Discriminant, J-SD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance). $\mathcal{R}^{\mu}$=Average Rand Index

	μ	σ	ν	Γ	ξ	ε	$\mathcal{G}_{\varsigma}$	$\mathcal{G}_{\upsilon}$	$\mathcal{G}_{\varepsilon}$	$\mathcal{G}_{\tau}$	$\mathcal{R}^{\mu}$
HLD	0.63	0.62	0.63	0.63	0.62	0.78	0.74	**0.81**	0.74	0.79	**0.70**
K-LD	0.58	0.62	0.63	0.63	0.63	0.77	0.74	**0.81**	0.73	0.79	0.69
FD	0.62	0.63	0.63	0.61	0.63	0.72	0.72	**0.75**	0.71	0.75	0.68
J-SD	0.58	0.62	0.63	0.63	0.63	0.77	0.77	**0.81**	0.73	0.79	**0.70**
BD	0.63	0.62	0.63	0.63	0.62	0.75	0.73	**0.82**	0.71	0.76	0.69
MD	0.62	0.59	0.63	0.61	0.59	0.68	0.70	0.76	0.72	**0.81**	0.67
MCMD	0.62	0.55	0.50	0.60	0.55	0.65	0.70	**0.78**	0.67	0.73	0.64
EneD	0.63	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.74	0.67
ReneD	0.58	0.47	0.62	0.64	0.60	0.73	0.72	**0.82**	0.67	0.77	0.66
Avg	0.61	0.60	0.61	0.63	0.61	0.72	0.73	**0.79**	0.70	0.77	0.68

Table 4.2: Rand Indices for various features of **Union** decomposition obtained using **Standard averaging** and various distance functions (HLD=Hellinger Distance, K-LD=Kullback-Leibler Distance, FD=Fishers Linear Discriminant, J-SD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance). $\mathcal{R}^{\mu}$=Average Rand Index

	μ	σ	ν	Γ	ξ	ε	$\mathcal{G}_{\varsigma}$	$\mathcal{G}_{\upsilon}$	$\mathcal{G}_{\varepsilon}$	$\mathcal{G}_{\tau}$	$\mathcal{R}^{\mu}$
HLD	0.62	0.62	0.62	0.64	0.62	0.67	0.73	**0.8**	0.72	0.76	**0.68**
K-LD	0.58	0.64	0.61	0.64	0.64	0.67	0.72	**0.81**	0.72	0.76	**0.68**
FD	0.63	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.74	0.67
JS	0.56	0.64	0.6	0.64	0.64	0.67	0.72	**0.81**	0.72	0.76	**0.68**
BD	0.62	0.64	0.6	0.65	0.65	0.73	0.72	**0.75**	0.68	0.75	**0.68**
MD	0.63	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.74	0.67
MCMD	0.62	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.75	0.67
EneD	0.62	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.74	0.67
ReneD	0.63	0.62	0.61	0.65	0.62	0.68	0.72	**0.75**	0.71	0.75	0.67
Avg	0.61	0.63	0.61	0.64	0.63	0.69	0.72	**0.77**	0.70	0.74	0.67

As seen from the tables the intersection of ADWPT decompositions $B^*_{\bigcap}$ produces slightly better results than union of ADWPT decompositions $B^*_{\bigcup}$. The main advantage of the intersection of ADWPT decompositions is that it reduces the length of the feature set greatly. We will also see in the next chapter that $B^*_{\bigcap}$ is a superior method for selection of subbands. The highest ARI is produced by Hellinger and Jensen-Shannon distance which is 0.70 while the highest Rand value of 0.82 is obtained for the GLCM correlation feature for the Bhattacharyya and Relative Energy distance based decompositions. Kullback-Leibler also produces relatively good Rand indices. Comparative analysis was made with LBP features. Two LBP based feature sets are acquired with neighbours 8 and radius 1 and neighbours 16 and radius 2. The Rand indices obtained after k-means clustering were 0.69 and 0.65 respectively.

On the other hand, in the case of $B^*_{\bigcup}$, the better performing distance functions are Hellinger, Kullback-Leibler, Jensen-Shannon and Bhattacharyya (as shown in Table 4.2). The highest ARI produced for $B^*_{\bigcup}$ is 0.68 which is lower than the highest ARI of 0.70 produced for intersection of decompositions. From the results, for union and intersection, it can be seen that some distance functions perform better than others. After careful analysis of the Tables 4.1 and 4.2, it can be seen that the Energy distance and Mahalanobis distance measures (in both modes i.e. union and intersection) consistently perform worse than Hellinger, Kullback-Leibler, Jensen-Shannon and Bhattacharyya distance functions.

It may also be observed that the difference between the various ARI's overall remains low but the difference between HRI's is high. It would be important here to discuss the various decompositions and how subband selection affects the overall clustering efficiency. It must be noted here that we would be referring only to the intersection decompositions in our discussion as they produce better Rand indices. Figure 4.3 shows the intersection decomposition obtained for the various distance functions. It is clear from Figure 4.3 (a-e) that the decompositions with a high value of $\mathcal{R}^{\mu}$ decompose subbands in the approximation, horizontal and vertical details i.e. the descendants of subbands $\mathcal{W}_{1,0,0}$, $\mathcal{W}_{1,1,0}$ and $\mathcal{W}_{1,0,1}$. The more discriminant subbands are found usually amongst the descendants of $\mathcal{W}_{1,1,0}$ and $\mathcal{W}_{1,0,1}$. The only exception is the Relative Energy distance for which the most discriminant subband is found in the descendants of the approximation subband

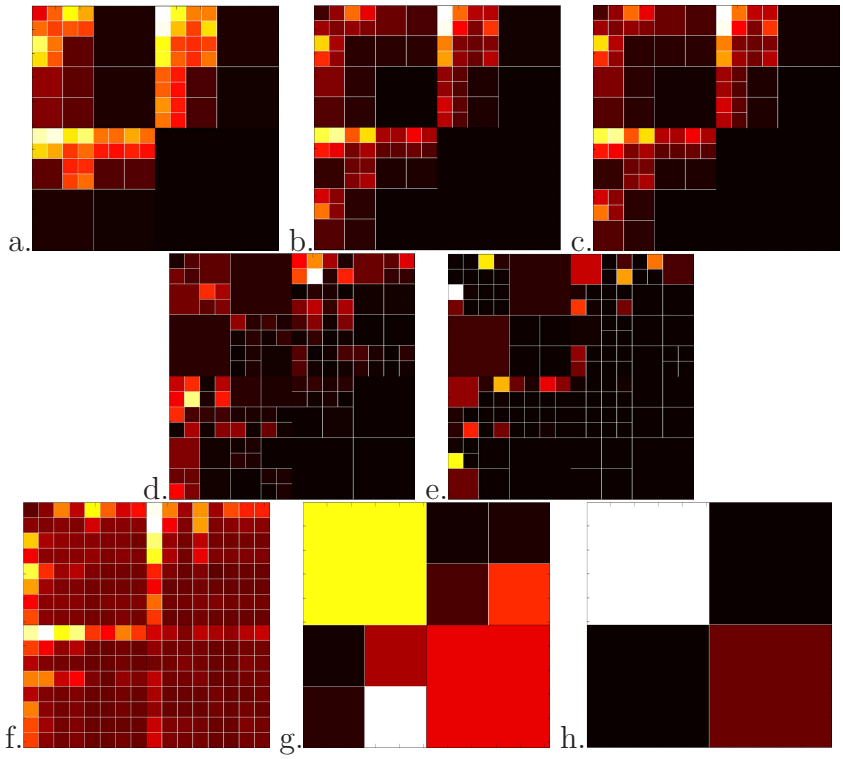


Figure 4.3: Standard averaging based intersection decompositions obtained for the distance functions a. Hellinger, b. Kullback-Leibler, c. Jensen-Shannon, d. Bhattacharyya, e. Relative Energy, f. Energy, g. Mahalanobis and h. Multiple-class Mahalanobis

$\mathcal{W}_{1,0,0}$. This is due to the fact that high energy is found in the descendants of $\mathcal{W}_{1,0,0}$ and since relative energy compares energies of two subbands, therefore more discriminant subbands are selected in the descendants of the more energetic subband. On the other hand, we can see that Energy distance does not do as well because it decomposes all subbands rendering the feature set too long and hence, adds features to the feature set which may not be the best for differentiating between Meningioma subtypes.

Mahalanobis distance fails to decompose subbands beyond 2 levels and hence fails to capture the intrinsic textural characteristics. Multi-class Mahalanobis is even worse as it fails to go beyond the first level. Multi-class Mahalanobis yields a feature set which is quite small and as the results indicate does not produce

desirable results. Hence, it can safely be deduced that subband selection in the descendants of the approximation subband $\mathcal{W}_{1,0,0}$, the vertical $\mathcal{W}_{1,1,0}$ and the horizontal detail subbands $\mathcal{W}_{1,0,1}$, is good for clustering as it keeps the feature set at a sufficient length and also captures the most useful discriminant textural characteristics. A FWPT decomposition as seen in the case of Energy distance is undesirable as it increases the feature set length and does not capture the more discriminant information. The Mahalanobis distance selects fewer subbands, in fact, the decomposition obtained is only up to the 2nd level which is not good for capturing textural characteristics as the Rand index indicates. The same applies to multi-class Mahalanobis which decomposes subbands only up to the 1st level. Next we discuss the results for the pseudo-averaging case.

4.4.2 Results: Pseudo Averaging

The pseudo averaging based approach produces results comparable to standard averaging. Tables 4.3 and 4.4 show the Rand indices obtained for various statistical features discussed previously. The highest ARI for pseudo-averaging is obtained in the case of intersection of decompositions for the Fisher discriminant as can be seen in Table 4.3. Fisher distance overall performs better than any other distance function for pseudo-averaging with an ARI of 0.70. The highest overall Rand value for the union decompositions is 0.69 which is slightly lower than the one obtained for intersection. The highest HRI obtained is 0.81 for Fisher Discriminant (correlation feature) which is much better than the highest Rand Index for the k-means clusters of LBP based features of 0.69. Intersection is again seen to perform better than union with an ARI of 0.70 compared to 0.69 obtained with union while the HRI produced with intersection is 0.81, much better than the union HRI of 0.77.

Like standard averaging Hellinger (0.69), Kullback-Leibler (0.69) and Jensen-Shannon (0.69) also produce good results in the case of pseudo averaging for $B^*_{\bigcap}$. The only difference is Fisher Discriminant which produces much better results for pseudo averaging (ARI=0.70) compared to those for standard averaging (ARI=0.67) while Bhattacharyya does better in the case of standard averaging (ARI of 0.69 compared to 0.63 for pseudo averaging). Relative energy performed

Table 4.3: Rand Indices for various features of **Intersection** decompositions obtained using **pseudo-averaging** and various distance functions (HLD=Hellinger Distance, K-LD=Kullback-Leibler Distance, FD=Fishers Linear Discriminant, J-SD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance). $\mathcal{R}^{\mu}$=Average Rand Index

	μ	σ	ν	Γ	ξ	ε	$\mathcal{G}_{\varsigma}$	$\mathcal{G}_{\upsilon}$	$\mathcal{G}_{\varepsilon}$	$\mathcal{G}_{\tau}$	$\mathcal{R}^{\mu}$
HLD	0.62	0.63	0.62	0.60	0.63	0.74	0.73	0.77	0.78	**0.80**	0.69
K-LD	0.62	0.63	0.62	0.60	0.63	0.74	0.74	0.77	0.75	**0.80**	0.69
FD	0.62	0.68	0.62	0.63	0.68	0.72	0.73	**0.81**	0.71	0.75	**0.70**
J-SD	0.63	0.63	0.62	0.60	0.63	0.74	0.73	0.77	0.75	**0.80**	0.69
BD	0.62	0.48	0.61	0.63	0.46	0.65	0.66	**0.76**	0.68	0.75	0.63
MD	0.62	0.60	0.62	0.60	0.60	0.72	0.72	0.78	0.72	**0.8**	0.68
MCMD	0.63	0.51	0.50	0.60	0.51	0.65	0.70	**0.78**	0.67	0.73	0.63
EneD	0.61	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.74	0.67
REneD	0.62	0.46	0.58	0.63	0.48	0.63	0.66	0.73	0.68	**0.76**	0.62
Avg	0.62	0.59	0.60	0.62	0.59	0.70	0.71	**0.77**	0.71	0.77	0.67

better in the case of standard averaging (HRI=0.82) but is seen to produce much worse results for pseudo-averaging (HRI=0.76). Figure 4.4 shows the intersection decomposition obtained for some of the distance functions.

Table 4.4: Rand Indices for various features of **Union** decompositions obtained using **pseudo-averaging** and various distance functions (HLD=Hellinger Distance, K-LD=Kullback-Leibler Distance, FD=Fishers Linear Discriminant, J-SD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance). $\mathcal{R}^{\mu}$=Average Rand Index

	μ	σ	ν	Γ	ξ	ε	$\mathcal{G}_{\varsigma}$	$\mathcal{G}_{\upsilon}$	$\mathcal{G}_{\varepsilon}$	$\mathcal{G}_{\tau}$	$\mathcal{R}^{\mu}$
HLD	0.58	0.62	0.61	0.64	0.62	0.68	0.73	0.75	0.72	**0.76**	0.67
K-LD	0.63	0.62	0.6	0.64	0.62	0.68	0.73	0.75	0.72	**0.76**	0.68
FD	0.63	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.74	0.67
J-SD	0.58	0.63	0.6	0.65	0.63	0.73	0.73	0.75	0.72	**0.76**	0.68
BD	0.63	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.74	0.67
MD	0.63	0.65	0.62	0.64	0.65	0.73	0.72	**0.77**	0.71	0.75	**0.69**
MCMD	0.62	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.75	0.67
EneD	0.62	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.75	0.67
REneD	0.63	0.63	0.61	0.64	0.63	0.68	0.72	**0.75**	0.68	0.74	0.67
Avg	0.62	0.63	0.61	0.64	0.63	0.69	0.72	**0.75**	0.70	0.75	0.67

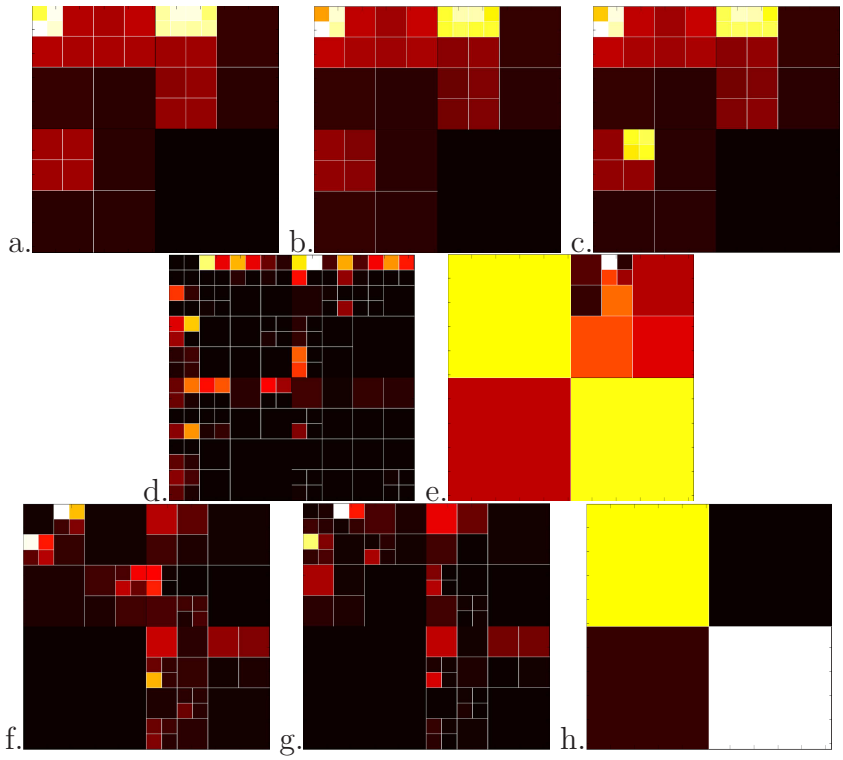


Figure 4.4: Pseudo averaged intersection decompositions obtained for the distance functions a. Hellinger, b. Kullback-Leibler, c. Jensen-Shannon, d. Fishers Discriminant, e. Mahalanobis, f. Bhattacharyya, g. Relative Energy and h. Multiple-class Mahalanobis

Figure 4.4(a)-(e) shows decompositions belonging to distance functions that perform comparably better in the case of pseudo-averaging while (f)-(h) produce lower clustering accuracies. It can be clearly seen from Figure 4.4 that the subbands $\mathcal{W}_{4,10,0}$, $\mathcal{W}_{4,11,0}$, $\mathcal{W}_{4,10,1}$ and $\mathcal{W}_{4,11,1}$ are the most critical subbands (this is true for both standard and pseudo-averaging). There inclusion in the Mahalanobis distance decompositions causes the clustering accuracy to rise (ARI=0.68 and HRI=0.80) while it can be clearly seen that the decompositions which do not include these subbands shown in Figure 4.4(f)-(h) do not perform well. On the other hand, the best clustering accuracies are obtained for the Fishers distance which includes subbands that are descendants of the approximation $\mathcal{W}_{1,0,0}$, the horizontal $\mathcal{W}_{1,1,0}$ and the vertical $\mathcal{W}_{1,0,1}$ detail subbands. The diagonal detail subbands decomposed i.e. the descendants of $\mathcal{W}_{1,1,1}$ do not seem to be too important as their inclusion or exclusion has little impact on clustering accuracies but in classification as we shall see in the next chapter these subbands can improve the classification accuracy.

4.4.3 Discussion

It can be safely concluded from the above analysis that standard averaging overall performs slightly better than pseudo-averaging in terms of clustering efficiency. Although the difference remains small with standard averaging producing the best HRI of 0.82 while pseudo-averaging producing a best HRI of 0.81 but the highest ARI for both remains equal to 0.70. However, different distance functions perform better for the two different types of averaging. Intersection is generally found to be better than the union of decompositions. Indeed, this trend can be seen for most of the distance functions with there being very few instances when the union performs better than intersection as seen in the case of Hellinger for standard averaging. The union decomposition represents the universal set or the maximum set whereas the intersection represents the minimum set. The intersection places a very stringent constraint that the subband included must be found in all the 625 decompositions obtained. From the above discussion, we conclude that there are two factors that affect the clustering accuracy obtained using a decomposition.

1. The frequency content of the subbands selected or which descendants of the approximation ($\mathcal{W}_{1,0,0}$) and detail ($\mathcal{W}_{1,1,0}$, $\mathcal{W}_{1,0,1}$, $\mathcal{W}_{1,1,1}$) subbands are selected.

2. Number of subbands selected.

From the above results, we can see that subbands that are the descendants of the approximation subband $\mathcal{W}_{1,0,0}$, vertical $\mathcal{W}_{1,1,0}$ and horizontal $\mathcal{W}_{1,0,1}$ detail subbands contains information that are effective for efficient clustering with subbands indices $\mathcal{W}_{4,10,0}$, $\mathcal{W}_{4,11,0}$, $\mathcal{W}_{4,10,1}$ and $\mathcal{W}_{4,11,1}$, identified as the most critical subbands.

The other important factor is the number of subbands selected. Next we would discuss, how number of subbands selected affects clustering accuracies. However, its extremely important that the subbands selected represent the spatial-frequency resolution which is best for differentiating between Meningioma subtypes. The discrimination power of a subband computed using a viable distance function would indicate how useful a certain subband is. Considering this fact, the more discriminant the subbands in a decomposition, the better it is for classification and clustering. We shall see in the classification results in the next chapter that usually a high number of subbands are better than a lower number provided they represent the right spatial frequency information. Subsequent discussion would aim to ascertain which number of subbands provide higher classification accuracies.

In this discussion, we would be referring to intersection decomposition as it provides better results for clustering. Our analysis shows that the best results depend upon the number of subbands selected and more importantly upon the frequency content of the subbands selected. Fisher discriminant produces the best results for pseudo-averaging. A total of 139 subbands are found in the intersection. While in the case of Bhattacharyya distance, intersection of decomposition produces a total of 58 subbands with a low ARI of 0.63. On the other hand Mahalanobis distance produces a relatively good ARI of 0.68 with just 13 subbands since the subbands selected are better for classification in comparison to those selected by Bhattacharyya, as explained earlier. Tables 4.5 and 4.6 show the number of subbands and the ARIs and HRIs obtained for each distance function.

Table 4.5: Number of subbands and Rand indices for **Standard Averaging** (HLD=Hellinger Distance, KLD=Kullback-Leibler Distance, FD=Fishers Linear Discriminant, JSD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance) (All values are correct to 2 decimal places)

	N_s	HRI	ARI
HLD	73	0.81	0.70
KLD	82	0.81	0.69
FD	178	0.75	0.68
JSD	82	0.81	0.70
BD	130	0.82	0.69
MD	10	0.81	0.67
MCMD	4	0.78	0.63
EneD	256	0.75	0.67
REneD	124	0.82	0.66

Table 4.5 shows the number of subbands and the ARIs and HRIs obtained for the various distance functions included in the study for the standard averaging case. In Table 4.5, it can be seen that the lowest ARI is obtained for the multi-class Mahalanobis distance. The number of subbands selected are quite low i.e. only four subbands are selected which is equivalent to a first level simple wavelet transform. Two class Mahalanobis distance is also not good for subband selection as it selects only 10 subbands and produces a comparatively low ARI of 0.67. The other distance for which a relatively low ARI is acquired is the Energy distance function. The ADWPT decompositions obtained for Energy distance functions are equivalent to a Full Wavelet Packet Transform (FWPT). As discussed in Chapter 2 and as per the results in Table 4.5 FWPT is not the best decomposition for Meningioma subtypes classification. The results indicate that a very low number of subbands (i.e. under 10 subbands) and a very high number of subbands such as 256 are not ideal for differentiating between Meningioma subtypes.

Again referring to the Table 4.5, the HRIs indicate that Bhattacharyya and the novel Relative Energy distance are the best distances for clustering Meningioma subtypes. Bhattacharyya and Relative Energy based decompositions produce 130 and 124 subbands respectively. Hence, subbands number in this range are good for differentiating between Meningioma subtypes provided the right spatial and frequency resolution subbands are selected. Mahalanobis distance performs well although only 10 subbands are selected but Fisher discriminant does not produce a too high an HRI although 178 subbands are selected. This is due to the non-selection of the right spatial-frequency resolution subbands as discussed earlier.

Table 4.6: Number of subbands and Rand indices for **Pseudo-averaging** (HLD=Hellinger Distance, KLD=Kullback-Leibler Distance, FD=Fishers Linear Discriminant, JSD=Jensen-Shannon Distance, ED=Euclidean Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MCMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance)

	N_s	HRI	ARI
HLD	37	0.80	0.69
K-LD	37	0.80	0.69
FD	139	0.81	0.69
J-SD	40	0.80	0.69
BD	58	0.76	0.63
MD	13	0.80	0.68
MCMD	4	0.78	0.63
EneD	256	0.75	0.67
REneD	64	0.76	0.62

Table 4.6 shows the HRIs and ARIs for the pseudo-averaging case for various distance functions. Hellinger, Kullback-Leibler, Fisher and Jensen-Shannon distances produce comparable results with Fisher being the best with an HRI of 0.81. Very few subbands, as obtained in the case of Multi-class Mahalanobis produce discouraging results which is consistent with the results for standard averaging. Moreover, 256 subbands selected in the case of Energy distance are

again found to be undesirable. Mahalanobis distance produces a good HRI for just 13 subbands as the subbands selected are good for clustering accuracy.

In the light of above discussion, we can conclude that generally more subbands are better than fewer subbands with the optimum lying somewhere in the range of 10 to 190, provided the right descendants of the approximation and detail subbands are selected. If more subbands are available but they do not represent the right frequency and spatial information then high accuracy results will not be obtained. If the right spatial and frequency information is selected then subbands as few as 13 produce high clustering efficiency results as seen for Mahalanobis distance (pseudo-averaging case). Therefore, selection of the right spatial-frequency resolutions represented by a subband is more important than the number of subbands selected. Hence, it can be safely concluded that we are interested in decompositions with 4 levels of decomposition in the vertical $\mathcal{W}_{1,0,1}$, horizontal $\mathcal{W}_{1,1,0}$ and also in the approximation $\mathcal{W}_{1,0,0}$ subband of a wavelet packet decomposition as these subbands are important for acquiring high clustering accuracies with subbands $\mathcal{W}_{4,10,0}$, $\mathcal{W}_{4,11,0}$, $\mathcal{W}_{4,10,1}$ and $\mathcal{W}_{4,11,1}$ being the most critical.

The above discussion describes how different distance functions produce different clustering results as different decompositions are acquired using them. In order to determine which discriminant functions are good and which are not, we develop a system so that we could accept or reject a distance function based upon its performance. Tables 4.7 and 4.8 present the results obtained taking into account factors such as stability $(\mathcal{S}_1, \mathcal{S}_2)$ defined in Chapter 3, maximum Rand index i.e. the HRI $(\mathcal{R}^*)$ obtained for a distance function, ARI $(\mathcal{R}^\mu)$ and subbands viability function (κ) which is defined as,

$$\mathcal{V} = 8\frac{N_s}{4^L}(1 - \frac{N_s}{4^L}) \tag{4.16}$$

$$\kappa = \begin{cases} \mathcal{V} & \text{if } \mathcal{V} < 1 \\ 1 & \text{otherwise} \end{cases}$$

where N_s is the number of subbands produced in the intersection of all decomposition for a given distance function and L is the level to which the transform is carried out. κ tends to 0 when the number of subbands are low (such as less than 10) or when the number of subbands are high i.e. 256 which is equivalent

to the number of subbands produced in FWPT. This is due to the fact that both these contingencies are undesirable as discussed previously. A very low number of subbands (less than 10) and a high number of subbands approaching or equal to the value 4^L have been shown to produce low clustering accuracies. Figure 4.5 shows how the κ function behaves in response to the number of subbands. As stated earlier, very low number of subbands and very high that is approximating the maximum at that decomposition level are not good for our problem and hence have low κ values.

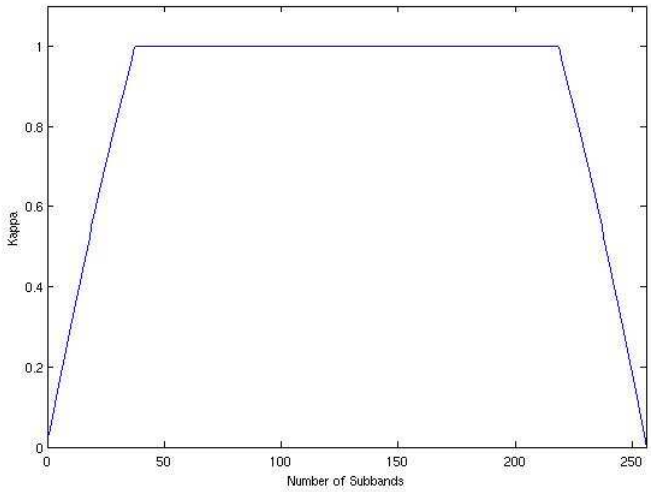


Figure 4.5: A graph showing the shape of the Kappa function

The overall measure of stability and clustering accuracy for a distance function is given by $\mathcal{M}$,

$$\mathcal{M} = \frac{\kappa\mathcal{S}\mathcal{R}^* + \kappa\mathcal{S}\mathcal{R}^\mu + \kappa\frac{\mathcal{R}^\mu}{\mathcal{R}^*}\mathcal{S}}{3} \tag{4.17}$$

The function describes the usefulness of a distance function in terms of its clustering efficiency and decomposition stability. Each factor of the above equation represents characteristics desirable in a discriminant function i.e. high decomposition stability $\mathcal{S}$ and good clustering efficiency $\mathcal{R}^*, \mathcal{R}^\mu$. $\kappa\mathcal{S}\mathcal{R}^*$ computes a relatively direct measure of clustering efficiency by taking into account the HRI obtained for a given function while $\kappa\mathcal{S}\mathcal{R}^\mu$ represents a viability measure more dependent

upon the ARI produced by the distance function. On the other hand, $\kappa\frac{\mathcal{R}^{\mu}}{\mathcal{R}^{*}}\mathcal{S}$ relies on a more important metric of overall clustering efficiency which relates ARI to the HRI produced. This is one of the most important factors in deciding the viability of a distance function. Each of these factors reflect the usefulness of a distance function based upon different clustering efficiency indices and the stability measure acquired. All of these are averaged to produce an overall distance function viability measure $\mathcal{M}$.

Table 4.7: Performance evaluation of the various distance functions for **Standard Averaging** (HLD=Hellinger Distance, KLD=Kullback-Leibler Distance, FD=Fishers Linear Discriminant, JSD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance) (All values are correct to 2 decimal places)

	$\mathcal{S}$	$\mathcal{R}^{*}$	$\mathcal{R}^{\mu}$	N_s	κ	$F(\mathcal{S}\mathcal{R}^{*})$	$F(\mathcal{S}\mathcal{R}^{\mu})$	$F(\mathcal{S}\mathcal{R})$	$\mathcal{M}$
HLD	0.70	0.81	0.70	73	1	0.57	0.49	0.60	**0.55**
KLD	0.71	0.81	0.69	82	1	0.58	0.49	0.60	**0.56**
FD	0.70	0.75	0.68	178	1	0.52	0.47	0.63	**0.54**
JSD	0.71	0.81	0.70	82	1	0.57	0.49	0.61	**0.56**
BD	0.35	0.82	0.69	130	1	0.29	0.24	0.29	0.27
MD	0.05	0.81	0.67	10	0.30	0.01	0.01	0.01	0.01
MMD	0.01	0.78	0.63	4	0.12	0	0	0	0
EneD	1	0.75	0.67	256	0	0	0	0	0
REneD	0.58	0.82	0.66	124	1	0.48	0.38	0.47	0.44

Table 4.8: Performance evaluation of the various distance functions for Pseudo-Averaging (HLD=Hellinger Distance, KLD=Kullback-Leibler Distance, FD=Fishers Linear Discriminant, JSD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, MMD=Multi-Class Mahalanobis Distance, EneD=Energy Distance, REneD=Relative Energy Distance)

	$\mathbb{S}$	$\mathcal{R}^*$	$\mathcal{R}^\mu$	N_s	κ	$F(\mathbb{S}\mathcal{R}^*)$	$F(\mathbb{S}\mathcal{R}^\mu)$	$F(\mathbb{S}\mathcal{R})$	$\mathcal{M}$
HLD	0.75	0.80	0.69	37	0.99	0.60	0.52	0.64	**0.59**
K-LD	0.75	0.80	0.69	37	0.99	0.59	0.52	0.64	**0.58**
FD	0.86	0.81	0.69	139	1	0.70	0.59	0.73	**0.67**
J-SD	0.74	0.80	0.69	40	1	0.59	0.51	0.64	**0.58**
BD	0.80	0.76	0.63	58	1	0.61	0.50	0.66	**0.59**
MD	0.49	0.80	0.68	13	0.39	0.15	0.13	0.16	0.15
MCMD	0.45	0.78	0.63	4	0.12	0.04	0.04	0.04	0.04
EneD	1	0.75	0.67	256	0	0	0	0	0
REneD	0.77	0.76	0.62	64	1	0.58	0.47	0.62	**0.56**

Tables 4.7 and 4.8 show the values for $\mathcal{M}$ obtained for various distance functions for the intersection of decompositions. The discriminant functions selected are shown in bold. The criterion for the selection of a distance function is based upon the measure $\mathcal{M}$ and is given as,

$$\mathcal{M} > \top \tag{4.18}$$

where $\top = 0.5$, which provides a good cut off point. It can be clearly seen from the tables that the other distance functions that suffer as per there stability or classification accuracy produce values which are relatively low in comparison to distance functions found good in our analysis. Most functions produce values for $\mathcal{M}$ sufficiently below the the threshold 0.5 except for the functions found good in our analysis which are safely above the threshold. Distance functions not selected do not provide any interesting or better results in terms of stability and clustering efficiency except for Bhattacharyya distance in case of pseudo-averaging which produces high clustering accuracies. Although Bhattacharyya distance is good for clustering but its stability is low, hence, it is not selected. The maximum value that can be attained for $\mathcal{M}$ is 1. A distance function for which $\mathcal{M}$ is equal to 1 would be the perfect distance function. This represents the theoretical limit.

Finally, we have come to the point where the most useful distance functions have been found. The functions that perform better in the case of pseudo-averaging are: Hellinger, Kullback-Leibler, Fishers Linear, Jensen-Shannon, Bhattacharyya and Relative Energy distance. Hence, there will be a total of 6 functions which would be analysed further in the rest of the thesis for pseudo-averaging. In case of standard averaging, the distance functions selected are: Hellinger, Kullback-Leibler, Fishers Discriminant and Jensen-Shannon distance. But it must be noted that Kullback-Leibler is equivalent to Jensen-Shannon in case of standard averaging whereas Kullback-Leibler produces the same decomposition as Hellinger in case of pseudo-averaging. In the next chapter, a comparison between the various distance functions for pseudo-averaging and standard averaging case are presented with classification results. We will include Relative Energy distance and Bhattacharyya distance for standard averaging in our analysis as they produce a relatively high value for $\mathcal{M}$ in case of pseudo-averaging and also

they produce the highest HRIs for standard averaging. They are included so that a comparison can be made between the results produced for standard and pseudo-averaging. The distance functions selected have been found useful for Meningioma subtypes classification. An analysis of this sort may be carried out for any other application and the best distance measures may be selected.

4.5 Summary and Conclusions

In this chapter, an analysis of the stability of ADWPT decompositions with respect to different distance functions is presented. 10 distance functions each for the two forms of pseudo averaging are compared. An objective criterion is proposed for selection of distance functions based upon stability and clustering efficiency. The viability of a distance function is defined in terms of the stability index and the clustering accuracy attained. 6 distance functions are selected for further analysis.

In terms of the averaging mechanism, it can be said that pseudo-averaging is relatively more stable than standard averaging but standard averaging produces slightly better clustering results. Therefore, its important that the classification results for both forms of averaging for the distance functions that perform better are presented. In the next chapter, we present the classification results and select the distance functions and feature type that is best for Meningioma subtype classification.

Chapter 5

Feature Selection for Subtype Classification

As stated earlier, the aim of our research is to differentiate amongst the different meningioma subtypes under study, i.e. Meningiotheliamatous, Fibroblastic, Transitional and Psammomatous. In this chapter, first the classification of feature sets obtained using ADWPT is carried out to investigate the efficacy of ADWPT features in differentiating between the various meningioma subtypes. The distance function that performs best in classifying amongst the various meningioma subtypes is identified. Subsequently, we show how the stability and the discrimination power of a subband impacts classification accuracy. An optimal subset of subband features is selected based upon subband discrimination power and stability of the subband that provide the best classification accuracies with almost perfect results in some instances.

Section 5.1 discusses the classification results and presents the best distance functions for meningioma subtype classification. Section 5.2 presents a feature selection approach based upon subband stability and subband discrimination power and shows that better classification accuracies are obtained with more stable and discriminant subbands.

5.1 Classification

The aim of our study has been principally to develop a technique to accurately differentiate between meningioma subtypes. After the acquisition of decompositions, features for each subband were computed and then used for clustering in the previous chapter. It was concluded from the clustering results that GLCM features generally provide better clusterings. In our work with classification of meningioma subtype images, we experimented with various features including the Grey Tone Difference Matrix features and the first order statistical features [151]. We found GLCM features to be better than all other features for classification. Hence, more GLCM features were included in the study. The GLCM features that were added for classification were GLCM Mean, Standard Deviation, Dissimilarity, Shade and Prominence as described by Haralick *et al.* [135].

GLCM Mean

The mean of a normalised GLCM matrix $Q_{d,p,q}^{\triangle x,\triangle y}$ may be computed as:

$$\mathcal{G}_{d,p,q}^{\mu} = \sum_{i,j=0}^{N-1} iQ_{d,p,q}^{\triangle x,\triangle y}(i,j) \tag{5.1}$$

The GLCM mean describes how the frequency content varies over a subband. A high value for mean would signal presence of high values or energy in texture represented by the subband whereas a low mean would correspond to low energy.

GLCM Standard Deviation

The standard deviation of a normalised GLCM matrix $Q_{d,p,q}^{\triangle x,\triangle y}$ may be computed as:

$$\mathcal{G}_{d,p,q}^{\sigma} = \sqrt{\sum_{i,j=0}^{N-1}(i-\mathcal{G}_{d,p,q}^{\mu})^2 Q_{d,p,q}^{\triangle x,\triangle y}(i,j)} \tag{5.2}$$

The GLCM Standard Deviation provides a measure of the variance in coefficient values over a subband i.e. how the coefficient values vary in relevance to each other.

GLCM Dissimilarity

The Dissimilarity feature of a normalised GLCM matrix $Q_{d,p,q}^{\triangle x,\triangle y}$ may be computed as:

$$\mathcal{G}_{d,p,q}^{\vartheta} = \sqrt{\sum_{i,j=0}^{N-1} |i-j| Q_{d,p,q}^{\triangle x,\triangle y}(i,j)} \tag{5.3}$$

The GLCM Dissimilarity is related to the contrast feature which provides a measure of the difference or the gradient in coefficient values over a subband. The only difference is that in computation of contrast the weights increase exponentially while in dissimilarity they increase linearly.

GLCM Shade

The Shade feature of a normalised GLCM matrix $Q_{d,p,q}^{\triangle x,\triangle y}$ may be computed as:

$$\mathcal{G}_{d,p,q}^{\Im} = sgn(A)|A|^{\frac{1}{3}} \tag{5.4}$$

where

$$A = \sum_{i,j=0}^{N-1} \frac{(i+j-2\mathcal{G}_{d,p,q}^{\mu})^3 Q_{d,p,q}^{\triangle x,\triangle y}(i,j)}{{\mathcal{G}_{d,p,q}^{\sigma}}^3(\sqrt{2(1+\mathcal{G}_{d,p,q}^{\upsilon})^3})} \tag{5.5}$$

and $sgn(A)$ denotes the sign on the value A. The GLCM Shade is related to the correlation feature which determines the relationship in terms of how the coefficients values vary over a subband. Shade feature as the name suggests tries to determine if the texture represented by a subband contains shading or darker regions in terms of how the coefficients are correlated.

GLCM Prominence

The Prominence feature of a normalised GLCM matrix $Q_{d,p,q}^{\triangle x,\triangle y}$ may be computed as:

$$\mathcal{G}_{d,p,q}^{\Im} = sgn(B)|B|^{\frac{1}{4}} \tag{5.6}$$

where

$$B = \sum_{i,j=0}^{N-1} \frac{(i+j-2\mathcal{G}_{d,p,q}^{\mu})^4 Q_{d,p,q}^{\triangle x,\triangle y}(i,j)}{4{\mathcal{G}_{d,p,q}^{\sigma}}^4(1+\mathcal{G}_{d,p,q}^{\upsilon})^2)} \tag{5.7}$$

and $sgn(B)$ denotes the sign on the value B i.e. either positive or negative. The GLCM Prominence is also related to the correlation feature ($\mathcal{G}_{d,p,q}^{\upsilon}$) which determines the relationship in terms of how the coefficients texture varies over a subband.

After feature extraction, the next step in any texture classification problem is classifier selection. Many classifiers have been used for texture classification. We experimented with a few classifiers and found Support Vector Machines (SVMs) to be the best for meningioma subtype classification using ADWPT based features. A discussion of various classifiers is presented next and the results obtained with each kind of classifier are discussed as well.

5.1.1 Classifiers

In any pattern classification problem, the choice of classifier is one of the critical aspects. The other important factor is feature selection. The success of a pattern recognition algorithm depends upon feature and classifier selection. In the previous chapter, we have shown that the features extracted using ADWPT and GLCM can produce good clusters. Rand indices as high as 0.82 were obtained for some of the features. This indicates that the features obtained are capable of differentiating between the various textures under study. Since our aim is classification, we must select an appropriate classifier. In this section, we discuss the choice of classifiers and present the results obtained with the one that produces the best results.

Over the last few decades, great progress has been made in the domain of classifiers. Many new classifiers have been invented. There are various types of classifiers. Mazhelis in [158] presents an overview of various types of classifiers and an analysis of the strengths and weaknesses of the classifiers in relation to a problem.

There are various types of binary-class and multi-class classifiers in use. A binary class classifier principally says whether a sample belongs to a class or not

i.e. it differentiates between two classes. Many binary-class classifiers can be combined to produce a multi-class classifier. A multi-class classifier as the name suggests classifies a test sample amongst a number of classes at the same time. In the process of classification, normally a process of training is involved with a number of training samples and subsequently test samples are used to test the classification accuracy of a classifier.

Types of Classifiers

The various classifiers whether one-class or multi-class can be categorised into one of the following based upon the internal model used by the classifier:

Density Methods

k-Nearest Neighbour (k-NN) is an example of a density based classifier. Other examples of such classifiers are Gaussian or Gaussian mixture based classifiers. In density methods an estimate of the probability density function of the feature values in the training samples is estimated. The training data is assumed to be representative of the texture class true data distributions. If a distribution of certain features is unknown then it is assumed to be Gaussian or approximated using techniques such as mixtures of Gaussians.

We have performed some analysis with k-Nearest Neighbour (k-NN) and the results were presented in [52]. A leave-one-out approach was used for test-trial runs while the classification accuracies produced were not very high. However, the k-NN results showed that GLCM features combined with ADWPT can be used to effectively differentiate between meningioma subtypes. ADWPT based GLCM features were found to be better than raw image based GLCM features for meningioma subtype classification.

Reconstruction Methods

In the reconstruction method, an underlying model of the data structure or feature space is estimated. The model is learnt. The parameters of the model are

computed in the training phase and the sampling error is estimated at each stage. The training is stopped when an acceptable error-rate is attained. Examples of such a classifier are neural networks such as Self Organizing Maps (SOMs) and Learning Vector Quantization (LVQ) [159].

Lessmann *et al.* [67] in their work presented the results of using SOMs for clustering of the meningioma subtypes under study in this thesis. We used LVQs for the purpose of classifying meningioma subtypes and presented the results in [140]. Relatively good classification accuracies were obtained although GLCM features were not used. Though transitional and fibroblastic subtypes were not adequately classified with classification accuracies as low as 50%.

Boundary Methods

The boundary method stipulates the estimation of the distance between test feature vectors and the boundary built around the training feature vectors. The test samples are classified based upon the minimum distance to a boundary representing a class. The boundary construction is obtained during the training phase. Boundary methods are particularly aimed for one-class classification. Support Vector Machines (SVMs) are an example of boundary method based classifiers. Our analysis, in the subsequent section, is carried out with an SVM as it is one of the most powerful and versatile classifiers in use [158]. We will present a detailed description of SVMs and how they can be used to obtain a multi-class classifier, since, it is a one-class classifier.

Classification with various kinds of classifiers was carried out to provide an idea as to how various methods of classification function for our problem. This analysis was by no means aimed at presenting a comprehensive and exhaustive analysis of classifiers that may be used for classifying the various meningioma subtypes. The aim of this thesis is to attain a set of feature vectors that are good representations of textural patterns under study as the quality of features is extremely important. A weak classifier with good and optimal feature set can obtain good classification accuracies but a sophisticated and powerful classifier would fail if the quality of the feature set is substandard.

5.1.2 Support Vector Machines (SVMs) for Multi-Class Classification

As mentioned previously SVM is a supervised classifier which approximates the decision surfaces of the theoretical Bayes classifier. SVM has found broad area of applications since its invention in 1995 by Vapnik [160] and it is based upon statistical learning theory. Its application area ranges from medical image analysis [161] [162] to spam categorization [163] and face recognition [164]. SVM has also been used for text classification [165] and time series prediction [166]. Support Vector Machines (SVMs) have been employed frequently in the literature to perform pattern classification. In [161], Nattkemper *et al.* show that SVM performs better than decision trees and nearest neighbor (k-NN) classifiers in classification of breast tumors. It has been proven to be a reliable and efficient tool in wide variety of applications.

There are various advantages to using an SVM over other classifiers including the fact that it is computationally less intensive (especially in comparison to neural networks) and performs well in high dimensional spaces and when training data is limited. These are good reasons for using an SVM for histological image classification as the training data available is usually not too large.

SVM uses a kernel trick to map the input space in to a higher dimensional feature space to make the non-linear hyperplane linear. To achieve this without increasing computational complexity, a kernel trick is employed. The kernel function $K(x_i, x_j)$ computes an equivalent kernel value for the value in the input space so that no explicit mapping is required. A few popular kernels are presented as under:

$$Linear \quad : \quad < x_i, x_j > \tag{5.8}$$

$$Gaussian \quad : \quad e^{-\gamma \|x_i - x_j\|^2} \tag{5.9}$$

$$Polynomial \quad : \quad (\gamma < x_i, x_j > + a)^d \tag{5.10}$$

where $< x, y >$ represents a dot product, while x and y are two arbitrary feature vectors. As seen in the equation, each of the kernel functions contains some free parameters. In the Gaussian kernel, parameter γ denotes the width of the

Gaussian radial basis function. As for the polynomial kernel, d represents the degree of the polynomial while γ is the coefficient of the polynomial function and a is the coadditive constant.

In the training phase, the statistical features based on the gray level co-occurrence matrix of the subbands of the ADWPT decomposition are used as input feature sets. SVMs are trained using the feature sets extracted from the training data. The Matlab version of LibSVM developed by Chang and Lin [167] was used in our analysis. The best results were obtained with a Gaussian SVM.

It is important to note here that SVM is a two-class classifier i.e. it can distinguish between two classes at a time whereas in our application we are trying to differentiate between four texture classes at the same time. To acquire multi-class classification, multiple SVMs are combined using a voting method. Since SVM is a two-class or binary classifier, a voting based method is followed to achieve multi-class classification. $\frac{N_c(N_c-1)}{2}$ SVMs are trained where N_c corresponds to the number of classes included in the study. A "one-against-one" approach is followed where two classes are compared at a time and the sample is assigned to the class which has the most votes. If a tie in votes is found, then the class with the smallest index is chosen. This method is often referred to as the wrapper method in literature.

5.1.3 Distance Functions and Classification Accuracies

As stated earlier, SVM is used to differentiate amongst the four meningioma subtypes included in the study. We experimented with various types of SVMs which included the Linear, Gaussian and Polynomial. Gaussian was found to be the best kernel for our problem. A manual search is carried out to find the best parameter γ for each test trial run. This implies that a search for the best γ parameter is performed from a set of predefined values for γ. A range of values for the γ parameter are advised by the authors of LibSVM [167] that range from $2^{-}13$ to 2^5 with an interval of 2^2. We search for the best value of γ in this range.

The MAWTT templates using pseudo and standard averaging are obtained. Each MAWTT subband is then used to extract GLCM features. These features are then provided to the SVM for classification. 5 different test-trial runs are

carried out with a different patient's data left out from the subband decomposition and the MAWTT template construction phase each time. The overall 5 fold cross validated results are presented in the next section with a detailed discussion on the best results acquired.

5.1.4 Classification using SVM

In order to evaluate the various distance functions, we present the 5-fold cross validated overall results for distance functions selected in the previous chapter. The classification accuracies for each meningioma subtype are presented later. Table 5.1 presents the results acquired for each GLCM feature that has been used in the analysis.

Table 5.1: Overall classification accuracies of the various distance functions (HLD=Hellinger Distance, KLD=Kullback-Leibler Distance, FD=Fishers Linear Discriminant, JSD=Jensen-Shannon Distance, BD=Bhattacharya Distance, MD=Mahalanobis Distance, REneD=Relative Energy Distance) for Intersection of Decompositions ($\overline{p}$=Pseudo-averaging, $\overline{s}$=Standard averaging)

	$HLD_{\overline{p}}$	$HLD_{\overline{s}}$	$KLD_{\overline{p}}$	$KLD_{\overline{s}}$	$FD_{\overline{p}}$	$FD_{\overline{s}}$	$JSD_{\overline{p}}$	$JSD_{\overline{s}}$	$BD_{\overline{p}}$	$BD_{\overline{s}}$
$\mathcal{G}^{\varsigma}$	76	82	76	82	77	75	81	81	77	79
$\mathcal{G}^{\upsilon}$	84	84	83	84	**89**	87	85	84	83	**89**
$\mathcal{G}^{\varepsilon}$	77	78	76	79	81	77	76	79	76	79
$\mathcal{G}^{\tau}$	83	85	83	85	85	80	83	85	83	85
$\mathcal{G}^{\mu}$	70	71	70	71	78	71	71	72	64	75
$\mathcal{G}^{\sigma}$	71	68	69	69	75	69	70	69	68	76
$\mathcal{G}^{\vartheta}$	77	81	76	81	80	78	77	81	77	79
$\mathcal{G}^{\Im}$	56	58	56	58	56	54	56	57	52	59
$\mathcal{G}^{\wp}$	63	66	67	66	66	67	64	67	58	69
Avg	73	74.78	72.89	75.00	**76.33**	73.11	73.67	75.00	70.89	**76.67**

As can be seen from the table, Fishers discriminant with pseudo averaging produces the highest classification accuracy of 89%. Furthermore, it also produces an average classification accuracy of 76.33% for the various GLCM features used.

Table 5.2: Overall classification accuracies contd...

	$MD_{\overline{p}}$	$MD_{\overline{s}}$	$REneD_{\overline{p}}$	$REneD_{\overline{s}}$	Avg
$\mathcal{G}^{\varsigma}$	68	66	76	78	**76.71**
$\mathcal{G}^{\upsilon}$	71	65	82	85	**82.50**
$\mathcal{G}^{\varepsilon}$	66	66	72	75	**75.50**
$\mathcal{G}^{\tau}$	64	65	82	83	**80.79**
$\mathcal{G}^{\mu}$	48	47	65	66	67.07
$\mathcal{G}^{\sigma}$	58	53	67	70	68.00
$\mathcal{G}^{\vartheta}$	67	64	76	79	**76.64**
$\mathcal{G}^{\Im}$	63	64	52	49	56.43
$\mathcal{G}^{\wp}$	51	52	60	59	62.50
Avg	61.78	60.22	70.22	71.56	71.79

This is consistent with our clustering results and the stability analysis as the Fishers distance produced one of the highest values for $\mathcal{M}$ and high Rand indices as well. The other function that does as well is the Bhattacharyya distance for standard averaging. Although Bhattacharyya distance produces very high accuracies but it produces one of the most unstable decompositions for standard averaging. Hence, it is undesirable for our problem. The other distance functions produce different classification accuracies for different features. The average classification accuracy varies from 60% to 76%, while the highest classification accuracy for a given distance function ranges from 66% to 89%. This variability is due to the different subbands selected by each distance function as discussed in the results for clustering.

The reason why Fishers discriminant (pseudo averaging) and Bhattacharyya (standard averaging) perform better than Hellinger or Kullback-Leibler is because many subbands at the 4th level are selected from amongst the descendants of approximation $\mathcal{W}_{1,0,0}$ and the vertical detail $\mathcal{W}_{1,1,0}$ for these distance functions. These subbands are not selected in the case of Hellinger or Kullback-Leibler distance measure based decompositions. A FWPT produces classification accuracies of around 86% for the correlation feature while the average classification accuracy over the various features remains at 74%. This indicates that adding subbands that are not the most discriminant to the feature set reduces classification accu-

racies. The fact that SVM is a good classifier, for applications where the feature set is long, is an important factor here. Most other types of classifiers respond negatively to the increase in feature length. Hence selecting a subset using various distance functions is viable as it reduces the size of the feature set but increases the overall classification accuracy for SVMs as well.

The GLCM features that were found to be most useful are correlation ($\mathcal{G}^{\upsilon}$), energy ($\mathcal{G}^{\varepsilon}$), homogeneity ($\mathcal{G}^{\tau}$) and dissimilarity ($\mathcal{G}^{\vartheta}$). Correlation produces the highest overall classification accuracy. Contrast produces better results for some distance functions such as Hellinger and Kullback-Leibler but is not good for other distance functions. Next, we present the classification accuracies for each of the meningioma subtypes for the best distance function and GLCM features.

Table 5.3 shows that relatively high classification accuracies are obtained for each meningioma subtype for the correlation feature for the Fishers discriminant. It must be noted here that these results are 5-fold cross validated. Table 5.4 shows the results obtained for Fishers discriminant using union of decompositions. It can be seen that results produced with intersection are consistently better than the results produced using union. This is consistent with our analysis in the clustering results. However, just as in the clustering results the difference between the classification accuracies for union and intersection remains small but as we shall see in the cross-validated results presented later, the feature reduction is substantial. Hence, better classification results are achieved with feature reduction of up to 50%.

Table 5.3: Classification accuracy for 5 Best GLCM features for $\mathcal{B}^{*}_{\bigcap}$ using Fishers Discriminant and pseudo averaging (F=Fibroblastic, M=Meningiotheliamatous, P=Psammomatous, T=Transitional)

Feature	F	M	P	T	Avg
$\mathcal{G}_{\upsilon}$	79	89	97	89	**89**
$\mathcal{G}_{\varepsilon}$	68	83	97	75	81
$\mathcal{G}_{\tau}$	81	89	95	75	85
$\mathcal{G}_{\vartheta}$	71	84	95	68	80

Table 5.4: Classification accuracy for 5 Best GLCM features for $\mathcal{B}^*_{\bigcup}$ using Fishers Discriminant and pseudo-averaging (F=Fibroblastic, M=Meningiotheliamatous, P=Psammomatous, T=Transitional)

Feature	F	M	P	T	Avg
$\mathcal{G}_{\upsilon}$	79	89	98	84	87
$\mathcal{G}_{\varepsilon}$	75	84	97	63	80
$\mathcal{G}_{\tau}$	78	90	98	70	84
$\mathcal{G}_{\vartheta}$	72	84	96	60	78

Table 5.5 shows the results for various test trial runs for Fishers discriminant. The results presented are for the correlation feature, as it is the best feature included in our study. We can see that the reduction in features from the union to intersection is substantial in all of the trial runs but the resulting classification accuracies are improved. The features produced using intersection and union are given by,

$$\mathcal{B}^*_{\bigcap} = \bigcap_{i=1}^{N_{\mathcal{B}}} \mathcal{B}^*_i = \mathcal{B}^*_1 \bigcap \mathcal{B}^*_2 \bigcap \ldots \mathcal{B}^*_{N_{\mathcal{B}}} \tag{5.11}$$

$$\mathcal{B}^*_{\bigcup} = \bigcup_{i=1}^{N_{\mathcal{B}}} \mathcal{B}^*_i = \mathcal{B}^*_1 \bigcup \mathcal{B}^*_2 \bigcup \ldots \mathcal{B}^*_{N_{\mathcal{B}}} \tag{5.12}$$

These equations imply that all the terminal subbands found in the intersection and union decomposition are used for obtaining GLCM features and then provided to a SVM Gaussian Kernel. The best value of γ is found for each trial run as described in Section 5.1.3.

The results in Tables 5.3, 5.4 and 5.5 show that high classification accuracies can be obtained for meningioma subtype classification using Fishers discriminant based intersection decompositions. The intersection decomposition acquires the subbands which have the highest probability of being decomposed for a texture classification problem using ADWPT.

The results show that the most stable subbands i.e. the subbands which have the highest probability of occurrence are the best for high classification accuracies. This is a very important result. The selection of most stable subbands

indicates that the subbands that are found most discriminant in all the different test/training runs with different data sets used for training and testing are the most viable for differentiating between meningioma subtypes. This result resolves the issue of intra-class variability in real world textures such as meningioma subtypes. Such textures suffer from the problem that changing test and training data in various test/trial runs causes variable results and hence, the technique often fails completely. Using the robust ADWPT which is equivalent to the intersection of decompositions, we not only achieve reduction in the length of the feature sets and the selection of a consistent set of features, but the classification and clustering accuracies improve as well. The intersection decomposition is a robust ADWPT decomposition which selects only those subbands that are selected for all the different permutations of test/training data.

Another important factor that effects the classification accuracies is the discrimination power of a subband. We shall see how subband discrimination power effects the classification accuracies in the next section. Moreover, the relationship between stability and subband discrimination power needs to be investigated. There is also the issue of selecting the subbands that have a high frequency of occurrence but are not selected for all the different test trial runs such as the subbands that have a stability index of 0.6 or above. In the next section, we analyse the affect of subband discrimination power and subband stability on classification accuracies.

5.2 Subband Discrimination Power and Stability for Features Selection

The selection of subbands from a given Full Wavelet Packet Decomposition tree based upon the subband discrimination power is inherently selection of features. Feature selection is one of the most critical aspects of any pattern recognition algorithm. Efficient selection of features can boost the classification accuracy. The better the features, the higher the classification accuracy. This could be seen in the context of the intersection and union of subbands. Intersection provides better results most often as it represents a smaller subset of features that were

found in all the different ADWPT decompositions whereas union represents a superset that has a significantly higher number of features but the clustering efficiency and classification accuracy is lower. Therefore, subband stability has a direct impact on clustering and classification accuracies as it affects feature selection. In this section, we evaluate stability and determine a mechanism for feature selection from an ADWPT representation using the distance function found best in the previous section. The aim is to maximise the classification accuracy Ξ,

$$F = \underset{\mathcal{S},\mathcal{P}}{\operatorname{argmax}}\, \Xi(\triangle_f(\mathcal{S},\mathcal{P})) \tag{5.13}$$

where Ξ represents clustering or classification accuracy. The above equation stipulates that stability and subband discrimination power may be used for feature selection $\triangle_f$, in order to consequently maximise accuracy. Maximising stability $\mathcal{S}$ and subband discrimination power $\mathcal{P}$ would lead to better classification accuracies. Results for selection of features based upon stability and subband discrimination power are presented to evaluate how they impact the classification of Meningioma subtypes.

5.2.1 Stability and Classification Accuracy

The results in the previous section indicate that the most stable subbands produced the most accurate results. This is due to the fact that the intersection of decompositions, which inherently contains all the subbands that have the maximum stability index of 1 produces better results than the union of decompositions. Union of decompositions contains all the subbands that are found in any decomposition for a distance function and hence, contains subbands which have a low probability of occurrence or low stability. These subbands may not be discriminant for all or the majority of the test trial runs.

It is important to ascertain how stability affects the classification accuracy. In order to do so, we use the stability index to select subbands from the union tree of ADWPT decompositions. This is because although high accuracies for intersection indicate that high stability is good for classification but many subbands that have a high stability index but not equal to 1 have been ignored. It is important

to ascertain how these subbands affect the classification accuracy. The union tree contains all the subbands that are found in the union decomposition along with their ancestors. We perform a bottom up search and exclude any subbands along with its descendants which have a stability index lower than the threshold being used. Though it is evident that the stability index of the children of a subband is proportional to the stability of the parent. The algorithm used for obtaining a subband stability based decompositions is presented below.

1. Compute the J-level union decomposition tree for a distance function $\mathcal{D}$.
2. Calculate the stability index of each subband using the equation $\mathcal{S}_{d,p,q} = N_{d,p,q}/N_T$ where $\mathcal{S}_{d,p,q}$ represents the stability index of a subband, $N_{d,p,q}$ is the number of times the p,qth subband at depth d is found in all the various decompositions and N_T is the total number of decompositions.
3. Initialize $d = J - 1$.
4. For all $0 \leq p < 2^j$, $0 \leq q < 2^j$, do the following:
   a. If $\mathcal{S}_{d,p,q} \leq \mathcal{T}_S$ (where $\mathcal{T}_S$ is the stability threshold)
      Remove the four child subbands at depth d
   b. otherwise keep the subband at depth d and all its sibling subbands.
5. Decrement d by 1.
6. If $d < 0$, then stop, otherwise goto step 4.

Analysis using various thresholds was carried out and the results are presented in Table 5.6.

It can be seen in Table 5.6 that the overall classification accuracies steadily improve as the stability threshold is raised. As the stability threshold rises, fewer and fewer subbands are selected but the classification accuracies still improve. A steady rise in classification accuracies is observed for all the different GLCM features except for homogeneity. Although the number of subbands is reduced from 243 to 148, the classification accuracies for the GLCM feature correlation rises from 87% to 89%, for energy from 79% to 81%, for homogeneity from 84% to

85% and for dissimilarity from 78% to 80%. The average classification accuracies over all the GLCM features included rises from 82% to 83.8%. Therefore, choosing subbands based upon their stability index is a good idea as the number of subbands are reduced whereas the classification accuracies improve. The results in Table 5.6 are 5-fold cross validated i.e. 5 different combinations of test-training data using the leave one patient out from each class method was used and the results presented in the table are the average of all the trial run results.

The results show that the classification accuracies improve as more stable subbands are selected. This is an important result, as it reasserts the fact that the subbands that are selected most often are the most discriminant as these have been selected due to their subband discrimination power in most test-trial runs with different test and training data. Therefore, we select the decomposition obtained when the stability is at 1.0 and use it for further analysis in the rest of this chapter. Next, we investigate how the discrimination power of the subbands in the most stable decomposition affects the classification accuracies.

5.2.2 Subband Discrimination Power and Classification Accuracy

The effect of subband discrimination power on classification accuracy must be positive i.e. the higher the discrimination power of a subband included, the higher the classification accuracy must be. Our analysis indicates that there is a direct correlation between subband discrimination power of a feature set and classification accuracy obtained. Pearson's linear correlation coefficient obtained for classification accuracy with subband discrimination power is 0.9 i.e. as the subband discrimination power rises the classification accuracy rises as well. This mean that subband discrimination power is highly correlated with classification accuracy as the maximum value for the Pearson's linear correlation coefficient is 1.0.

Table 5.5: 5 fold cross validation results for Union feature set $F_{\bigcup}$ and Intersection feature set $F_{\bigcap}$ of Decompositions for Fishers discriminant (F=Fibroblastic, M=Meningiotheliamatous, P=Psammomatous, T=Transitional)

Test Trial #	Features	N	Acc_F	Acc_M	Acc_P	Acc_T	$Acc_{Overall}$
1	$\mathcal{B}_{\bigcup}$	247	65	83	94	73	79
	$\mathcal{B}_{\bigcap}$	154	65	81	92	83	80
2	$\mathcal{B}_{\bigcup}$	247	92	98	100	75	91
	$\mathcal{B}_{\bigcap}$	169	92	100	100	81	93
3	$\mathcal{B}_{\bigcup}$	244	83	67	100	92	85
	$\mathcal{B}_{\bigcap}$	139	88	67	98	94	87
4	$\mathcal{B}_{\bigcup}$	244	67	100	96	94	89
	$\mathcal{B}_{\bigcap}$	139	71	100	96	92	90
5	$\mathcal{B}_{\bigcup}$	244	88	98	98	88	93
	$\mathcal{B}_{\bigcap}$	139	81	96	100	94	93
Avg	$\mathcal{B}_{\bigcup}$	245	79	89	97	84	87
	$\mathcal{B}_{\bigcap}$	148	79	89	97	89	89

$\mathcal{T}_S$	Subbands	$\mathcal{G}_\upsilon$	$\mathcal{G}_\varepsilon$	$\mathcal{G}_\tau$	$\mathcal{G}_\vartheta$	Avg
0.1	243	87	79	84	78	82.0
0.3	218	88	79	83	79	82.3
0.5	205	88	80	83	80	82.8
0.7	200	88	80	84	80	83.0
0.9	153	88	80	85	80	83.3
1.0	148	89	81	85	80	83.8

Table 5.6: 5 fold cross-validated classification accuracies for the decompositions obtained using stability based thresholding ($\mathcal{G}_\upsilon$=GLCM Correlation, $\mathcal{G}_\varepsilon$=GLCM Energy, $\mathcal{G}_\tau$=GLCM Homogeneity, $\mathcal{G}_\vartheta$=GLCM Dissimilarity)

To investigate how selection of the more discriminant subbands affects the classification accuracies, we select a subset of subbands from the most stable decomposition i.e. intersection decomposition, using the subband discrimination power as a basis. The subbands are selected in the way that first the subbands representing 50% of the total subband discrimination power are selected with the most discriminant subbands selected first and then the less discriminant subbands and so forth. The results are presented in Tables 5.7 and 5.8. It must be noted here that the results in these Tables have been 5-fold cross validated.

% $\mathcal{P}$	$N_{\mathcal{D}}$	F	M	P	T	Avg
50	12	78	72	83	68	76
60	15	80	81	89	71	80
70	20	82	79	89	65	79
80	27	79	90	88	69	81
85	31	82	82	93	75	83
90	38	81	81	93	73	82
94	47	83	85	93	70	83
96	55	84	86	95	72	84
99	82	78	86	98	80	85

Table 5.7: 5 fold cross-validated classification accuracies for the subband selection using the most discriminant subbands $N_{\mathcal{D}}$ for the correlation feature (F=Fibroblastic, M=Meningiotheliamatous, P=Psammomatous and T=Transitional)

The results in Table 5.7 show that the overall classification accuracy improves with the inclusion of more features. Important thing to note is that with just 12 most discriminant features representing 50% of the total subband discrimination power, an overall classification accuracy of 76% is obtained. As we add the other features, the classification accuracy improves but there are instances when the overall classification accuracy falls as seen in the case when subbands representing 70% and 90% of the subband discrimination power are selected. The maximum classification accuracy attainable is 89% with 148 subbands. We can see in the table that with 82 features that represent 99% of the subband discrimination

power, a classification accuracy of 85% is achieved while the rest of the 56 subband features improve the classification accuracy by only 4% to raise it to 89%. Another important aspect of the results presented in the table is that the classification accuracies of individual meningioma subtypes do not always improve as the subband discrimination power of the feature set increases. As mentioned earlier, this is due to the fact that overall subband discrimination power is a sum of the subband discrimination power of pairwise comparisons between meningioma subtypes. Hence, each subband selected may not be good for differentiating all meningioma subtypes included in the study.

% Disc	No. Subbands	F	M	P	T	Avg
50	12	81	80	90	71	80
60	15	66	89	93	76	81
70	20	79	87	90	73	82
80	27	83	90	84	81	85
85	31	73	88	94	82	85
90	38	79	87	94	83	86
94	47	81	88	93	80	86
96	55	75	90	96	81	86
99	82	80	87	95	78	85

Table 5.8: 5 fold cross-validated classification accuracies for the subband selection using the most discriminant subbands for the homogeneity feature (F=Fibroblastic, M=Meningiotheliamatous, P=Psammomatous and T=Transitional)

The results for the homogeneity feature in Table 5.8 show that for the homogeneity feature, the 12 most discriminant subbands produce an overall classification accuracy of as high as 80%. The subbands are ordered based upon there subband discrimination power with the most discriminant subbands features added to the feature set first and then subbands with lower subband discrimination power are added to the feature set. For the homogeneity feature, addition of less discriminant subbands causes the classification accuracy to fall as can be seen in the table. Hence, 38 subbands which represent 90% of the subband discrimina-

tion power produce the highest classification accuracy in case of the homogeneity feature which is 86%.

From the results in Tables 5.7 and 5.8, it is evident that the discrimination power of a subband has an effect on the classification accuracy. The higher the discrimination power of a subband the higher the classification accuracy obtained using it. However, it may be observed that the most discriminant subbands alone may not be sufficient, as its a collection of subbands representing information from all the different frequency channels that together achieve high classification accuracies. This is obvious from the results in Tables 5.7 and 5.8, since a few most discriminant subbands alone are not sufficient for obtaining the highest classification accuracies but a sizable collection of them representing subbands from various frequency channels provide the best results. Another issue, as mentioned earlier is that all highly discriminant subbands may not be able to differentiate between all the meningioma subtypes included in the study with equal efficiency. Hence, a fall in classification accuracy for some subbands is observed for individual meningioma subtypes while the overall classification accuracy improves. In the next section, we will investigate how the selection of a subset of subbands that always improve the classification accuracy impacts the overall results.

5.2.3 Feature Selection

The selection of the wavelet based representation that best differentiates between meningioma subtypes is essentially selection of subband features. This is analogous to dimensionality reduction. Many techniques exist for dimensionality reduction, including linear techniques such as Principal Component Analysis (PCA) and non-linear techniques such as Diffusion Maps (DM).

However, as our results in the previous section indicate, there may still be room for improvement. As we see from the results in the previous section, not all subbands with a higher subband discrimination power yield higher classification accuracies. The theory of feature selection states that for any given problem with a defined number of elements N_e in a feature set X, a subset of the said feature set Y may be found that would improve the classification accuracy Ξ. Therefore for any feature set X, a feature set $Y \subset X$ exists such that $\Xi(Y) > \Xi(X)$, where

$\Xi(X)$ and $\Xi(Y)$ represent the classification accuracies obtained using the feature set X and Y respectively.

Many studies have been proposed advocating various approaches for feature selection [168] [169]. Lei *et al.* [169] propose feature selection based upon relevance and redundancy of the features. While Jain *et al.* [168] found floating selection as the best approach for feature selection in texture analysis. We combine the two approaches into a technique that adds 1 subband feature to the feature set at a time and keeping it if it improves the classification accuracy or discarding it if the classification accuracy falls. The relevance of a subband is measured based upon its discrimination power. A subband that is highly discriminant but causes the classification accuracy to fall is considered redundant. It is a greedy approach and finds a local optima as per the set of features used for differentiating between meningioma subtypes.

We employ two approaches for selection of subbands using the greedy algorithm, so that the usefulness of more discriminant subbands can be evaluated. In the first instance, we provide subband features in the frequency ordered mode while in the second instance we sort the subbands in terms of their discrimination power and then use them for classification.

Table 5.9 shows the results obtained for various test trial runs when subbands are selected using the greedy approach, in which the subbands that improve or maintain the classification accuracy are retained.

It can be clearly seen from the table that greedy selection using the most discriminant subbands often produces higher classification accuracies with fewer features. Hence, selection of features based upon the subband discrimination power aids in obtaining higher classification accuracies. The results obtained are encouraging as our technique is able to classify correctly all the various meningioma subtypes with greater than 90% accuracy. The psammomatous subtype is classified correctly almost 100% of the time with just one case producing a classification accuracy of 98% for this meningioma subtype.

The results in this section indicate that ADWPT is a robust and efficient technique that provides high classification accuracies by selecting subbands representing the most important spatial and frequency information. It not only achieves feature reduction by retaining the best features but also achieves high

Table 5.9: 5 fold cross validation results for Greedy feature set with frequency sorted subbands $\mathcal{F}$ and the Greedy feature set with subbands sorted based upon there discrimination power computed using Fishers Discriminant $\mathcal{F}_{\mathcal{D}}$ (F=Fibroblastic, M=Meningiotheliamatous, P=Psammomatous, T=Transitional)

Test Trial #	Feature Selection	N	Acc_F	Acc_M	Acc_P	Acc_T	$Acc_{Overall}$
1	$\mathcal{F}$	35	**87.5**	77.1	95.8	87.5	87
	$\mathcal{F}_{\mathcal{D}}$	**32**	85.4	**79.2**	**97.9**	**93.8**	**89.1**
2	$\mathcal{F}$	**77**	**95.8**	100	100	**91.7**	**96.9**
	$\mathcal{F}_{\mathcal{D}}$	80	93.8	100	100	87.5	95.3
3	$\mathcal{F}$	54	89.6	72.9	97.9	**95.8**	89.1
	$\mathcal{F}_{\mathcal{D}}$	**49**	**93.8**	**77.1**	**100**	89.6	**90.1**
4	$\mathcal{F}$	47	**93.8**	100	95.8	**97.9**	**96.9**
	$\mathcal{F}_{\mathcal{D}}$	**38**	89.6	100	100	93.8	95.8
5	$\mathcal{F}$	60	83.3	97.9	100	93.8	93.8
	$\mathcal{F}_{\mathcal{D}}$	**50**	**89.6**	**100**	100	**97.9**	**96.9**
Avg	$\mathcal{F}$	55	90	89.58	97.9	**93.34**	92.74
	$\mathcal{F}_{\mathcal{D}}$	**50**	**90.44**	**91.26**	**99.58**	92.52	**93.44**

classification accuracies. ADWPT also provides a means of ranking features based upon the subband discrimination power and stability of the subbands the features are acquired from. The results in this chapter indicate that there is a direct relationship between subband discrimination power and classification accuracy, and between stability and classification accuracy. Hence, ADWPT is an highly effective technique which is not only robust to variation in the data samples but also provide a means of ranking features and at the same time provides high classification accuracies.

5.3 Conclusions

In this chapter, we presented the classification accuracies that may be obtained using the various distance functions for the two different mechanisms for MAWTT template construction. Some distance functions are found to be better than other

distance functions. The classification accuracies using SVMs are obtained for each of the distance functions and the results for the two types of MAWTT template construction modes is investigated. The best results are obtained for Fishers discriminant using pseudo-averaging. Overall classification accuracies as high as 89% are obtained for the intersection of ADWPT decompositions. Hence, we provide a technique that resolves the problem of texture variation by selecting the subbands that are selected for all the different permutations of test/training data i.e. the most stable subbands. Although substantial reduction in features is acquired using the intersection of decompositions, the classification accuracies for the meningioma subtypes are still improved. Hence, a robust ADWPT decomposition is acquired using the intersection of decompositions which not only resolves the issue of intra-class texture variation but also improves classification accuracies.

In this chapter, we also analysed how selection of subbands based upon their discrimination power and stability index affects the classification performance. We have shown that increasing the stability of a decomposition i.e. selecting the most stable subbands improves the classification accuracy. Therefore, the more stable a subband, the better it is for discriminating between meningioma subtypes. Another important factor that affects the classification accuracy is the discrimination power of a subband. The results show that there is a direct relationship between the discrimination power of a subband and its effectiveness in differentiating between meningioma subtypes. We show that selecting the more discriminant subbands improves the classification accuracy more substantially in comparison to selecting any other subband. A feature selection technique using subband discrimination power as a guide is presented which performs better than greedy selection of subband features.

Hence, ADWPT is a powerful approach for subband selection to acquire features for meningioma subtype classification. It not only makes it possible to acquire high classification accuracies by selecting the most relevant spatial frequency resolution subbands but also overcomes the issue of wavelet packet decomposition variability due to high intra-class texture variation in meningioma subtypes. ADWPT is a robust mechanism for selection of a consistent set of fea-

tures that provide high classification accuracies. It also provide a mechanism for sorting features based upon subband discrimination power and stability index.

Chapter 6

Conclusions

The thesis presents a wavelet packet transform based technique for histological image texture classification. The approach overcomes the issue of high intra-class and low inter-class variation in the texture found in meningioma images. High classification accuracies are obtained using the technique. In this chapter, we will present a summary review of the technique and also discuss the main contributions of this thesis. The limitations of the technique are also presented with a section on future directions.

6.1 Thesis Summary

In this thesis, we have developed a technique that borrows from various approaches to pattern classification. The target application is a real world problem from the domain of histological image analysis.

Since human beings are the best pattern recognizers, therefore the Human Visual System (HVS) holds important clues and ideas for the construction of a viable pattern recognition technique. It is known that the HVS is composed of various types of cells at various layers that perform different functions. The visual stimuli is processed by the different layers before interpretation by the brain. There is sufficient evidence to suggest that the layers acquire the data and process it so that the edge information is separated from the rest of the texture. Texture orientation is also extracted separately and the visual scene is interpreted in terms of statistical properties as determined by Julesz *et al.*

[107, 170]. These statistical features include first-order, second-order and higher-order features. These features make different contributions to the overall texture description in the process of visual cognition.

Many theories that represent the process of pattern recognition have been presented in the first chapter. These include template matching, statistical and syntactic features, structural approaches and topological or transform based techniques. Just like the HVS, these techniques aim at simplifying the process of pattern recognition by breaking up the visual stimuli into constituent features (statistical and structural) and transforming the signal into different more descriptive representations. Hence, it is ascertained that efficient recognition of complex visual patterns from the real world involves acquiring different more meaningful representations of the pattern and extraction of statistical features that represent the pattern. The idea is to develop a technique that would transform the data so that the textural edge and the non-edge information is extracted separately and a mechanism is evolved to capture this information in terms of the statistical properties just as the HVS does. This forms the basis of our technique.

In this thesis, we present a technique that we call the Adaptive Discriminant Wavelet Packet Transform (ADWPT). ADWPT transforms an input signal, in the case of this thesis, histological images of Meningioma samples into a multiresolution representation that decomposes the image in to various spatial and frequency resolutions. This decomposition represents the edges and other textural information in the image at multiple resolutions. The ADWPT finds the best spatial-frequency resolutions to describe the signal so that the most discriminant features that differentiate between the various meningioma subtypes are selected. These spatial-frequency subbands are used to extract Gray Level Co-occurrence Matrix (GLCM) based statistical features. These features are used to describe the texture and to perform classification using various pattern classifiers. The technique aims to mimic the HVS as it acquires a textural representation with edge and non-edge information in different subbands and subsequently uses the statistical features extracted (as per the statistical approach), for texture description and classification.

The technique is shown to be robust as it finds the texture representation that does not change when the data samples for meningioma subtypes used to

construct it are changed. These representations provide features that acquire high classification accuracies. Various features are extracted from the subbands in the ADWPT decomposition. The features that provide the best results are Gray Level Co-occurrence Matrix (GLCM) features. Other features such as first-order and second order statistical features were also used for analysis.

Various classifiers were used for analysis with Support Vector Machines (SVMs) providing the best results. Other classifiers used were Learning Vector Quantization (LVQ) and k nearest neighbours (k-NN). Classification accuracies of as high as 90% may be achieved using SVMs for the various meningioma subtypes with the psammomatous subtype providing near perfect results. Clustering using k-means clustering is also carried out and the results are presented.

In essence, this thesis presents a technique that aims to mimic the HVS in terms of how visual images are interpreted. This is a very simplistic approximation to the HVS which is far more complex and sophisticated. The results produced indicate that the technique is good for differentiating between complex real world textures.

6.2 Main Contributions

The main contributions of the work presented in this thesis are:

1. Discriminant Subband Selection: In Chapter 3, an algorithm for the selection of most discriminant subbands is presented. The algorithm compares how various distance functions capture the ability of a subband to differentiate between two meningioma subtypes. The results indicate that certain distance functions are better than others for the problem of meningioma subtype classification with Fisher's discriminant providing the best results. The approach selects subbands from an overcomplete wavelet packet representation and acquires high accuracy results as shown in Chapter 5.

2. Stability Analysis: Chapter 3 introduces the concept of ADWPT stability and Chapter 4 presents a detailed analysis of the stability of decompositions for a distance function. Stability represents the probability of a decomposition to remain consistent when the data used to construct the ADWPT

decomposition changes. This is an important factor because if the decompositions change greatly when the data changes then it would be difficult to identify the subbands that contain information important for meningioma subtype classification. Our analysis showed that the higher the stability of the ADWPT decomposition, the higher the classification accuracy.

3. ADWPT Feature Analysis: Chapter 4 describes the various feature extraction techniques for capturing the information contained in a subband. Apart from other features this thesis presents a detailed discussion and analysis of various feature extraction techniques from subbands in the ADWPT decompositions. The results show that GLCM features are the best features from amongst those tried for meningioma subtype classification when combined with ADWPT.

4. Subband Discrimination Power and Classification Accuracy: In chapter 5, we have presented a detailed analysis of the relationship between the discrimination power of a subband and its impact on the overall classification accuracy. The results show that there is a direct relationship between the discrimination power of a subband and the classification accuracy achieved when features extracted from it are used for classification. As the more discriminant subband features are added to the feature set, a jump in the overall classification accuracy is observed in most cases. This implies that the approach followed in this thesis for subband selection based upon the discrimination power is valid since the discrimination power is a good measure for describing the ability of a subband to differentiate between meningioma subtypes.

6.3 Limitations of the Work

The main limitation of the work is in the method used to compute the discrimination power of a subband. As stated earlier, the discrimination power of a subband is a sum of the distances computed for six different comparisons amongst the 4 different meningioma subtypes being analysed. Although the results in Chapters 4 and 5 indicate that the subbands with higher discrimination power most often

impact the classification accuracy positively i.e. they increase the accuracies, but certain subbands that have a relatively high discrimination power are seen to decrease the overall classification accuracy. This is due to the fact that a subband that is good for differentiating between two subtypes may not be as good for discriminating between all the other meningioma subtypes. Therefore, the computation of discrimination power is susceptible to the bias introduced due to a high distance value between two textures under study and low values for comparison between other texture types. This problem would be compounded as the number of textures for comparison was increased. Hence, the technique is suited to instances where the number of texture classes under study is limited. The results are near perfect for most two class cases using meningioma subtype images.

In future work, we will try to overcome this problem and make the technique more efficient. There are various avenues of further research that may be explored in order to improve the technique and the classification accuracy obtained using it.

6.4 Future Directions

There are various future directions in which the research may be carried out. These are presented in a summary form as follows:

1. Averaging for MAWTT construction: In this thesis we have explored two means of averaging of subband coefficients to construct the MAWTTs. As stipulated in the chapter 3, there could be various other mechanisms to acquire the templates. These could be various other averaging approaches that use different averaging windows. In averaging the size of the window would be an important factor as a small window would be more suited to frequent changes in coefficient values whereas a long window would be more suited for instances where the coefficient values are more consistent. These averaging windows could be different for the various subbands depending upon the frequency content they represent which in turn depends upon the texture samples being analysed. Hence, the search for the optimal averaging

window could be a subject of further research. Something conclusive can be ascertained with thorough analysis. Although it is important to realize here that the ultimate aim is to obtain stable decompositions which represent subbands that can be used to attain high classification accuracies.

2. Distance functions analysis: In this thesis, we have presented various different distance functions and obtained ADWPT representations using each for classifying meningioma subtypes. More distance functions may be used with an emphasis on obtaining a more accurate estimation of the discrimination power of each subband. The important issue to resolve is the accurate estimation of discrimination power for differentiating between multiple texture classes, so that a high distance value for two texture classes does not bias the discrimination power. It may also be found that out of the various short listed distance functions included in the classification analysis in chapter 4 different distance functions may be found better for different texture classification problems.

3. Binary classification and combination: Currently we have tried to compare all the various meningioma subtypes at the same time. A viable alternative would be to compare them two at a time i.e. computation of the various decompositions and classification using the various classifiers. This would raise the issue of ties and generation of multiple ADWPT decompositions for different pairs of meningioma textures. However, more elaborate analysis and experimentation may be carried out to explore all possible modalities of ADWPT and classifier combination in a binary classification mode.

4. Segmentation and structural analysis: Segmentation has been a more popular approach for classification in the domain of histological image analysis with textural analysis not being explored more vigorously. For meningioma subtype classification, segmentation may provide information that could facilitate meningioma subtype classification. Various segmentation schemes may be applied to extract structural information from the histological image samples and these may be combined with our textural analysis technique for better classification results.

5. Colour information: In our study, we have not explored colour information at all for image analysis. Colour based techniques such as colour features, colour histograms and colour wavelets may be applied to Meningioma subtype classification.

This thesis presents a novel wavelets based technique that is robust to intra-class variation and achieves efficient feature selection for high accuracy classification results in image classification problems that suffer with low inter-class differences such as Meningioma subtype classification.

Appendix A

List of Publications

H Qureshi, N. Rajpoot, T. Nattkemper, Volkmar Hans, "A Robust Adaptive Wavelet-based Method for Classification of Meningioma Histology Images", in *Proceedings MICCAI'2009 Workshop on Optical Tissue Image Analysis in Microscopy, Histology, and Endoscopy (OPTIMHisE)*, London (UK), 2009

H Qureshi, O. Sertel, R. Wilson, N. Rajpoot, M. Gurcan, "Adaptive Discriminant Wavelet Packet Transform and Local Binary Patterns for Meningioma Subtype Classification", in *Proceedings Medical Image Computing and Computer Assisted Intervention (MICCAI'08)*, New York (USA), 2008

H Qureshi, R. Wilson, N. Rajpoot, "Optimal Wavelet Basis for Wavelet Packets based Meningioma Subtype Classification", in *Proceedings Medical Image Understanding and Analysis (MIUA'08)*, Dundee (UK), July 2008

H Qureshi, N Rajpoot, R Wilson, T Nattkemper, V Hans, "Comparative analysis of Discriminant Wavelet Packet Features and Raw Image Features for Classification of Meningioma Subtypes", in *Proceedings Medical Image Understanding and Analysis (MIUA'07)*, Aberystwyth (UK), July 2007

H Qureshi, NM Rajpoot, K Masood, V Hans, "Classification of Meningiomas using Discriminant Wavelet Packets and Learning Vector Quantization", in *Proceedings Medical Image Understanding and Analysis (MIUA'06)*, Manchester (UK),

July 2006

Khalid Masood, Nasir Rajpoot, Hammad Qureshi and Kashif Rajpoot, "Hyperspectral Texture Analysis for Colon Tissue Biopsy Classification", in *International Symposium on Health Informatics and Bioinformatics*, Turkey 2007

H Qureshi, NM Rajpoot, K Masood, V Hans, "Classification of Meningiomas using Discriminant Wavelet Packets and Learning Vector Quantization", in *Poster Session of Wave 2006*, EPFL, Lausanne, Switzerland, July 2006

H Qureshi, N Rajpoot, T Nattkemper, V Hans, "Comparison: Meningioma Classification using Wavelet Packets and Normal Texture based Classification", in *1st British Machine Vision Association Student Chapters Meeting*, London (UK), March 2007

Appendix B

Distance Functions, Decompositions and Stability Graphs

B.1 Distance Functions

The calculation of the most discriminant decomposition is of paramount importance in our study. The structure of the ADWPT decomposition is dependent on the discriminant function $\mathcal{D}$. We have used some of the popular linear and statistical measures to compute $\mathcal{D}$ and have also derived some of our own measures. A discussion of these functions is as follows:

B.1.1 Hellinger Distance

The pairwise discrimination power of the (p, q)th subband located at depth d is calculated using the Hellinger distance [171] as follows:

$$\mathcal{D}_{d,p,q}^{a,b} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[\sqrt{\mathcal{A}_{d,p,q}^{a}(m,n)} - \sqrt{\mathcal{A}_{d,p,q}^{b}(m,n)} \right]^2 \tag{B.1}$$

where $\mathcal{A}_{d,p,q}^{a}(m,n)$ and $\mathcal{A}_{d,p,q}^{b}(m,n)$ denote the standard average or pseudo-average ppdes of the (p, q)th subband $\mathcal{T}_{d,p,q}^{a}$ and $\mathcal{T}_{d,p,q}^{b}$ at depth d for classes a and b respectively. The m and n represent the (m, n)th coefficient of the template subband.

B.1.2 Kullback-Leibler Distance

The pairwise discrimination power of the (p,q)th subband located at depth d is calculated using the Kullback-Leibler distance [129] as follows:

$$\mathcal{D}_{d,p,q}^{a,b} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left(\mathcal{A}_{d,p,q}^{a}(m,n)log\left(\frac{\mathcal{A}_{d,p,q}^{a}(m,n)}{\mathcal{A}_{d,p,q}^{b}(m,n)}\right)+\mathcal{A}_{d,p,q}^{b}(m,n)log\left(\frac{\mathcal{A}_{d,p,q}^{b}(m,n)}{\mathcal{A}_{d,p,q}^{a}(m,n)}\right)\right) \tag{B.2}$$

B.1.3 Fishers Linear Discriminant

The pairwise discrimination power of the (p,q)th subband located at depth d is calculated using the Fishers Linear Discriminant [129] as follows:

$$\mathcal{D}_{d,p,q}^{a,b} = \frac{(\mu_{d,p,q}^{a}-\mu_{d,p,q}^{b})^2}{(\sigma_{d,p,q}^{a})^2+(\sigma_{d,p,q}^{b})^2} \tag{B.3}$$

where $\mu_{d,p,q}^{a}$ and $\mu_{d,p,q}^{b}$ represent the mean of the (p,q)th template subbands $\mathcal{T}_{d,p,q}^{a}$ and $\mathcal{T}_{d,p,q}^{b}$ at depth d for classes a and b and $\sigma_{d,p,q}^{a}$ and $\sigma_{d,p,q}^{b}$ represent the variance of the (p,q)th MAWTT subband at depth d for classes a and b respectively.

B.1.4 Jensen-Shannon Distance

The pairwise discriminating power of the (p,q)th subband located at depth d is calculated using the Jensen-Shannon distance [172] as follows:

$$\mathcal{D}_{d,p,q}^{a,b} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left[\mathcal{A}_{d,p,q}^{a}(m,n)log\left(\frac{\mathcal{A}_{d,p,q}^{a}(m,n)}{\hat{\mathcal{A}}_{d,p,q}^{a,b}(m,n)}\right)+\mathcal{A}_{d,p,q}^{b}(m,n)log\left(\frac{\mathcal{A}_{d,p,q}^{b}(m,n)}{\hat{\mathcal{A}}_{d,p,q}^{a,b}(m,n)}\right)\right]/2 \tag{B.4}$$

where $\hat{\mathcal{A}}_{d,p,q}^{a,b}(m,n)$ denotes the average of the two ppdes being analysed, given by

$$\hat{\mathcal{A}}_{d,p,q}^{a,b}(m,n) = \frac{(\mathcal{A}_{d,p,q}^{a}(m,n)+\mathcal{A}_{d,p,q}^{b}(m,n))}{2} \tag{B.5}$$

The Jensen-Shannon distance is unique as it incorporates the averaging of the ppdes of the respective subbands.

B.1.5 Bhattacharyya Distance

The pairwise discrimination power of the (p, q)th subband located at depth d is calculated using the Bhattacharyya distance [129] as follows:

$$\mathcal{D}_{d,p,q}^{a,b} = \frac{1}{4}\left[log\left(\frac{(\sigma_{a,d,p,q}^2 - \sigma_{b,d,p,q}^2)^2}{4\sigma_{a,d,p,q}^2\sigma_{b,d,p,q}^2}\right) + FD_{d,p,q}^{a,b}\right] \tag{B.6}$$

where $\sigma_{d,p,q,a}$ and $\sigma_{d,p,q,b}$ represent the standard deviations of the (p, q)th MAWTT subband $\mathcal{T}_{d,p,q}$ at depth d for classes a and b respectively and $FD_{d,p,q}^{a,b}$ is the value of the Fishers discriminant between the respective subbands as defined above in Eq. 3.10.

B.1.6 Mahalanobis Distance

The pairwise discrimination power of the (p, q)th subband located at depth d is calculated using the Mahalanobis distance [173] as follows:

$$\mathcal{D}_{d,p,q}^{a,b} = \sqrt{|X_{d,p,q}^{a,b} C_{d,p,q}^{a,b} (X_{d,p,q}^{a,b})^t|} \tag{B.7}$$

where $|.|$ denotes the determinant and X is a mean matrix containing the differences of μ of the ppdes under study and is given by,

$$X_{d,p,q}^{a,b} = \begin{pmatrix} \mu_{d,p,q}^{a,a} & \mu_{d,p,q}^{a,b} \\ \mu_{d,p,q}^{b,a} & \mu_{d,p,q}^{b,b} \end{pmatrix} \tag{B.8}$$

where $\mu_{d,p,q}^{a,b} = \mu_{d,p,q}^{a} - \mu_{d,p,q}^{b}$ denotes the mean of the ppde of the (p, q)th subband at depth d for classes a and b respectively and $C_{d,p,q}^{a,b}$ represents the covariance of the distributions of the MAWTT subbands $\mathcal{T}_{d,p,q}^{a}$ and $\mathcal{T}_{d,p,q}^{b}$ representing the two classes a and b respectively. The standard deviation of the subband would be given by $\nu_{d,p,q}^{a} = \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\sqrt{(\mathcal{A}_{d,p,q}^{a}(m,n) - \mu_{d,p,q}^{a})^2}}{M \times N}$. The covariance matrix $C_{d,p,q}^{a,b}$ for two classes can then be represented as,

$$C_{d,p,q}^{a,b} = \begin{pmatrix} \nu_{d,p,q}^{a}\nu_{d,p,q}^{a} & \nu_{d,p,q}^{a}\nu_{d,p,q}^{b} \\ \nu_{d,p,q}^{b}\nu_{d,p,q}^{a} & \nu_{d,p,q}^{b}\nu_{d,p,q}^{b} \end{pmatrix} \tag{B.9}$$

For n classes, the covariance matrix is given as,

$$C_{d,p,q}^{c_1,c_2,\ldots,c_n} = \begin{pmatrix} \nu_{d,p,q}^1\nu_{d,p,q}^1 & \nu_{d,p,q}^1\nu_{d,p,q}^2 & \cdots & \nu_{d,p,q}^1\nu_{d,p,q}^n \\ \nu_{d,p,q}^2\nu_{d,p,q}^1 & \nu_{d,p,q}^2\nu_{d,p,q}^2 & \cdots & \nu_{d,p,q}^2\nu_{d,p,q}^n \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{d,p,q}^n\nu_{d,p,q}^1 & \nu_{d,p,q}^n\nu_{d,p,q}^2 & \cdots & \nu_{d,p,q}^n\nu_{d,p,q}^n \end{pmatrix} \tag{B.10}$$

This also brings us to an important result. From the equations above, it could be implied that the Mahalanobis distance could be used to compare any number of textural classes. The Mahalanobis distance for four classes can be written as,

$$\mathcal{D}_{d,p,q}^{a,b,c,d} = \sqrt{|X_{d,p,q}^{a,b,c,d} C_{d,p,q}^{a,b,c,d} (X_{d,p,q}^{a,b,c,d})^t|} \tag{B.11}$$

where

$$X_{d,p,q}^{u,v,w,x} = \begin{pmatrix} \mu_{d,p,q}^{u,u} & \mu_{d,p,q}^{u,v} & \mu_{d,p,q}^{u,w} & \mu_{d,p,q}^{u,x} \\ \mu_{d,p,q}^{v,u} & \mu_{d,p,q}^{v,v} & \mu_{d,p,q}^{v,w} & \mu_{d,p,q}^{v,x} \\ \mu_{d,p,q}^{w,u} & \mu_{d,p,q}^{w,v} & \mu_{d,p,q}^{w,w} & \mu_{d,p,q}^{w,x} \\ \mu_{d,p,q}^{x,u} & \mu_{d,p,q}^{x,v} & \mu_{d,p,q}^{x,w} & \mu_{d,p,q}^{x,x} \end{pmatrix} \tag{B.12}$$

denotes an array of the difference of means of the (p, q)th subband at depth d for classes u, v, w and x respectively and $C_{d,p,q}^{u,v,w,x}$ represents the covariance matrix of the ppdes representing the classes u, v, w and x.

B.1.7 Energy Distance

The energy distance is fundamentally the difference in energy of respective subbands of the decompositions representing classes a and b and is given by,

$$\mathcal{D}_{d,p,q}^{a,b} = (\mathcal{E}_{d,p,q}^{a} - \mathcal{E}_{d,p,q}^{b})^2 \tag{B.13}$$

where $\mathcal{E}_{d,p,q}^{a} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}(\mathcal{A}_{d,p,q}^{a}(m,n))^2$. Hence, $\mathcal{E}_{d,p,q}^{a}$ represents the energy of the (p, q)th MAWTT subband at depth d for class a.

B.1.8 Relative Energy Distance

In this section, we propose a novel metric for comparison of subbands. This metric has not been used for subband selection yet to the best of our knowledge. The relative energy distance may be given as,

$$\mathcal{D}^{a,b}_{d,p,q} = \left(\frac{\mathcal{E}^{a}_{d,p,q}}{\mathcal{E}^{b}_{d,p,q}} + \frac{\mathcal{E}^{b}_{d,p,q}}{\mathcal{E}^{a}_{d,p,q}} \right) /2 - 1 \tag{B.14}$$

The relative energy distance function may be derived as,

$$F(\mathcal{E}^{a}_{d,p,q}, \mathcal{E}^{b}_{d,p,q}) = \left(\frac{\mathcal{E}^{a}_{d,p,q}}{\mathcal{E}^{b}_{d,p,q}} + \frac{\mathcal{E}^{b}_{d,p,q}}{\mathcal{E}^{a}_{d,p,q}} \right) /2 - 1 \tag{B.15}$$

$$\Rightarrow F(\mathcal{E}^{a}_{d,p,q}, \mathcal{E}^{b}_{d,p,q}) = \left(\frac{(\mathcal{E}^{a}_{d,p,q})^2 + (\mathcal{E}^{b}_{d,p,q})^2}{2\mathcal{E}^{a}_{d,p,q}\mathcal{E}^{b}_{d,p,q}} \right) - 1 \tag{B.16}$$

$$\Rightarrow F(\mathcal{E}^{a}_{d,p,q}, \mathcal{E}^{b}_{d,p,q}) = \left(\frac{(\mathcal{E}^{a}_{d,p,q})^2 + (\mathcal{E}^{b}_{d,p,q})^2 - 2\mathcal{E}^{a}_{d,p,q}\mathcal{E}^{b}_{d,p,q}}{2\mathcal{E}^{a}_{d,p,q}\mathcal{E}^{b}_{d,p,q}} \right) \tag{B.17}$$

$$\Rightarrow F(\mathcal{E}^{a}_{d,p,q}, \mathcal{E}^{b}_{d,p,q}) = \frac{(\mathcal{E}^{a}_{d,p,q} - \mathcal{E}^{b}_{d,p,q})^2}{2\mathcal{E}^{a}_{d,p,q}\mathcal{E}^{b}_{d,p,q}} \tag{B.18}$$

Therefore, the relative energy distance differs from the simple energy distance given in the previous section.

B.2 Union of Decompositions

This section presents the union decompositions (defined in section 4.1.1) acquired for various distance functions.

B.2.1 Hellinger Distance

Figures B.1 B.2 show the union of the ADWPT decompositions obtained using Hellinger distance.

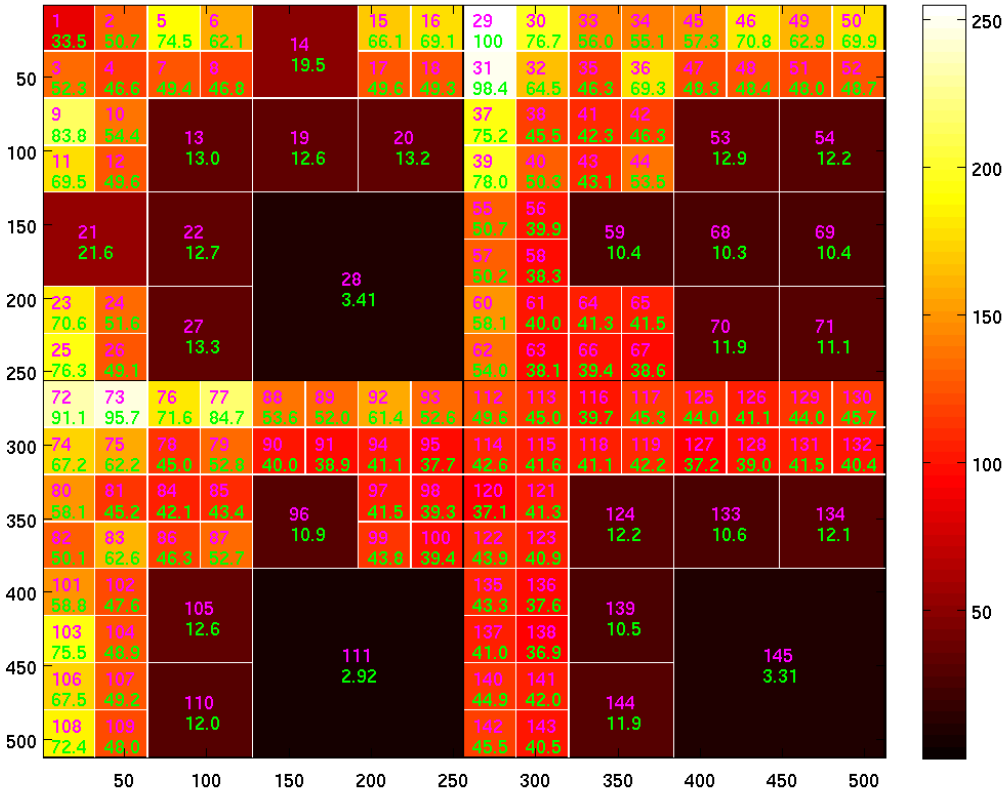


Figure B.1: Union of ADWPT decompositions obtained using the Hellinger distance (Standard averaging)

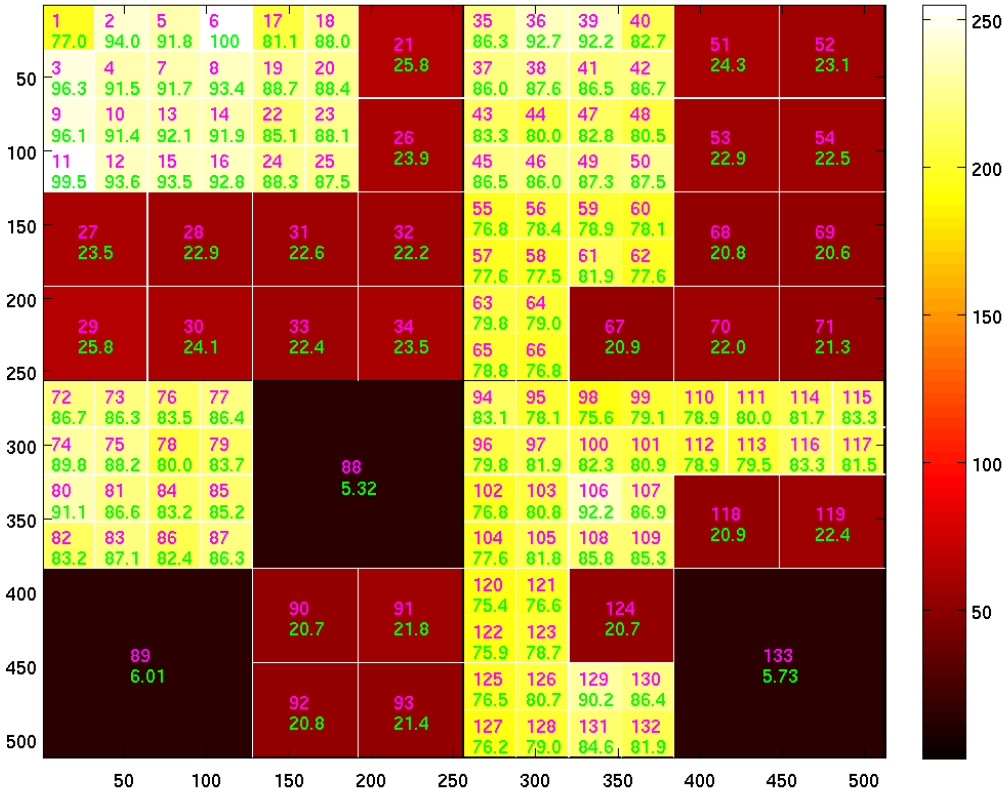


Figure B.2: Union of ADWPT decompositions obtained using the Hellinger distance (Pseudo averaging)

The green number indicates the scaled discriminance value of a subband (maximum being 100 and minimum being 0) whereas the magenta is the number of the subband ordered in terms of increasing frequency.

B.2.2 Kullback-Leibler Distance

Figures B.3 B.4 show the union of the ADWPT decompositions obtained using Kullback-Leibler distance.

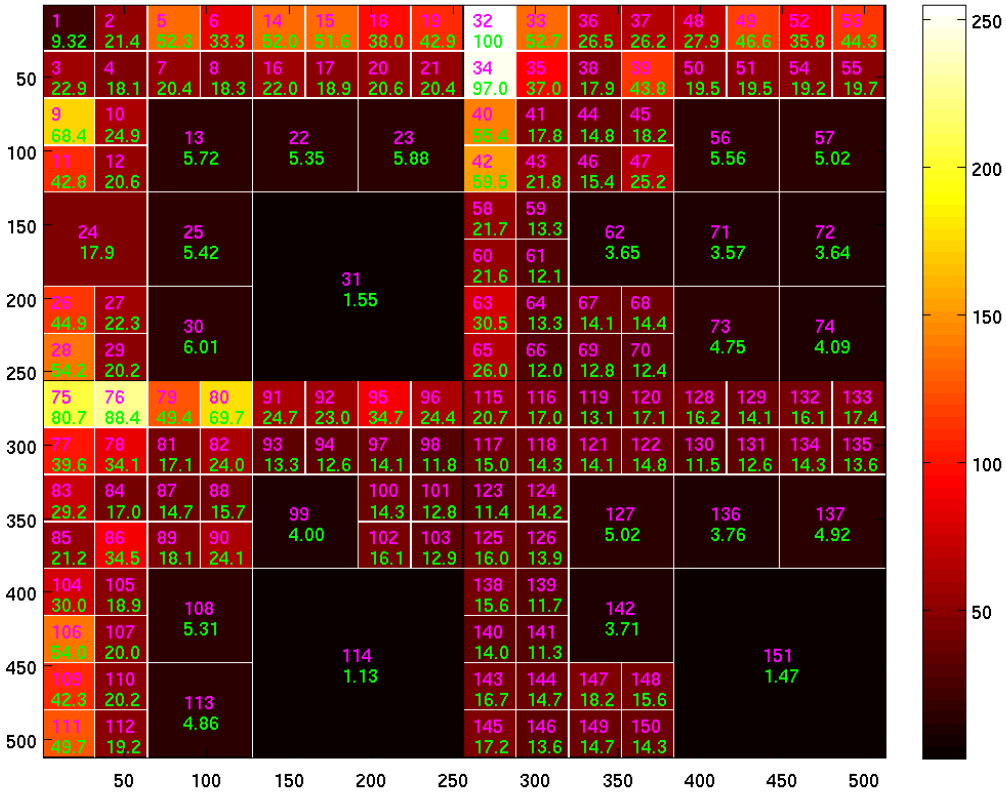


Figure B.3: Union of ADWPT decompositions obtained using the Kullback-Leibler distance (Standard averaging)

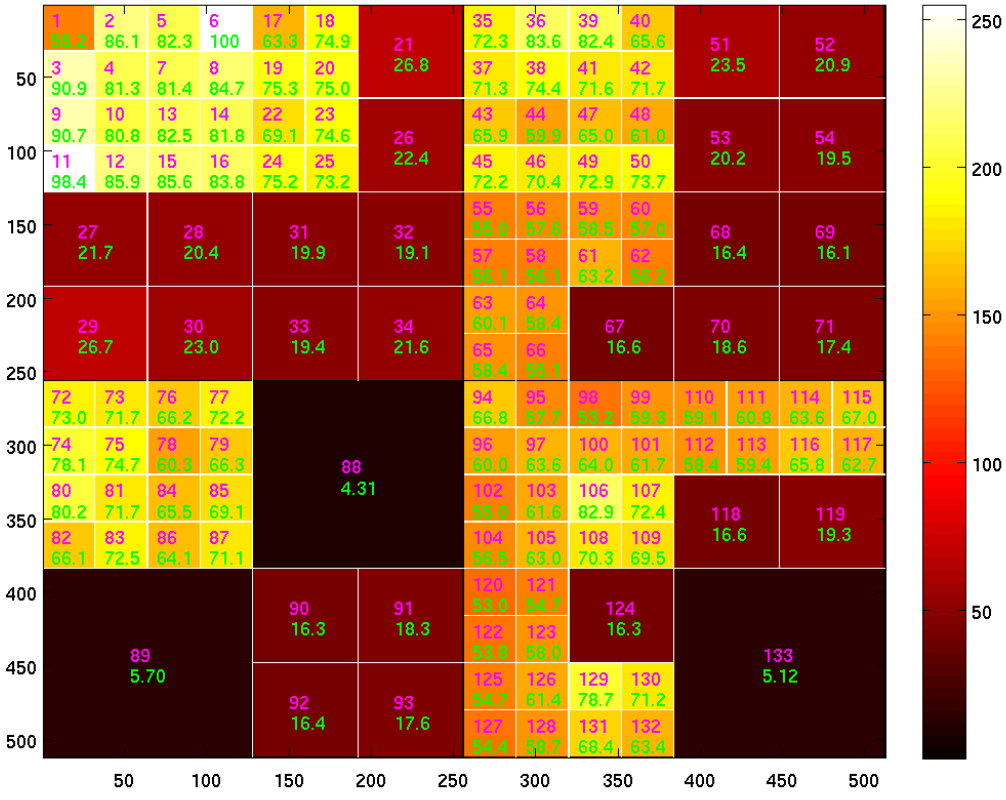


Figure B.4: Union of ADWPT decompositions obtained using the Kullback-Leibler distance (Pseudo averaging)

B.2.3 Fishers Linear Discriminant

Figures B.5 B.6 show the union of the ADWPT decompositions obtained using Fishers Linear Discriminant.

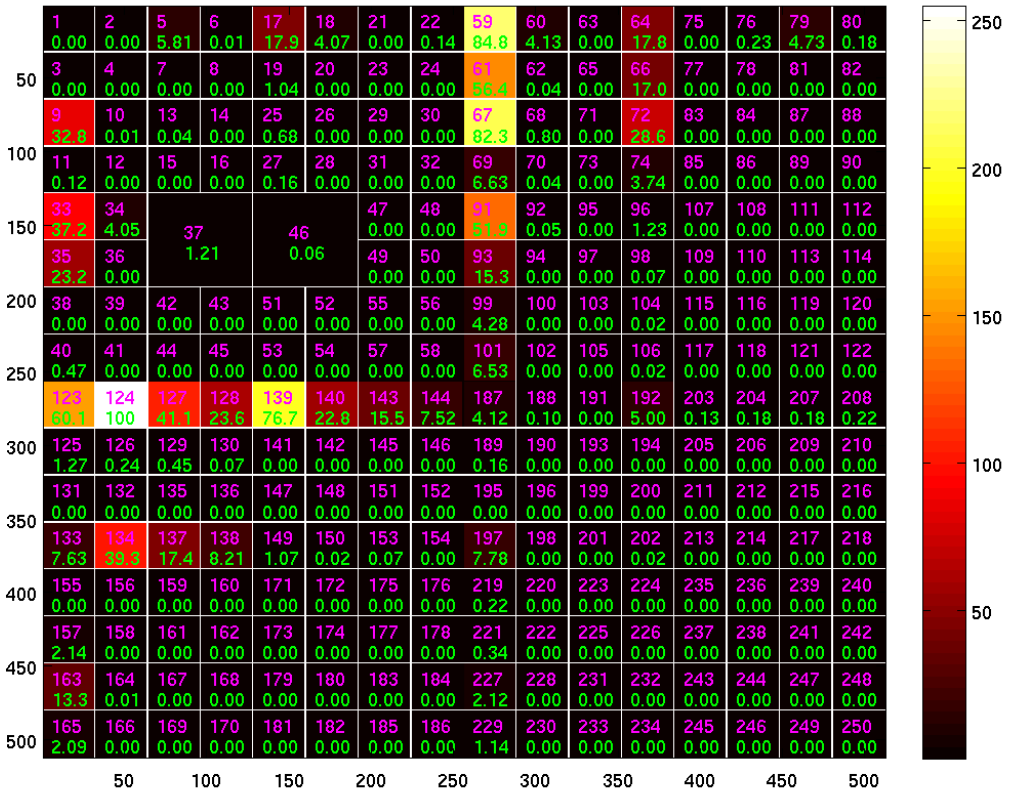


Figure B.5: Union of ADWPT decompositions obtained using the Fishers Linear Discriminant based distance (Standard averaging)

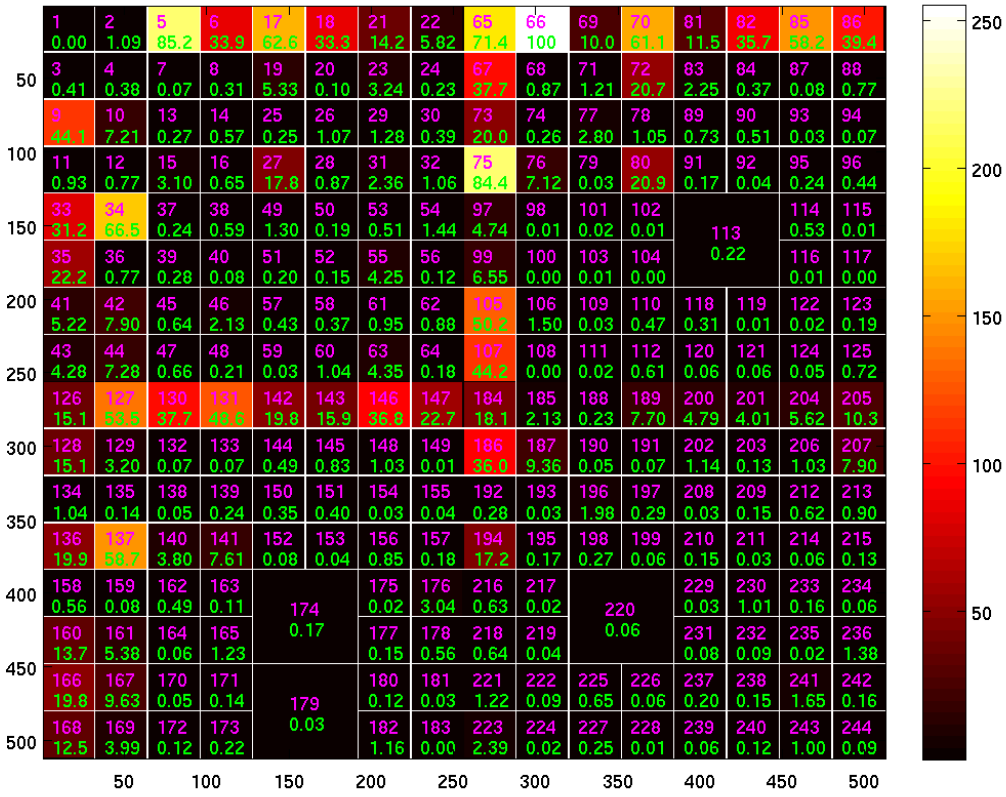


Figure B.6: Union of ADWPT decompositions obtained using the Fishers Linear Discriminant based distance (Pseudo averaging)

B.2.4 Jensen-Shannon Distance

Figures B.7 B.8 show the union of the ADWPT decompositions obtained using Jensen-Shannon distance.

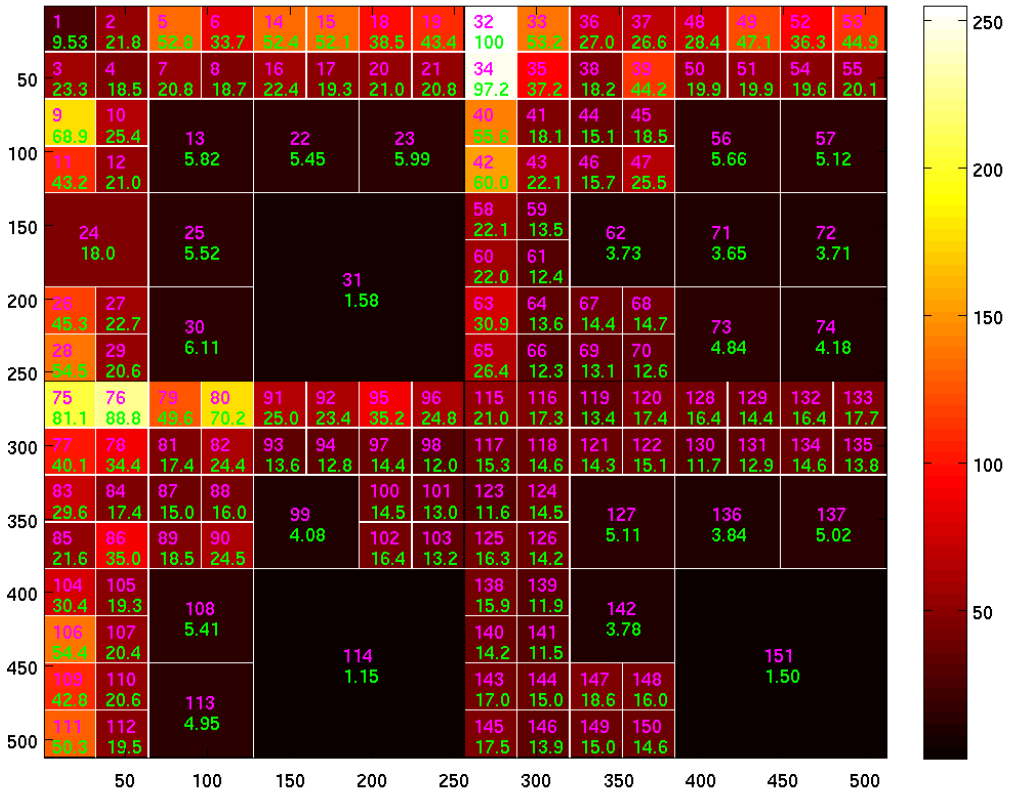


Figure B.7: Union of ADWPT decompositions obtained using the Jensen-Shannon distance (Standard averaging)

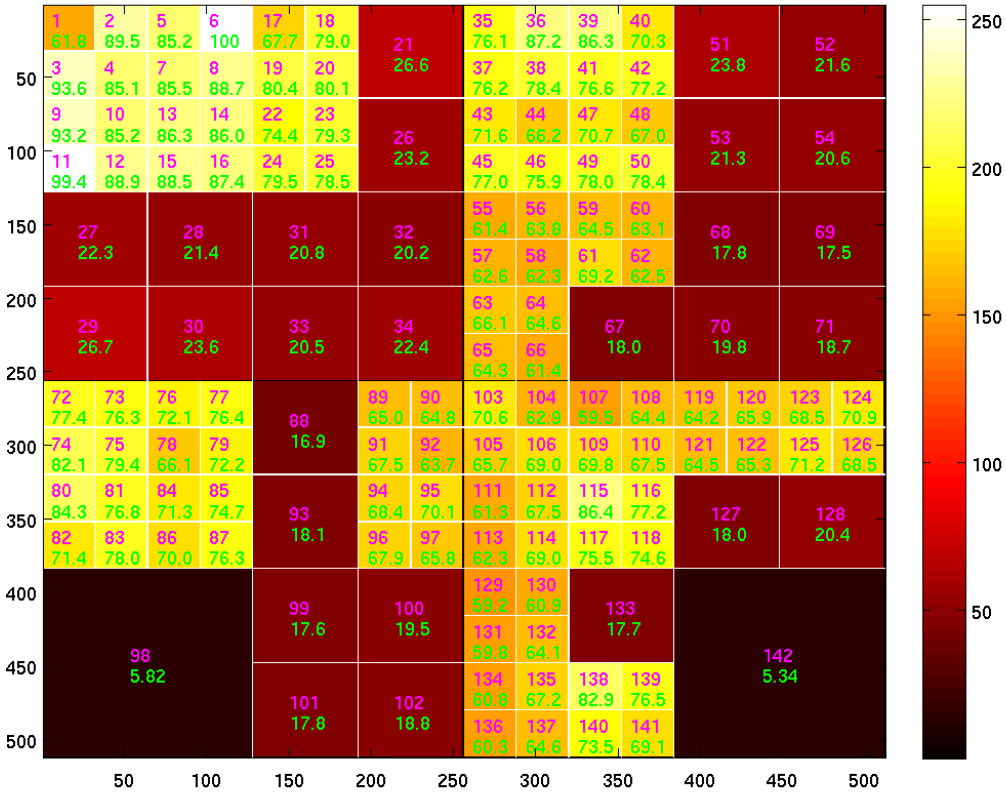


Figure B.8: Union of ADWPT decompositions obtained using the Jensen-Shannon distance (Pseudo averaging)

B.2.5 Bhattacharya Distance

Figures B.9 B.10 show the union of the ADWPT decompositions obtained using Bhattacharya distance.

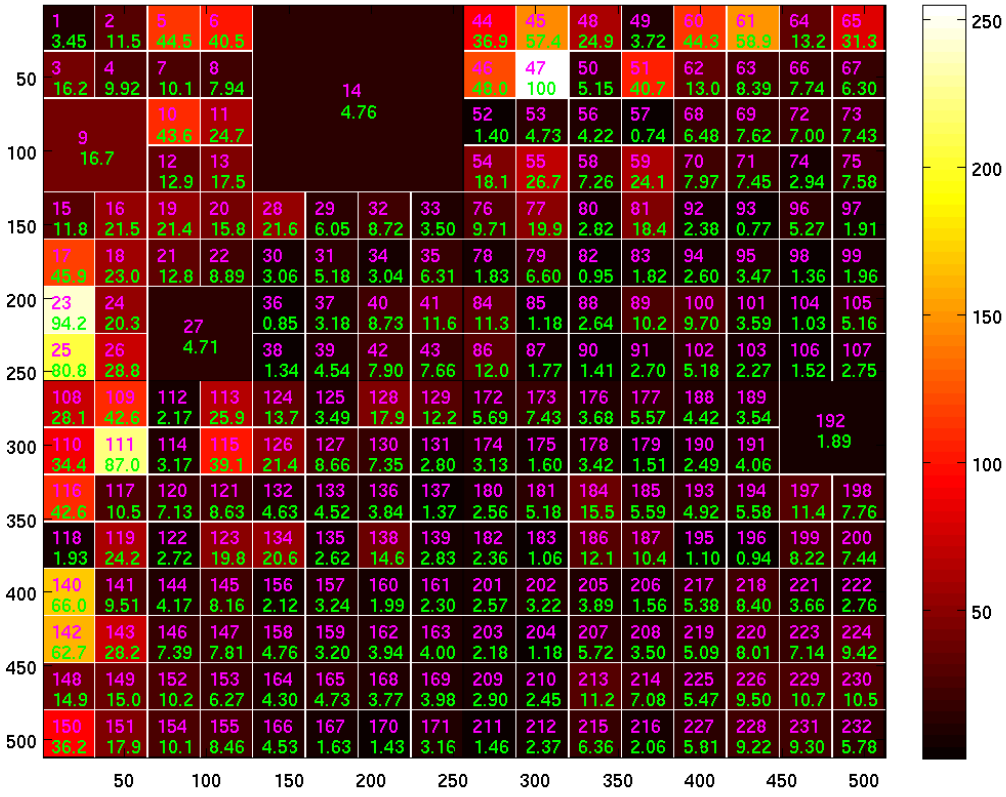


Figure B.9: Union of ADWPT decompositions obtained using the Bhattacharya distance (Standard averaging)

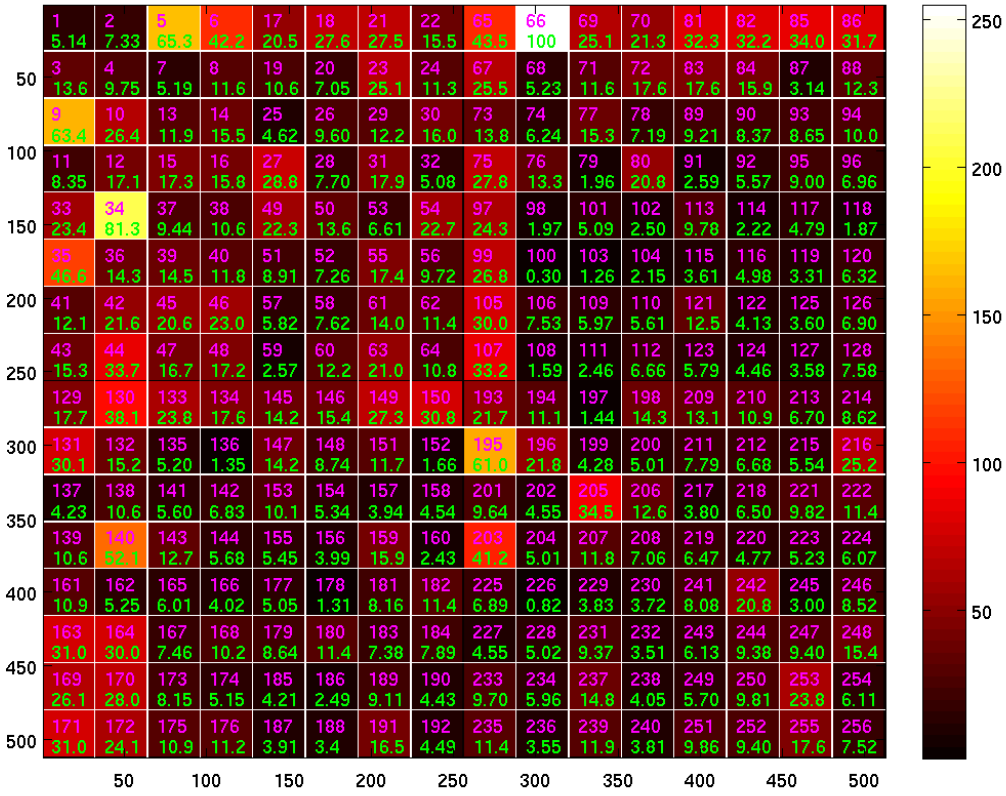
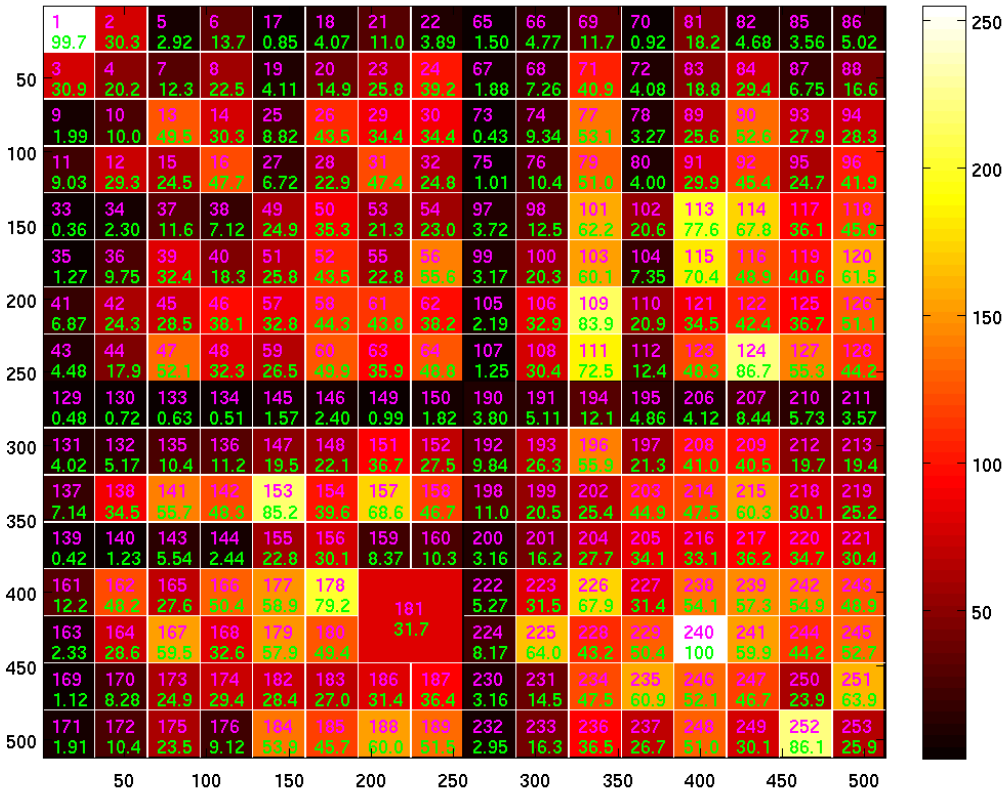


Figure B.10: Union of ADWPT decompositions obtained using the Bhattacharya distance (Pseudo averaging)

B.2.6 Mahalanobis Distance

Figures B.11 B.12 show the union of the ADWPT decompositions obtained using Mahalanobis distance.



a.

Figure B.11: Union of ADWPT decompositions obtained using the Mahalanobis distance (Standard averaging)

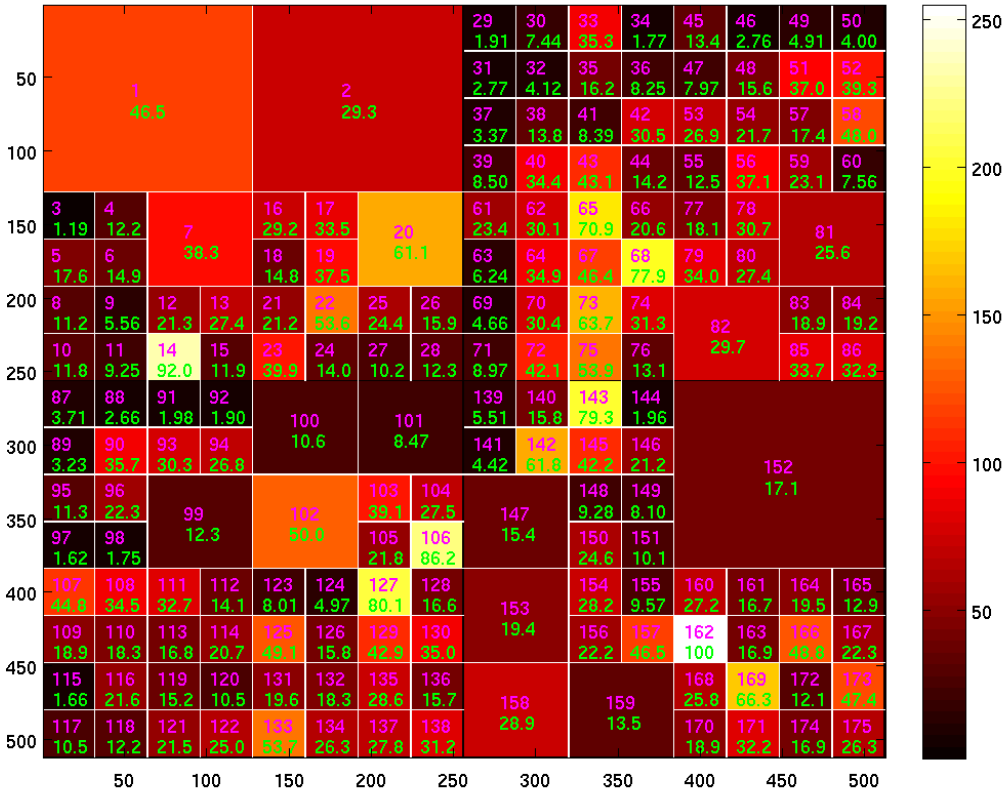


Figure B.12: Union of ADWPT decompositions obtained using the Mahalanobis distance (Pseudo averaging)

Figures B.13 B.14 show the union of the ADWPT decompositions obtained using Mahalanobis distance for multiple classes i.e. 4 in our case.

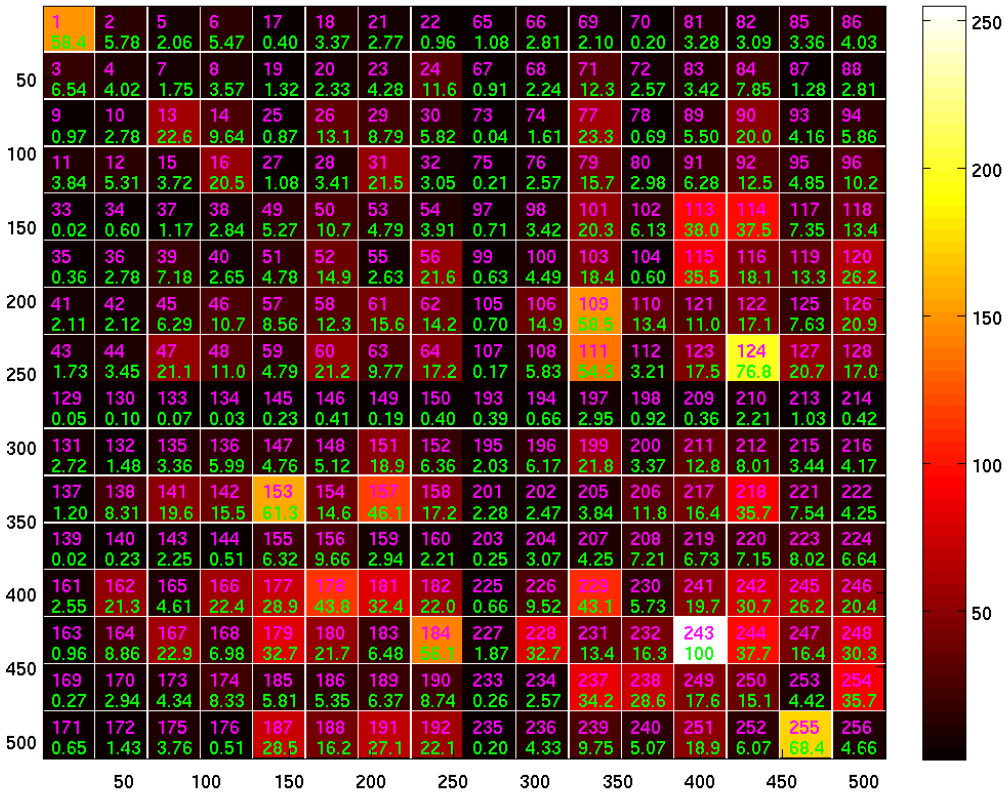


Figure B.13: Union of ADWPT decompositions obtained using the Multiple Class Mahalanobis distance (Standard averaging)

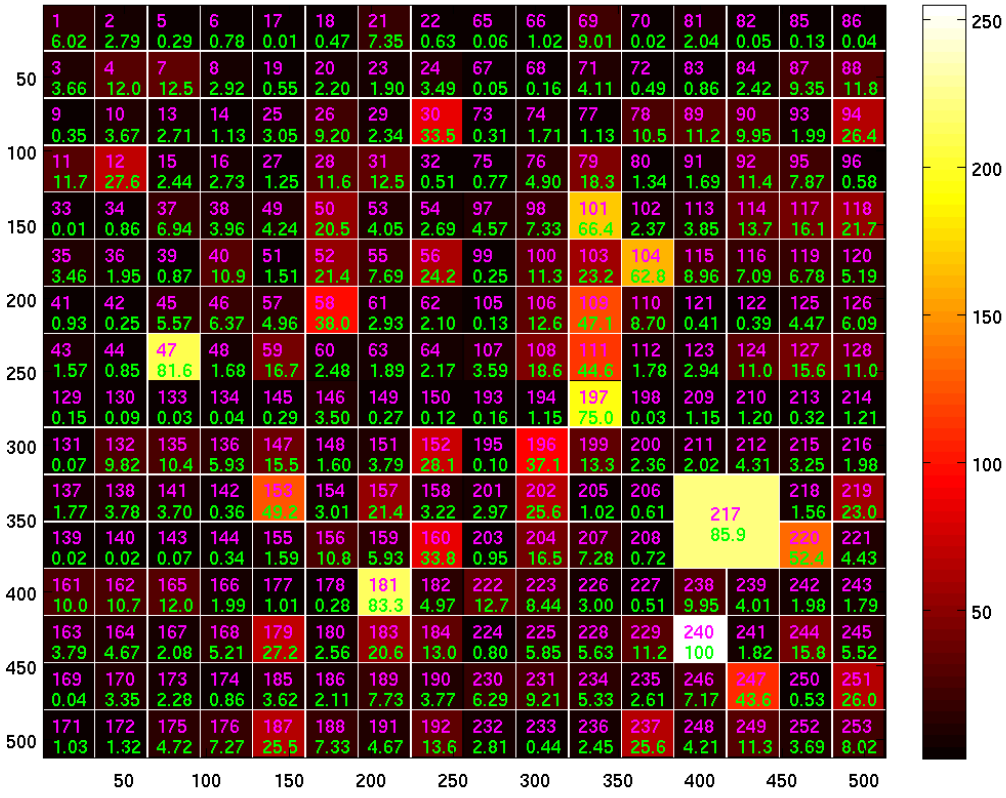


Figure B.14: Union of ADWPT decompositions obtained using the Multiple Class Mahalanobis distance (Pseudo averaging)

B.2.7 Energy Distance

Figurea B.15 B.16 show the union of the ADWPT decompositions obtained using Energy distance.

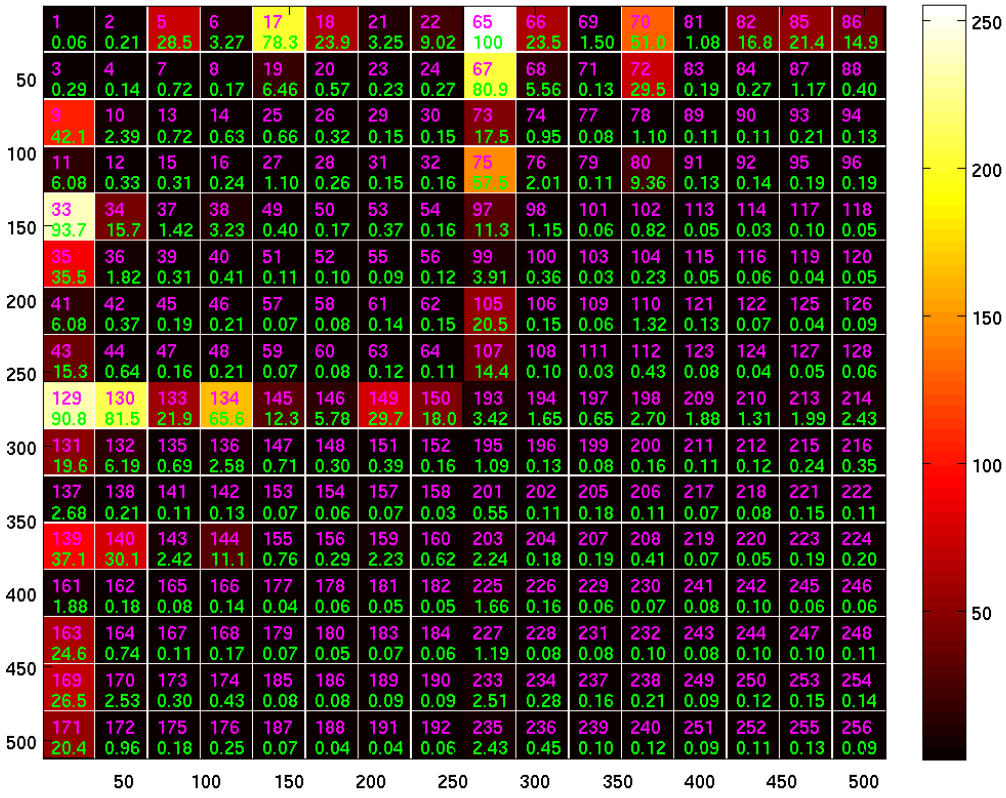


Figure B.15: Union of ADWPT decompositions obtained using the Energy distance (Standard averaging)

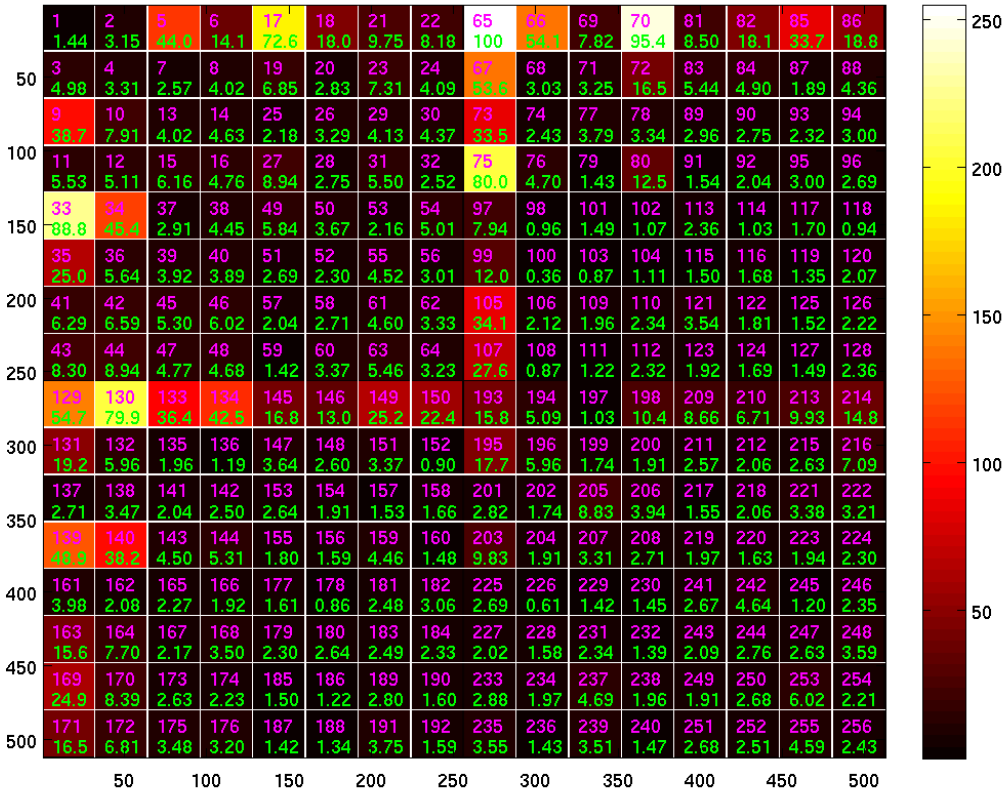


Figure B.16: Union of ADWPT decompositions obtained using the Energy distance (Pseudo averaging)

B.2.8 Relative Energy Distance

Figures B.17 B.18 show the union of the ADWPT decompositions obtained using the Relative Energy distance.

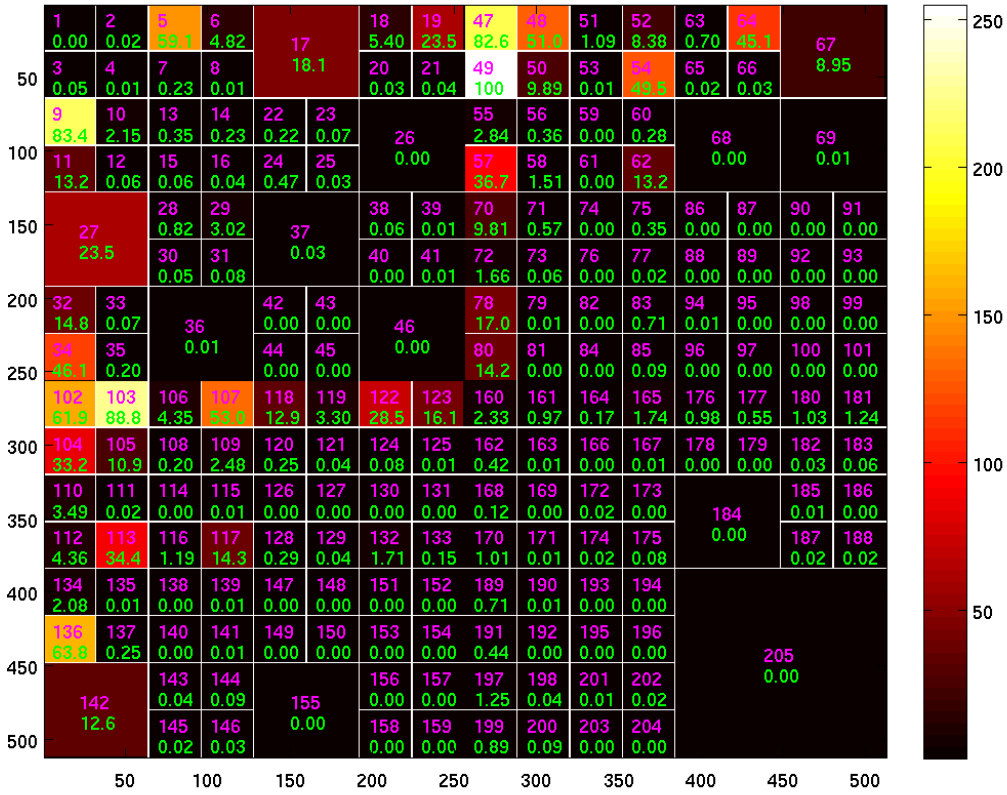


Figure B.17: Union of ADWPT decompositions obtained using the Relative Energy distance (Standard averaging)

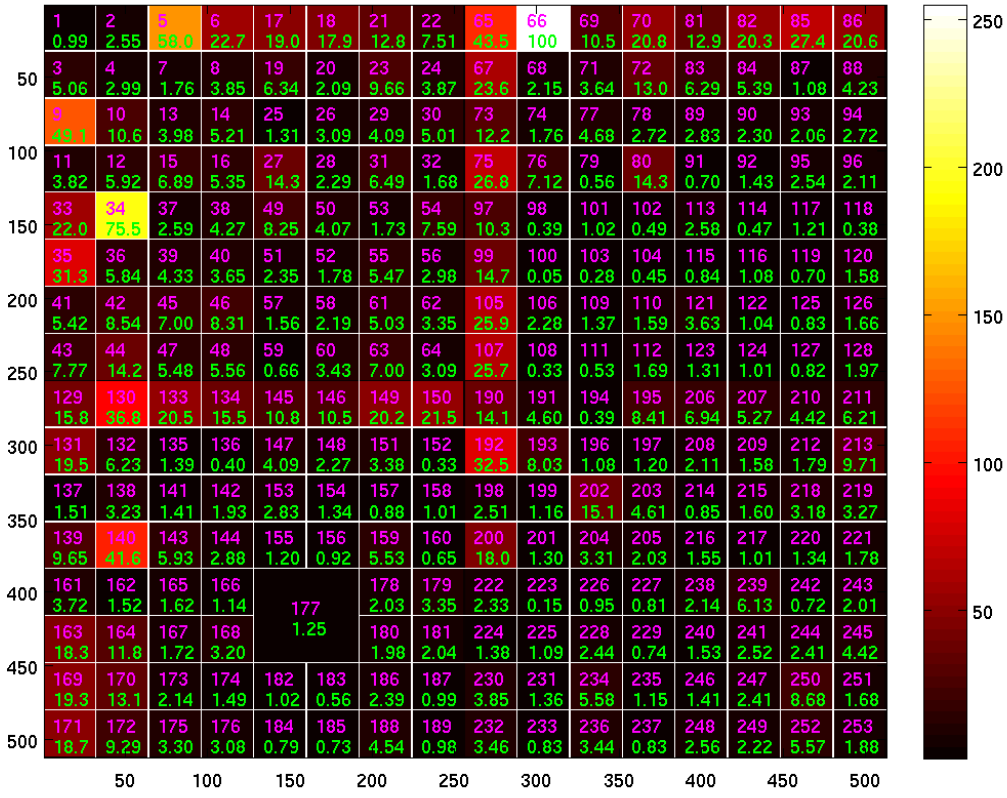


Figure B.18: Union of ADWPT decompositions obtained using the Relative Energy distance (Pseudo averaging)

B.3 Intersection of Decompositions

This section presents the intersection decompositions (defined in section 4.1.2) acquired for various distance functions.

B.3.1 Hellinger Distance

Figures B.19 B.20 show the intersection of the ADWPT decompositions obtained using Hellinger distance.

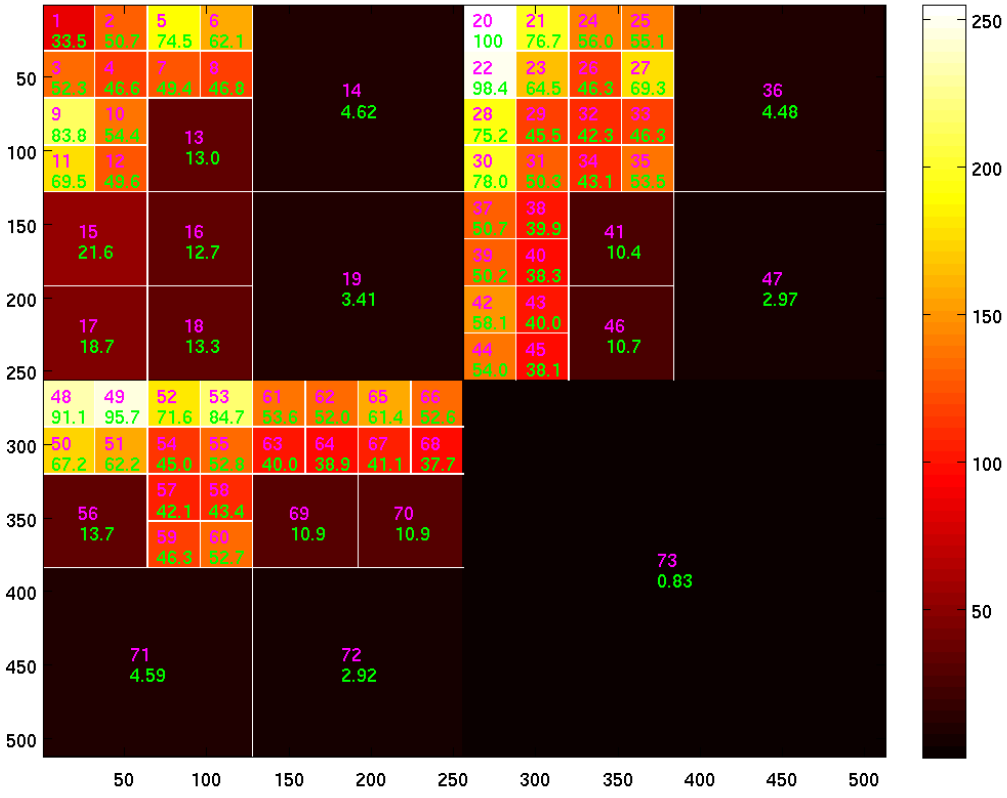


Figure B.19: Intersection of ADWPT decompositions obtained using the Hellinger distance (Standard averaging)

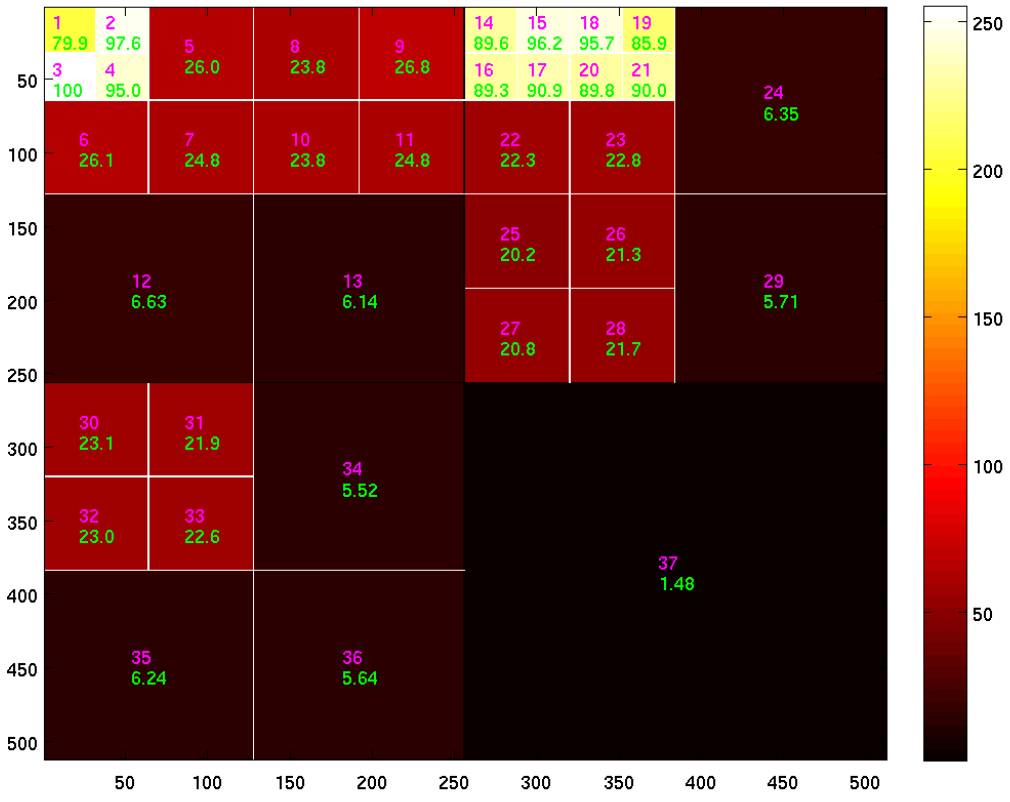


Figure B.20: Intersection of ADWPT decompositions obtained using the Hellinger distance (Pseudo averaging)

Again, the magenta number indicates the number of the subband when they are ordered in terms of increasing frequency content whereas the green number is the discriminant value of the subband concerned, scaled so that the maximum value is 100 and minimum 0.

B.3.2 Kullback-Leibler Distance

Figures B.21 B.22 show the intersection of the ADWPT decompositions obtained using Kullback-Leibler distance.

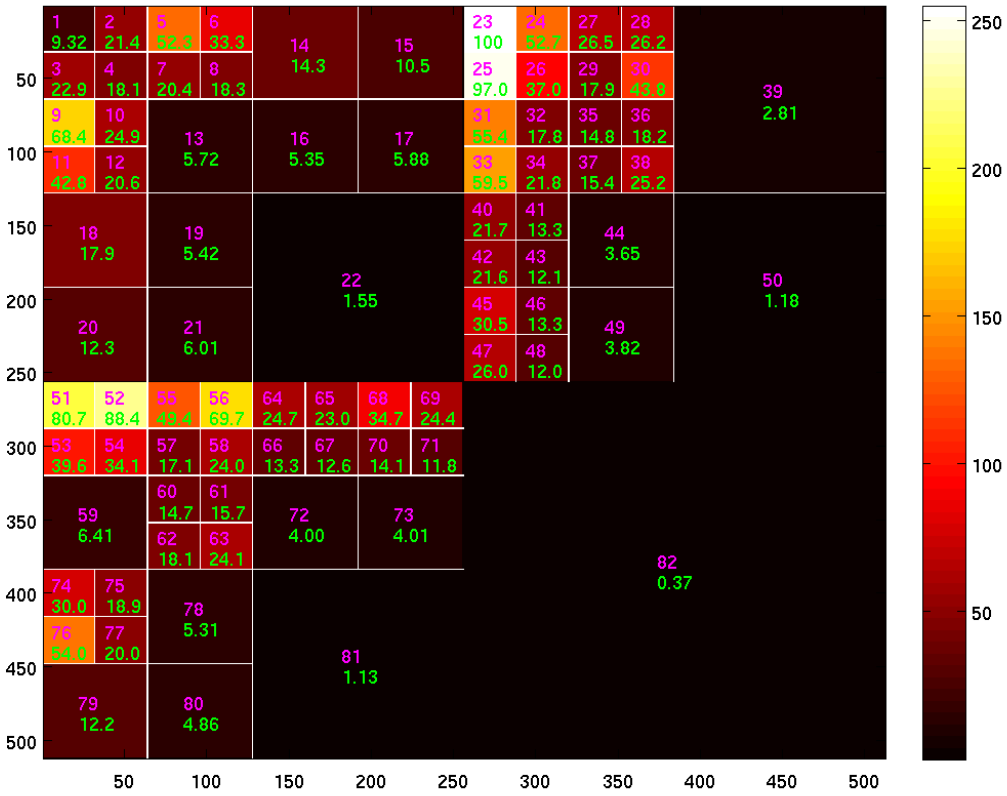


Figure B.21: Intersection of ADWPT decompositions obtained using the Kullback-Leibler distance (Standard averaging)

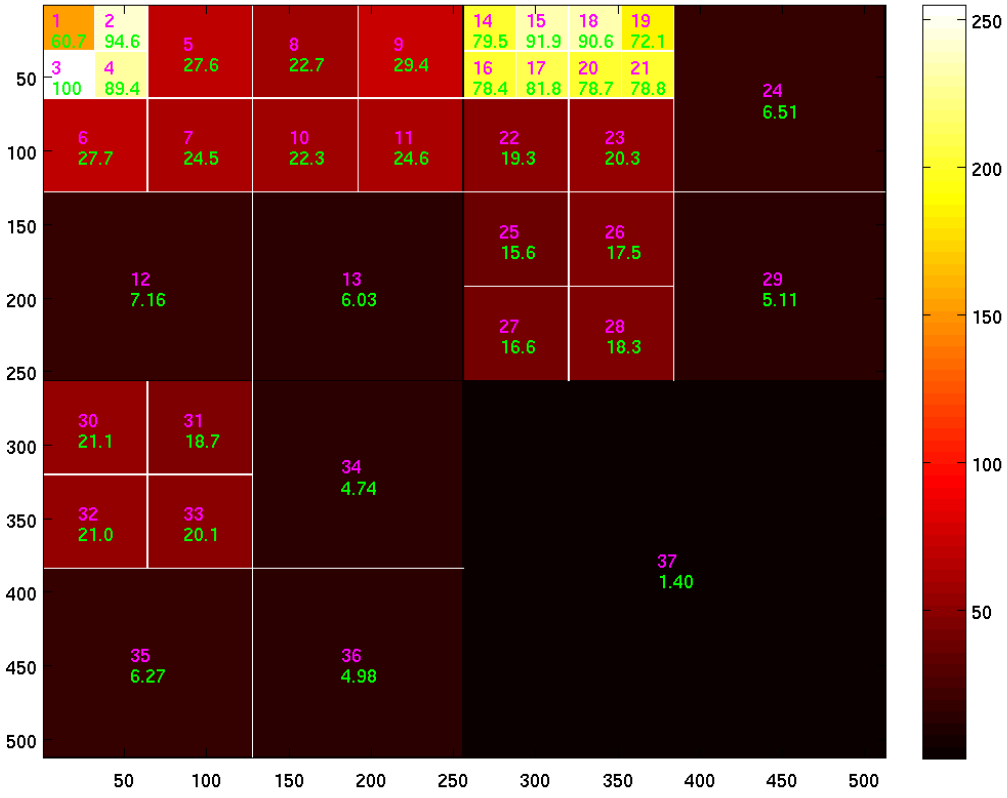


Figure B.22: Intersection of ADWPT decompositions obtained using the Kullback-Leibler distance (Pseudo averaging)

B.3.3 Fishers Linear Discriminant

Figures B.23 B.24 show the intersection of the ADWPT decompositions obtained using Fishers Linear Discriminant distance.

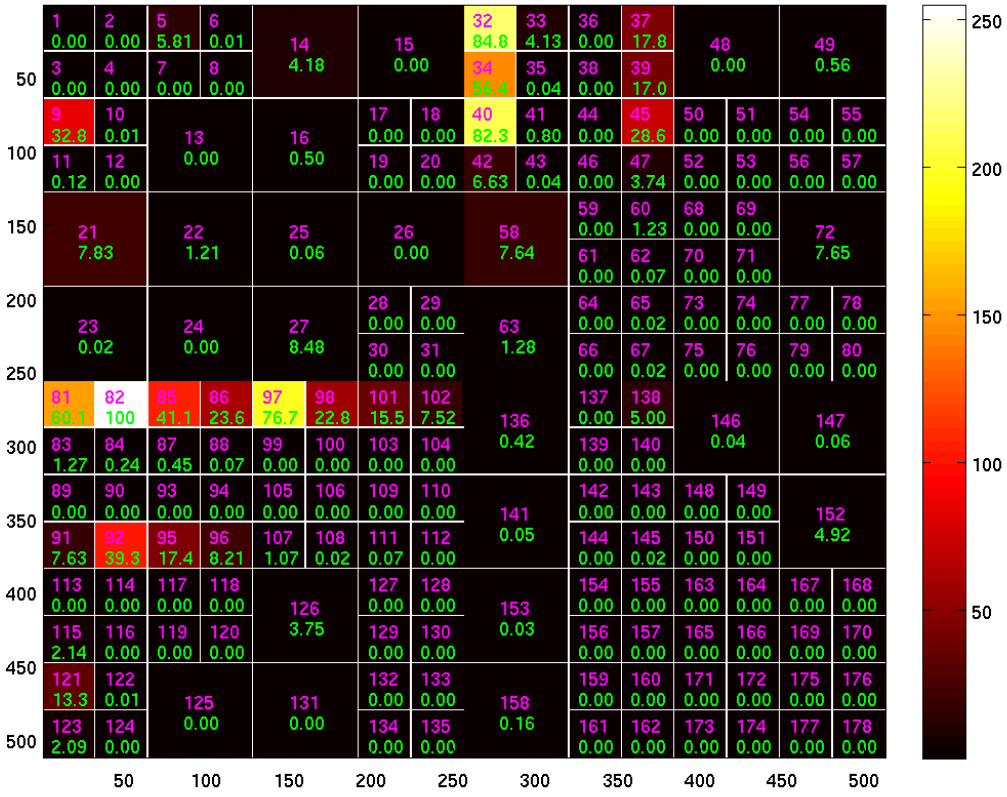


Figure B.23: Intersection of ADWPT decompositions obtained using the Fishers Linear Discriminant (Standard averaging)

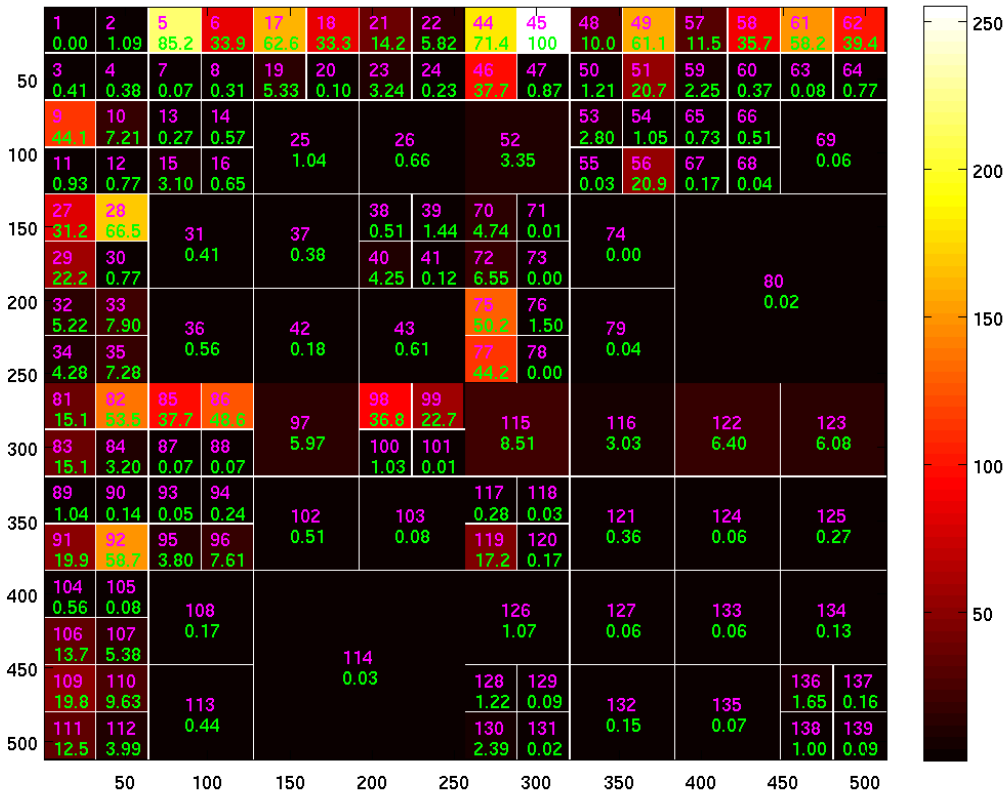


Figure B.24: Intersection of ADWPT decompositions obtained using the Fishers Linear Discriminant (Pseudo averaging)

B.3.4 Jensen-Shannon Distance

Figurea B.25 B.26 show the intersection of the ADWPT decompositions obtained using Jensen-Shannon distance.

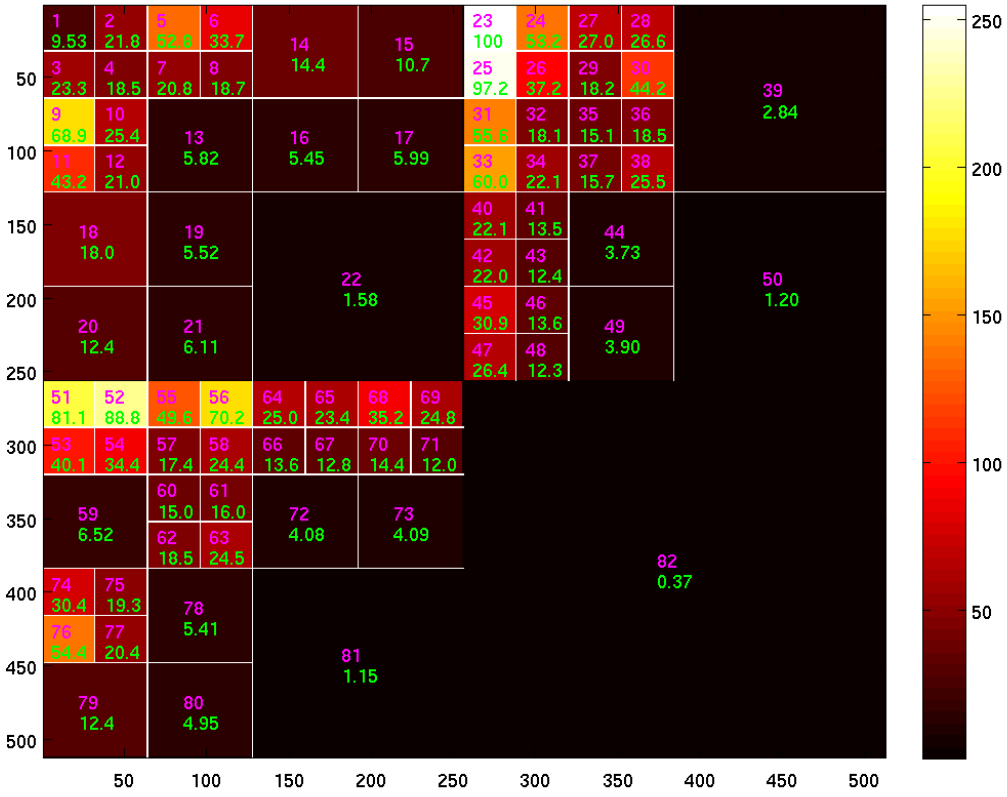


Figure B.25: Intersection of ADWPT decompositions obtained using the Jensen-Shannon distance (Standard averaging)

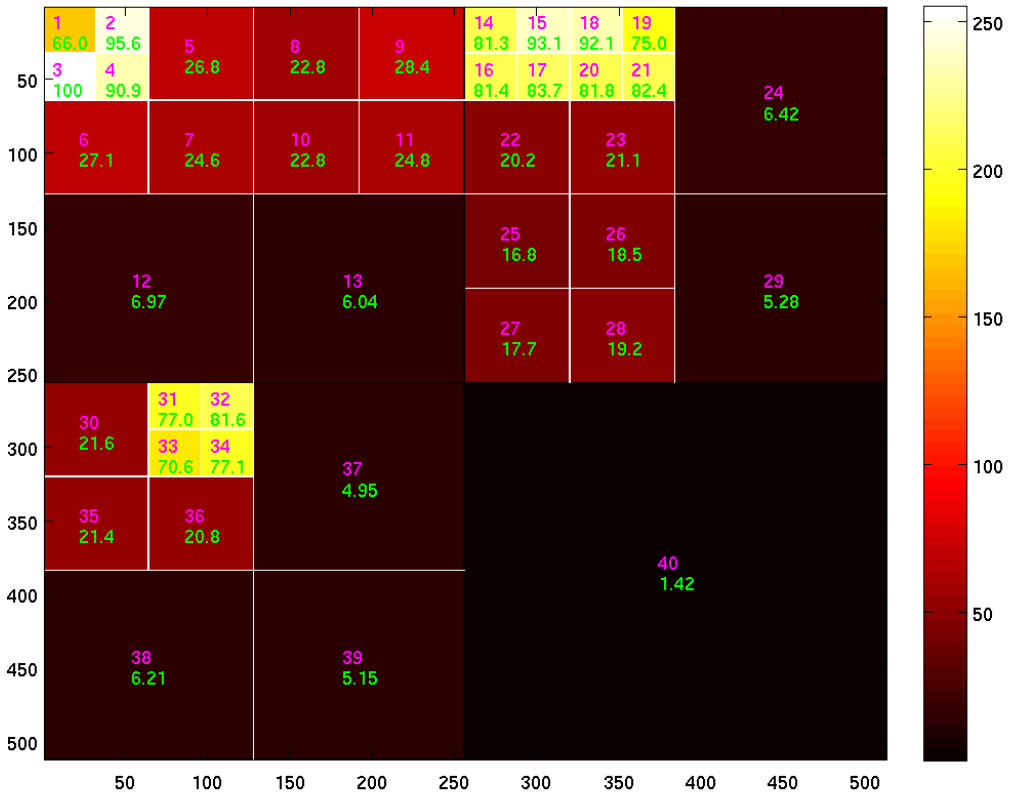


Figure B.26: Intersection of ADWPT decompositions obtained using the Jensen-Shannon distance (Pseudo averaging)

B.3.5 Bhattacharya Distance

Figures B.27 B.28 show the intersection of the ADWPT decompositions obtained using Bhattacharya distance.

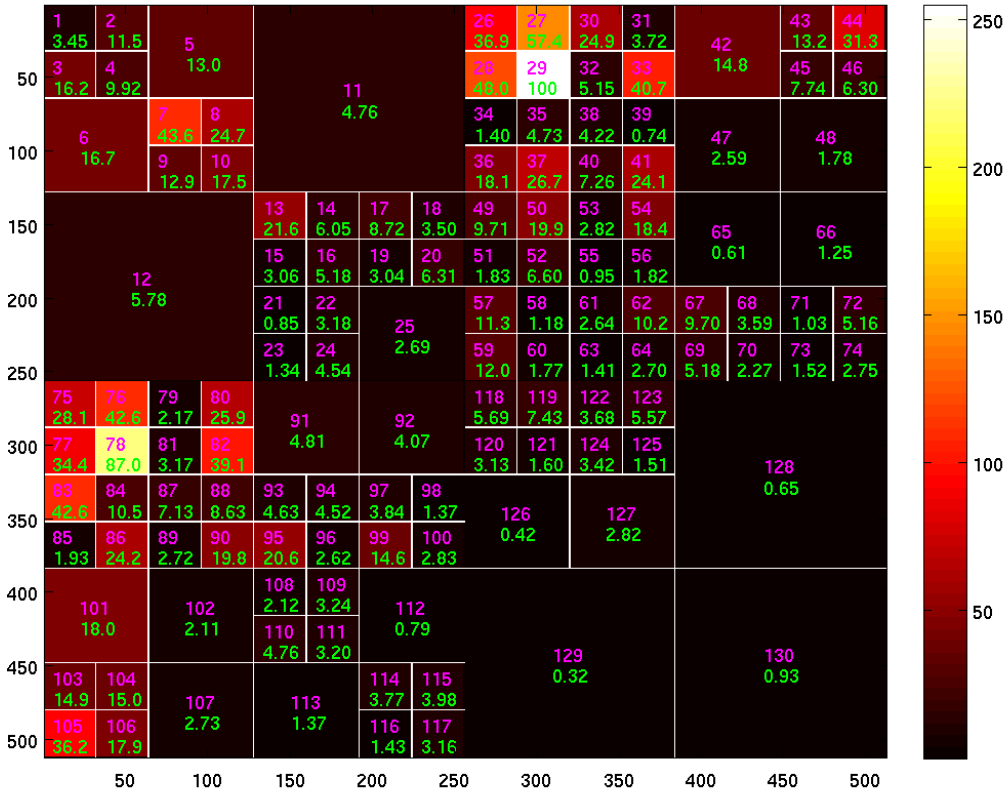


Figure B.27: Intersection of ADWPT decompositions obtained using the Bhattacharya distance (Standard averaging)

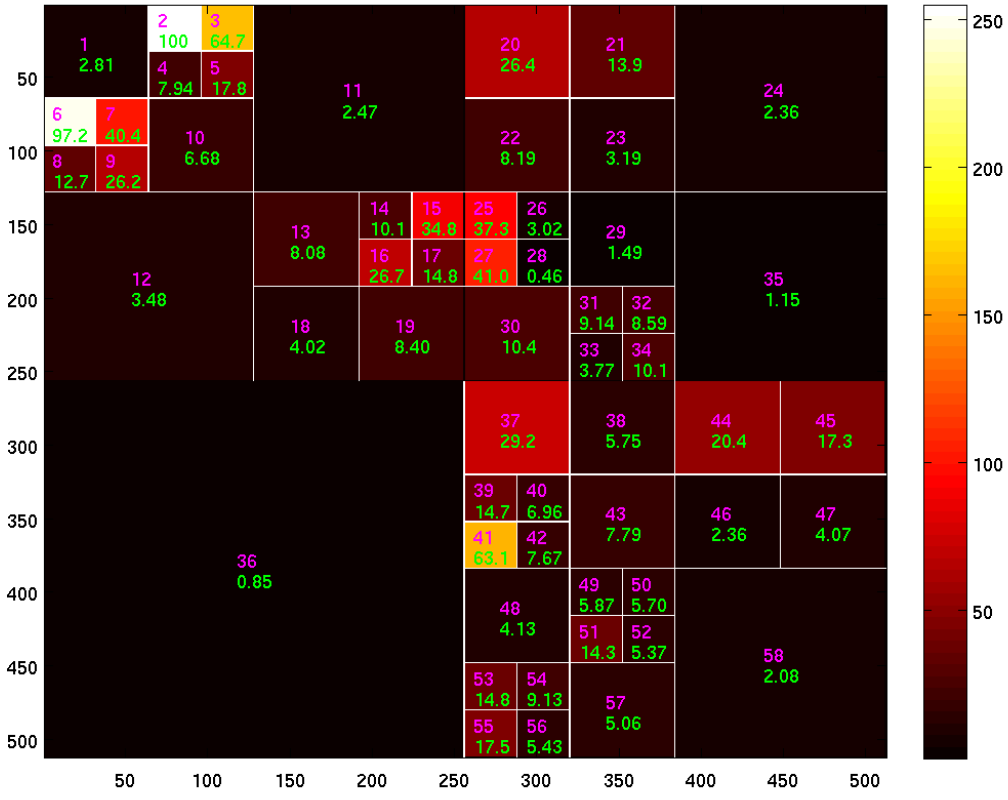


Figure B.28: Intersection of ADWPT decompositions obtained using the Bhattacharya distance (Pseudo averaging)

B.3.6 Mahalanobis Distance

Figures B.29 B.30 show the intersection of the ADWPT decompositions obtained using Mahalanobis distance.

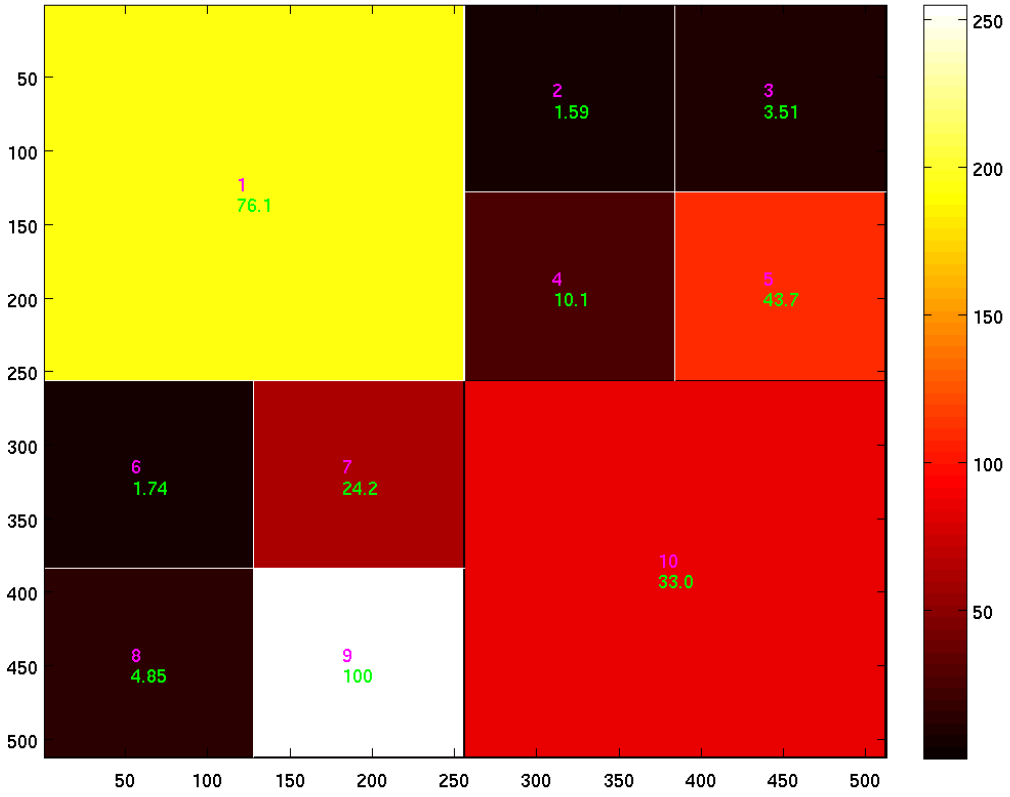


Figure B.29: Intersection of ADWPT decompositions obtained using the Mahalanobis distance (Standard averaging)

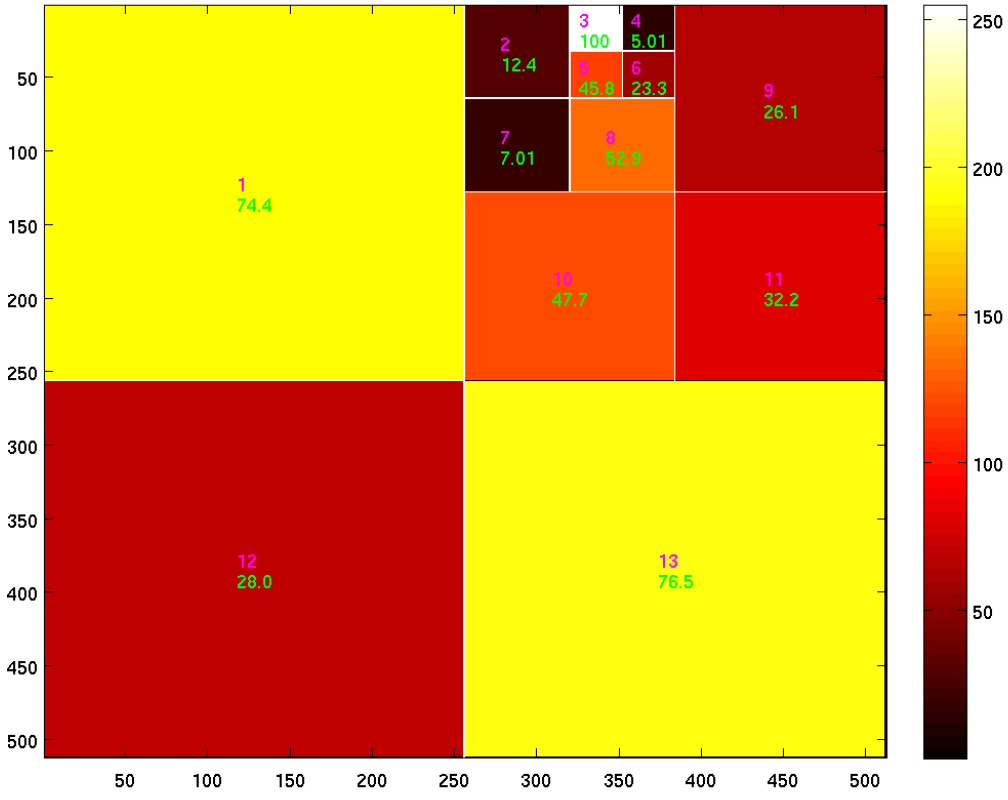


Figure B.30: Intersection of ADWPT decompositions obtained using the Mahalanobis distance (Pseudo averaging)

Figures B.31 B.32 show the intersection of the ADWPT decompositions obtained using Mahalanobis distance for multiple classes i.e. 4 in our case.

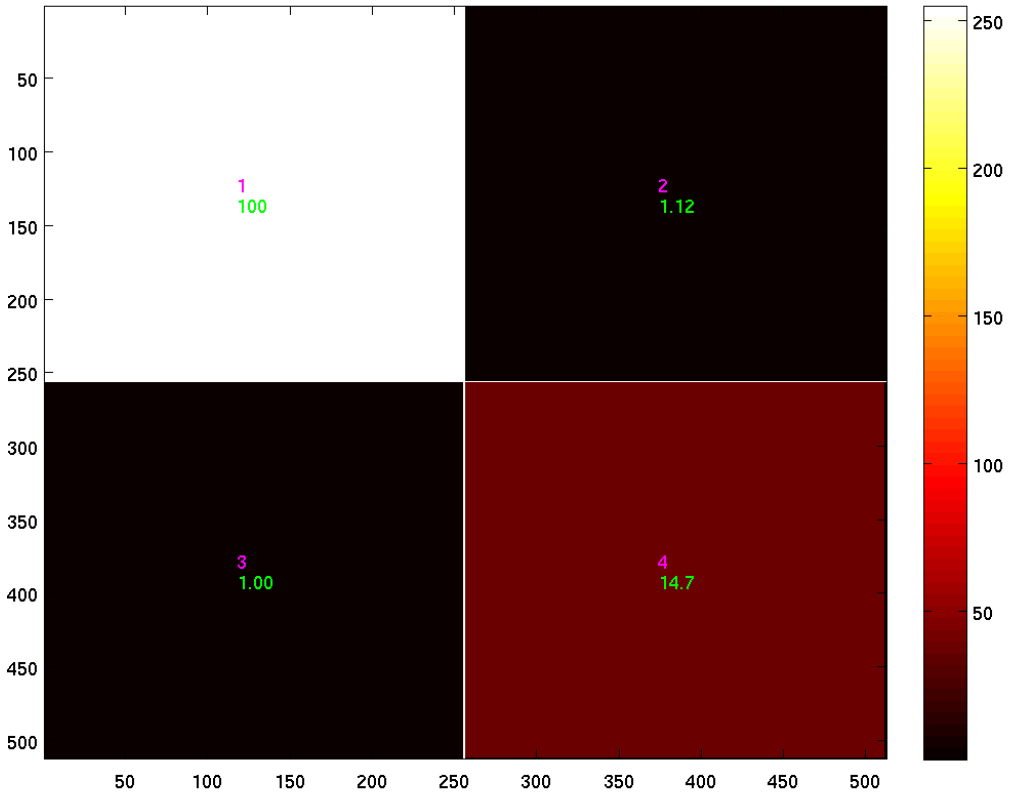


Figure B.31: Intersection of ADWPT decompositions obtained using the Multiple Class Mahalanobis distance (Standard averaging)

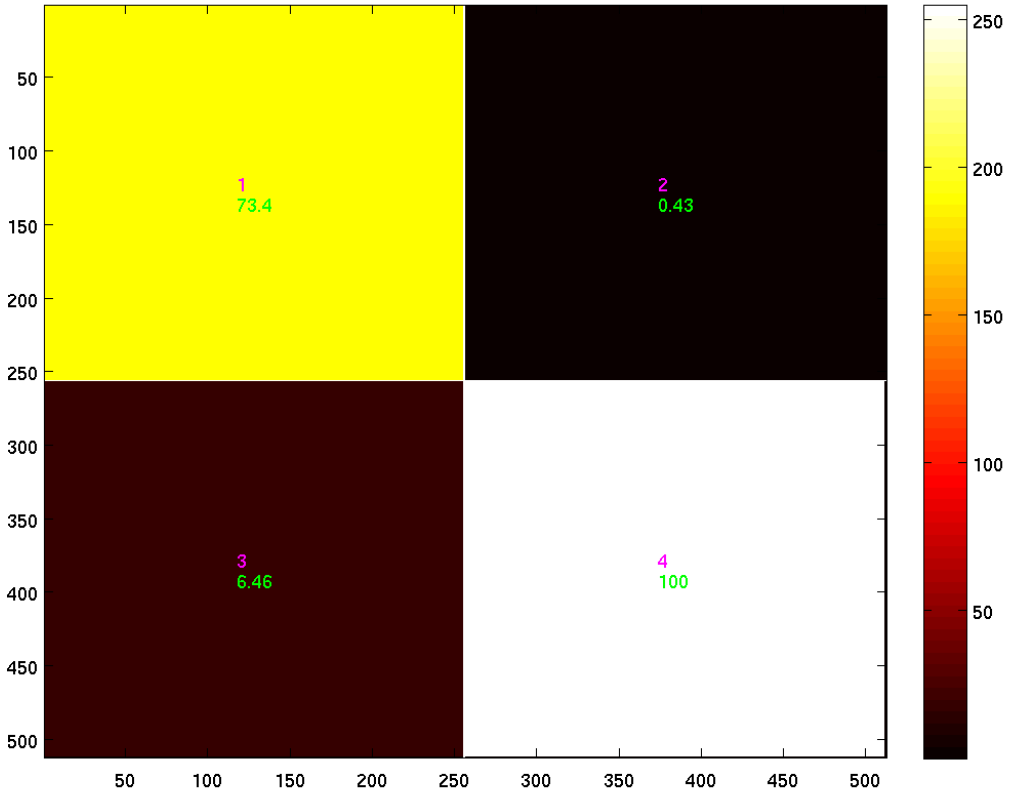


Figure B.32: Intersection of ADWPT decompositions obtained using the Multiple Class Mahalanobis distance (Pseudo averaging)

B.3.7 Energy Distance

Figures B.33 B.34 show the intersection of the ADWPT decompositions obtained using Energy distance.

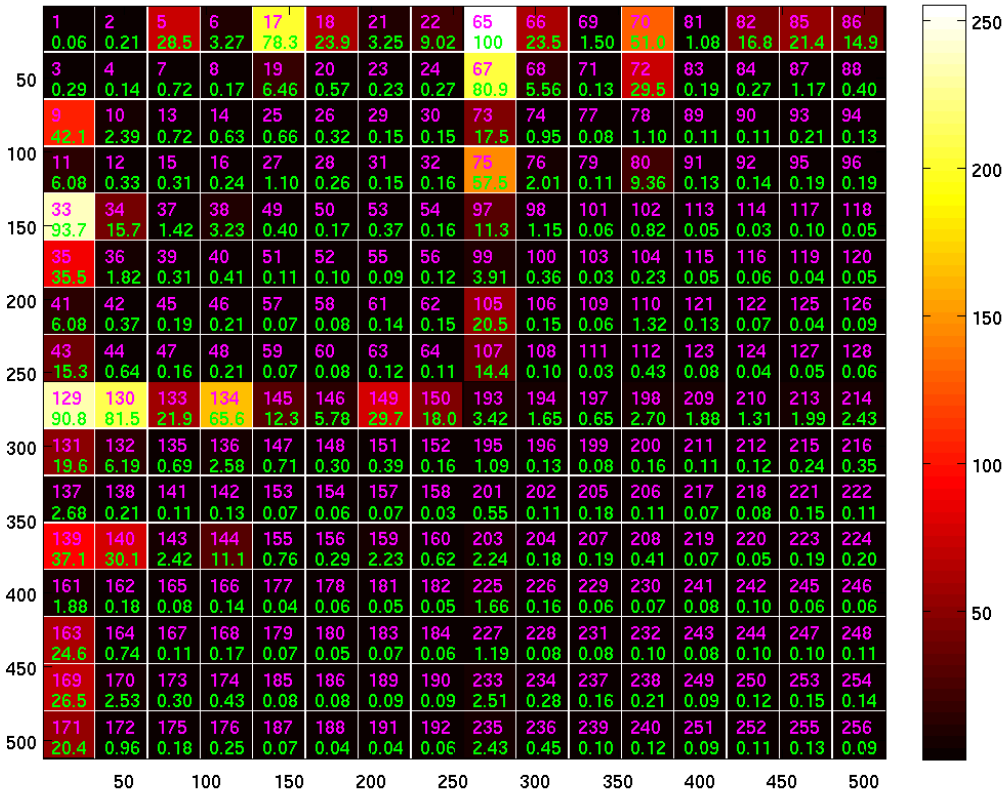


Figure B.33: Intersection of ADWPT decompositions obtained using the Energy distance (Standard averaging)

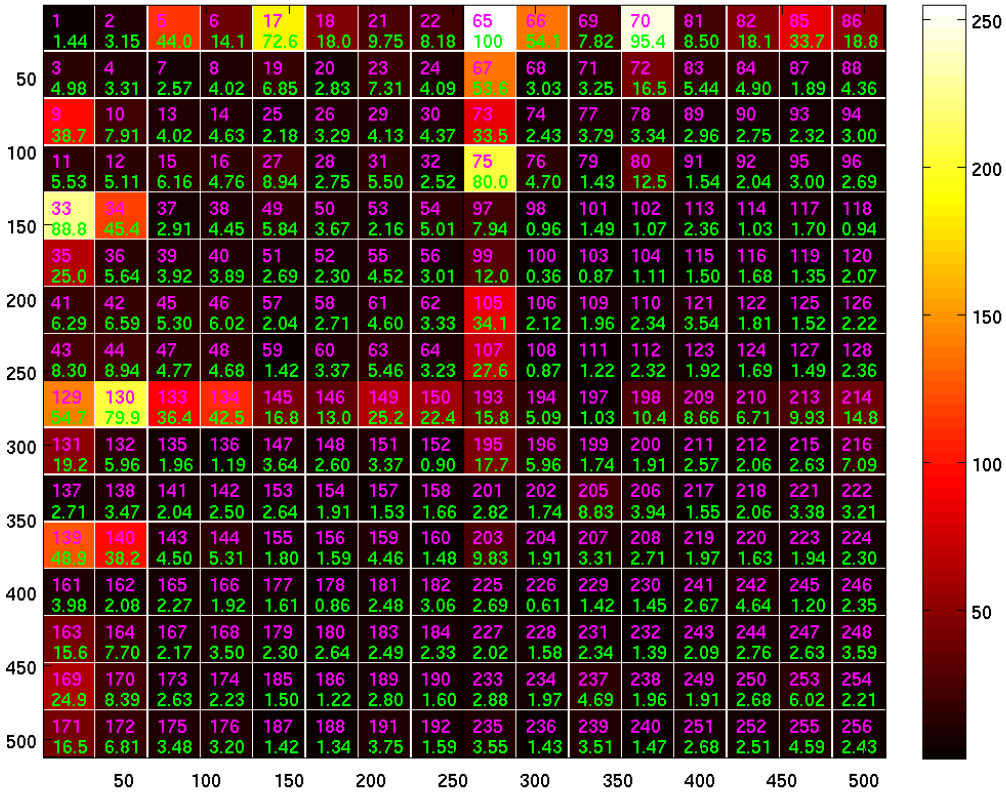


Figure B.34: Intersection of ADWPT decompositions obtained using the Energy distance (Pseudo averaging)

B.3.8 Relative Energy Distance

Figures B.35 B.36 show the intersection of the ADWPT decompositions obtained using Relative Energy distance.

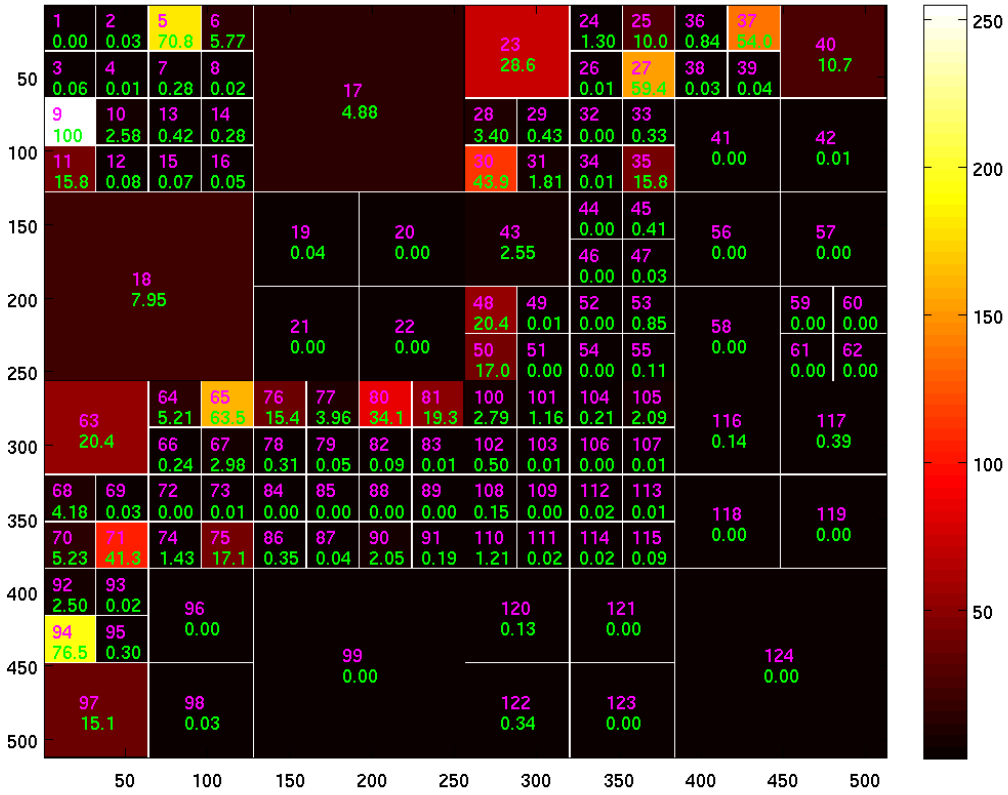


Figure B.35: Intersection of ADWPT decompositions obtained using the Relative Energy distance (Standard averaging)

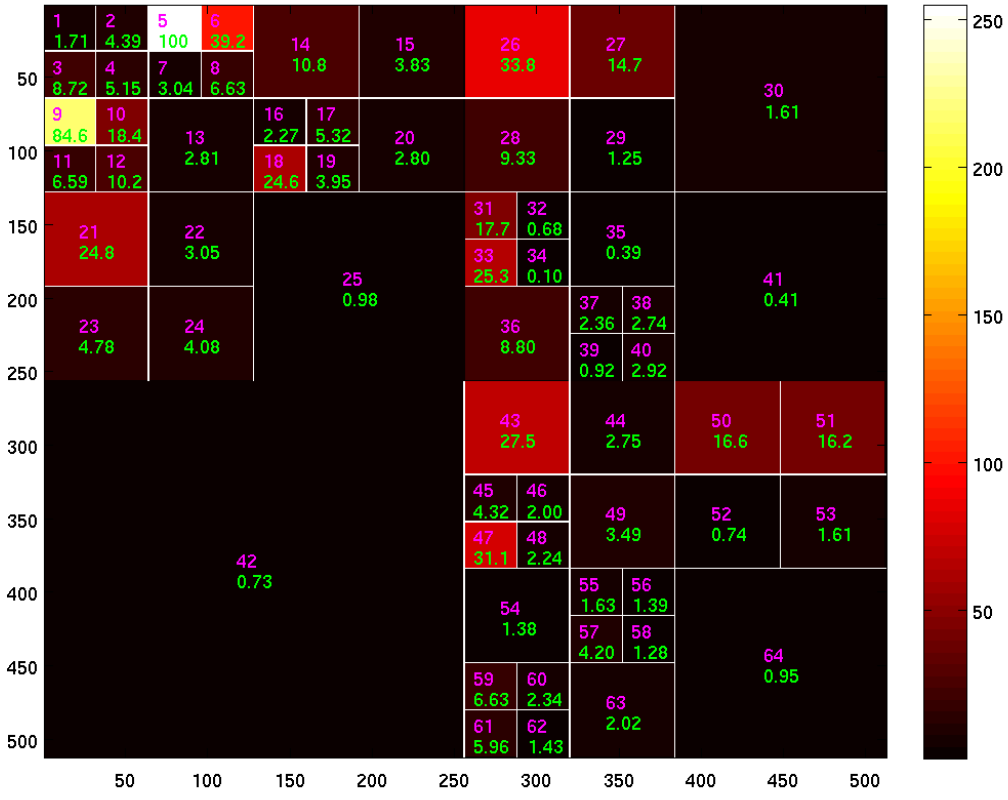


Figure B.36: Intersection of ADWPT decompositions obtained using the Relative Energy distance (Pseudo averaging)

B.4 Stability of a Decomposition

B.4.1 Standard Averaging

B.4.1.1 Kullback-Leibler Distance

A total of 114 different decompositions for Kullback-Leibler distance are computed in the case of standard averaging of the probability density estimates. Most of the decomposition again occur only once or twice. This can be seen in the stability decomposition. The stability decomposition obtained using the Kullback-Leibler distance is given in the Figure B.37.

B.4.1.2 Fishers-Linear Discriminant

As many as 316 different decompositions were obtained for Fishers Linear Discriminant incase of the standard averaging of the probability density estimates. Most of the decomposition occur only once or twice. The stability decomposition obtained using the Fishers Linear distance function is given in the Figure B.38.

B.4.1.3 Jensen-Shannon Distance

A total of 118 different decompositions for the Jensen-Shannon distance are computed in the case of standard averaging of the probability density estimates. Most of the decompositions occur only a few times and this could be seen from the stability decomposition. The stability decomposition obtained using the Jensen-Shannon distance is given in the Figure B.39.

B.4.1.4 Bhattacharya Distance

Using the Bhattacharya distance based analysis, our algorithm produces 557 different decompositions with each decomposition not repeating more than twice. Hence, the decompositions were very unstable. The stability decomposition obtained using the Bhattacharya distance is given in the Figure B.40.

B.4.1.5 Two Class Mahalanobis Distance

Again a great variety of decompositions, as many as 618, are obtained using the Mahalanobis distance over MAWTTs using standard averaging. Most of the

decompositions are not repeated i.e. they are only found once. The stability of the decomposition using the two class Mahalanobis distance is given in the Figure B.41.

B.4.1.6 Multi-class Mahalanobis Distance

Mahalanobis distance in the multi-class mode produces 619 different decompositions. The stability of the decomposition using the Multi-class Mahalanobis distance is given in the Figure B.42.

B.4.1.7 Energy Distance

The Energy distance again produces only one decomposition which is equivalent to the Full Wavelet Packet Transform. The stability decomposition obtained using the Energy distance is given in the Figure B.43.

B.4.1.8 Relative Energy Distance

The Relative Energy distance produces 340 different decompositions. The stability decomposition obtained using the Relative Energy distance is given in the Figure B.44.

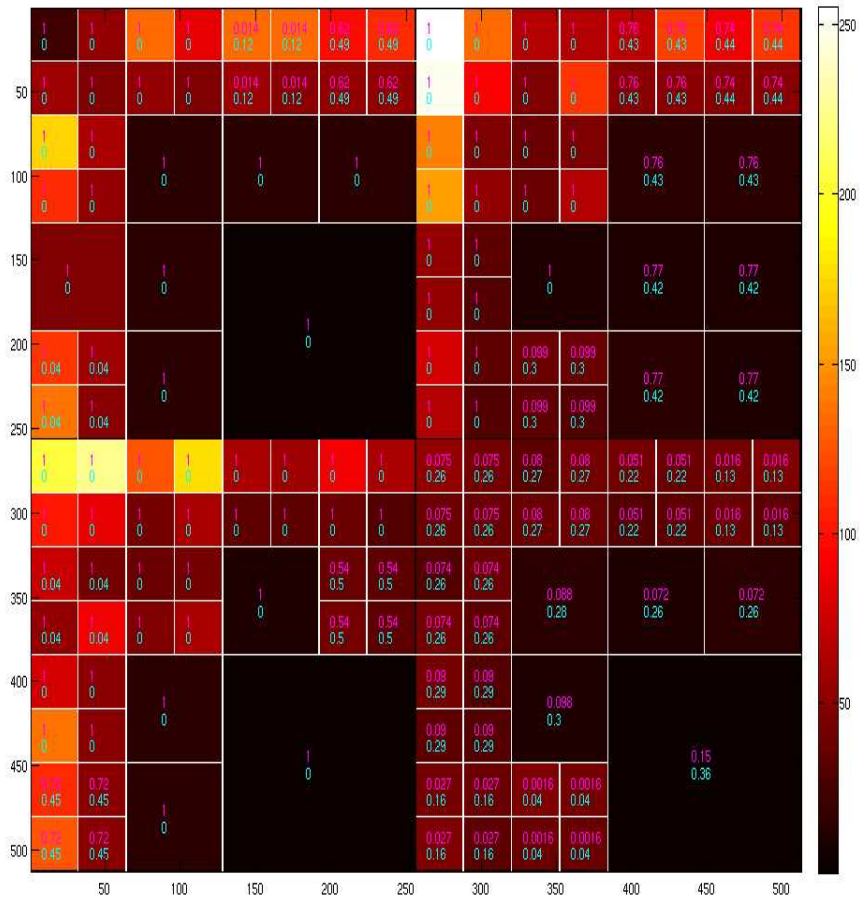


Figure B.37: Decomposition stability obtained using the Kullback-Leibler distance function over Standard Averaging based MAWTTs

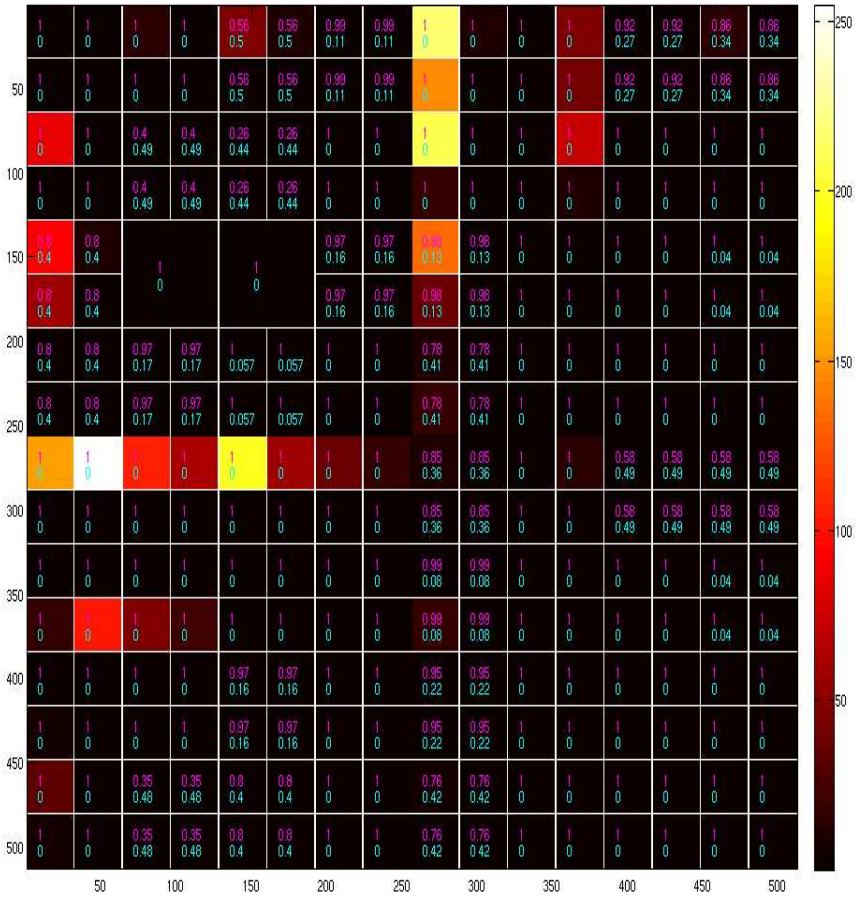


Figure B.38: Decomposition stability obtained using the Fishers Linear distance function over Standard Averaging based MAWTTs

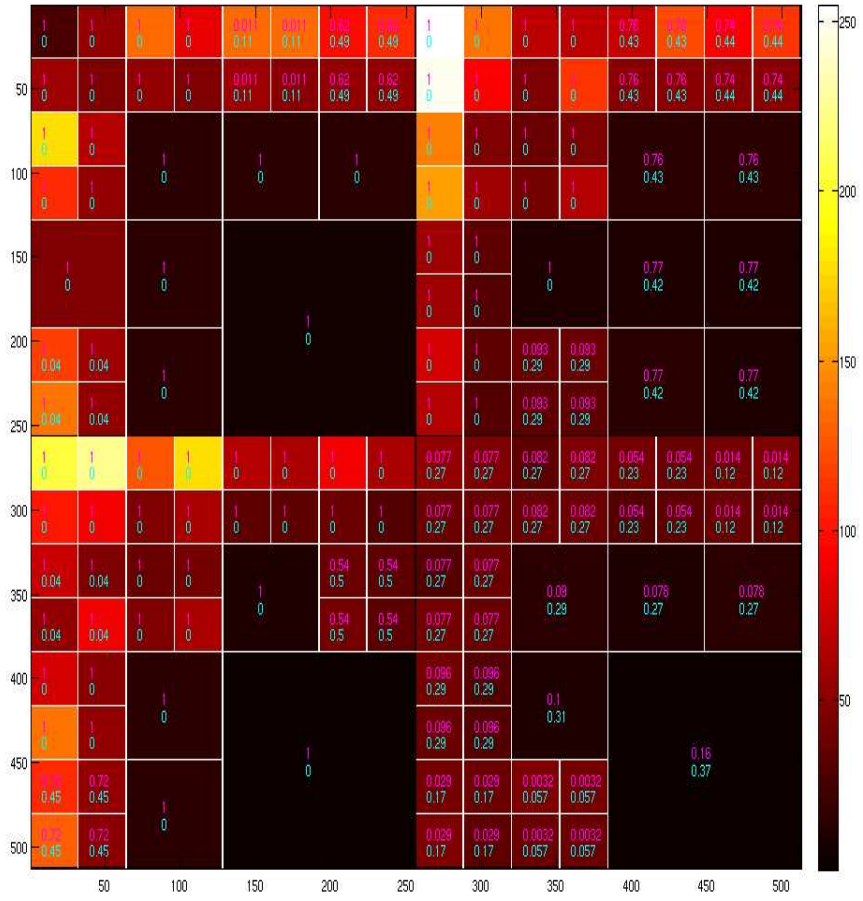


Figure B.39: Decomposition stability obtained using the Jensen-Shannon distance function over Standard Averaging based MAWTTs

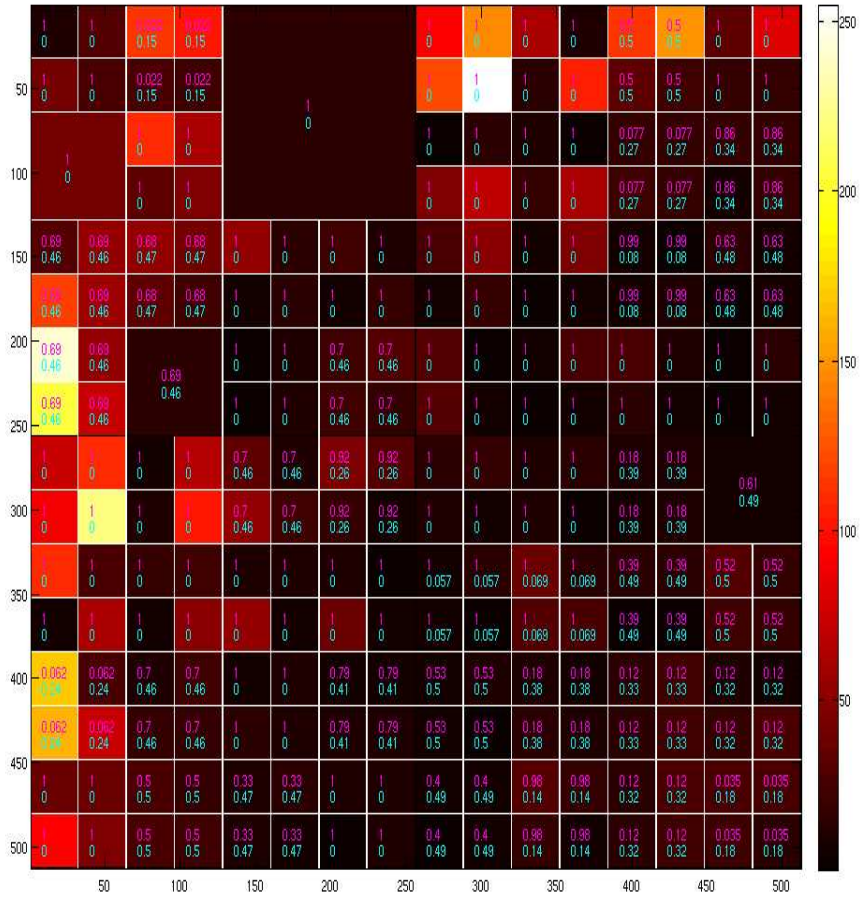


Figure B.40: Decomposition stability obtained using the Bhattacharya distance function over Standard Averaging based MAWTTs

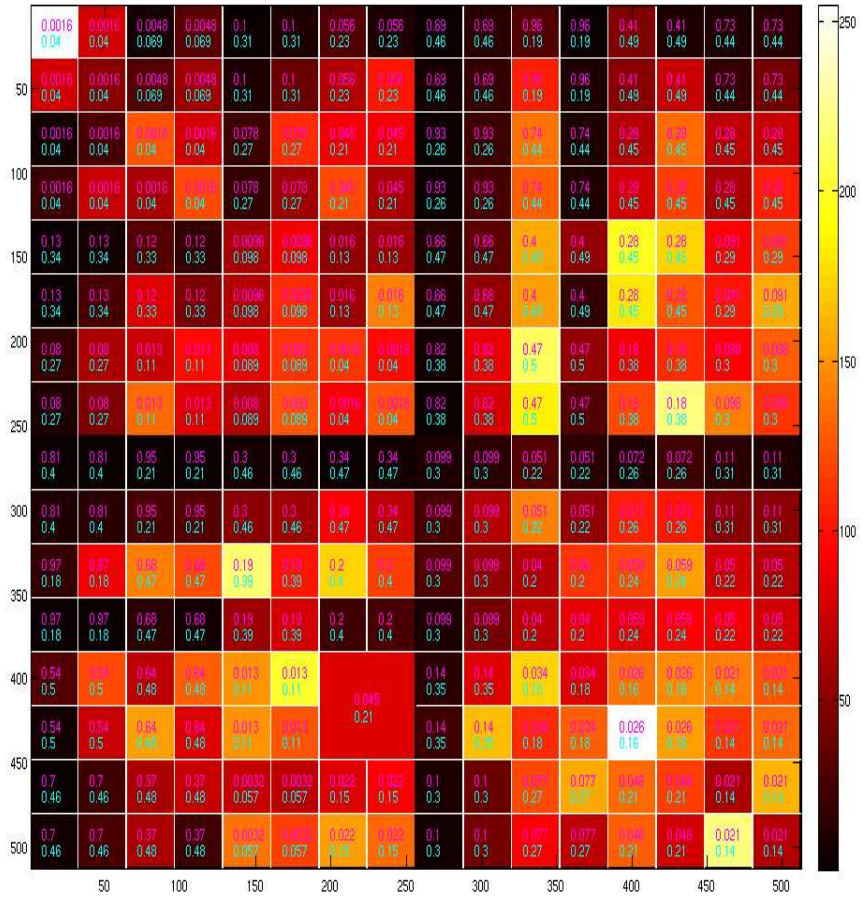


Figure B.41: Decomposition stability obtained using the Mahalanobis distance function over Standard Averaging based MAWTTs

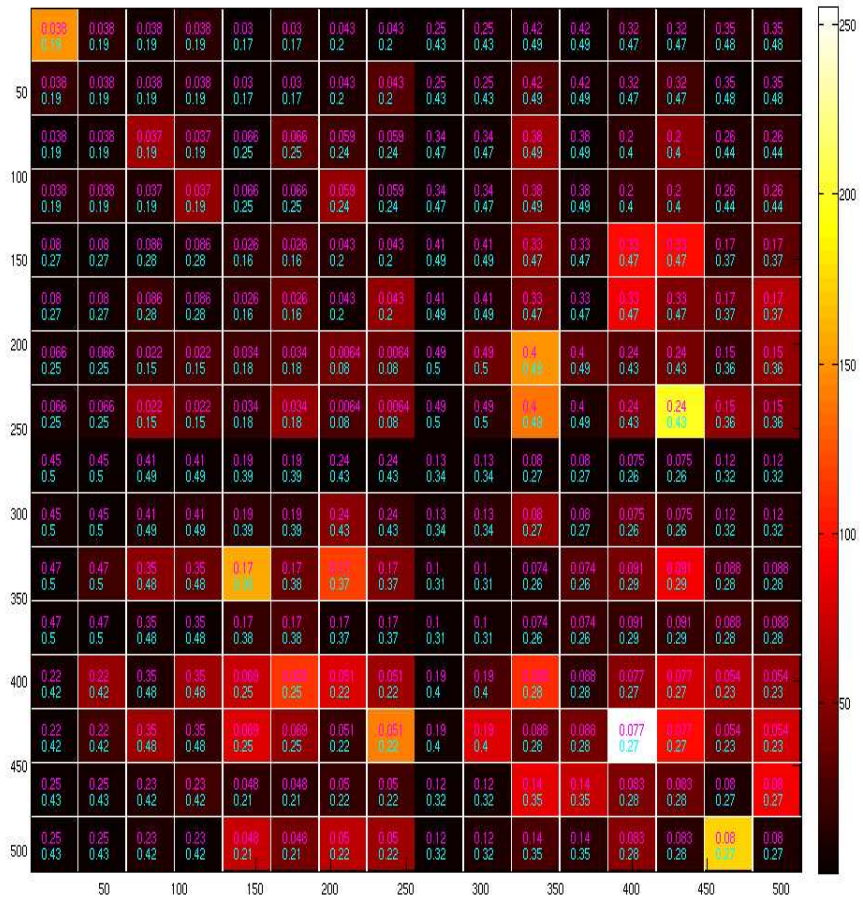


Figure B.42: Decomposition stability obtained using the Multi-class Mahalanobis distance function over Standard Averaging based MAWTTs

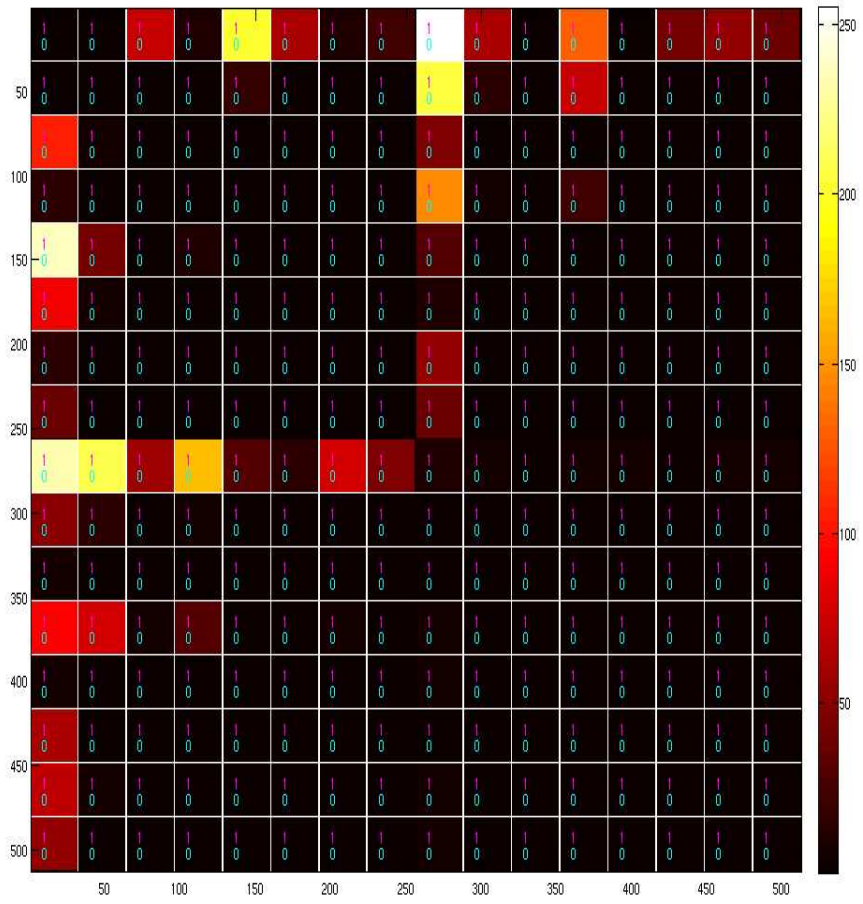


Figure B.43: Decomposition stability obtained using the Energy distance function over Standard Averaging based MAWTTs

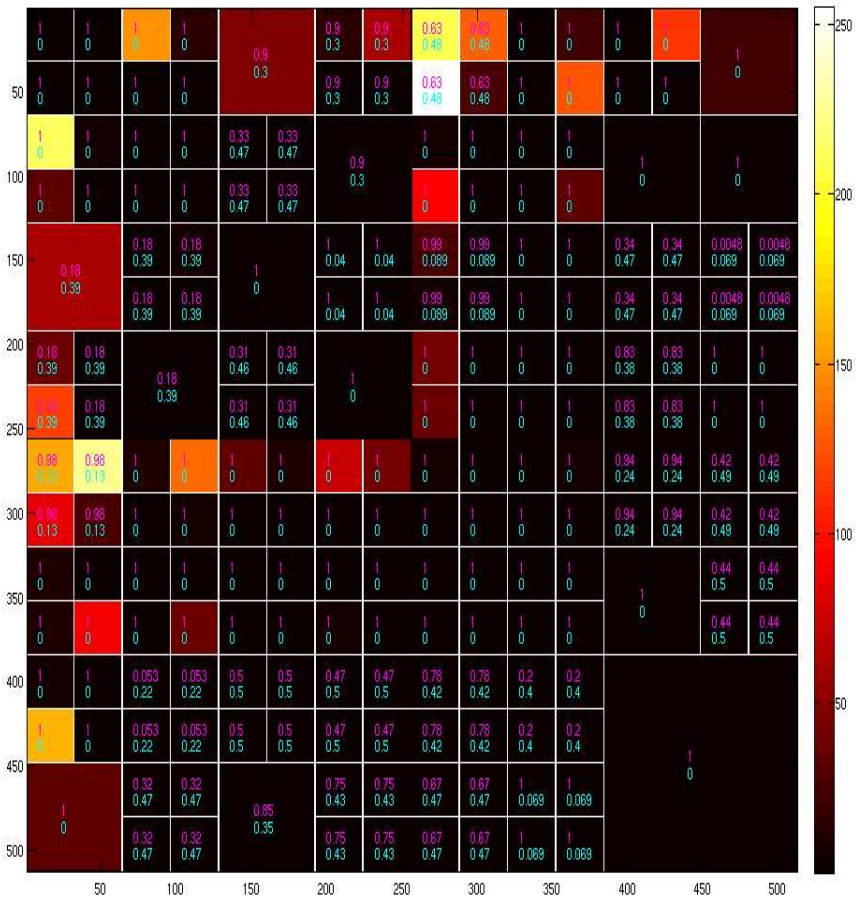


Figure B.44: Decomposition stability obtained using the Relative Energy Subbands distance function over Standard Averaging based MAWTTs

B.4.2 Pseudo-averaging

B.4.2.1 Kullback-Leibler Distance

A total of 16 different decompositions are obtained for the Kullback-Leibler distance function. The Figure B.45 displays the stability decomposition for Kullback-Leibler distance function.

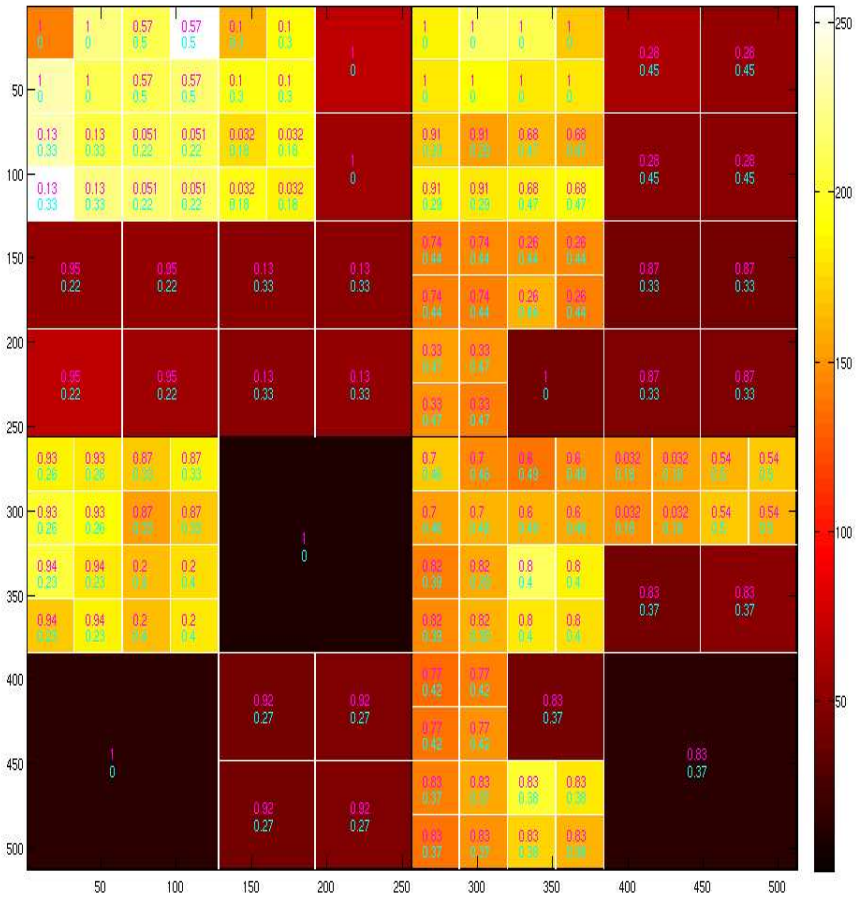


Figure B.45: Decomposition stability obtained using the Kullback Leibler distance over Pseudo-averaging based MAWTTs

B.4.2.2 Fishers Linear Discriminant

The Fishers Linear Discriminant produces a total of 16 different decompositions. The stability decomposition obtained using the Fishers Linear Discriminant is given in the Figure B.46.

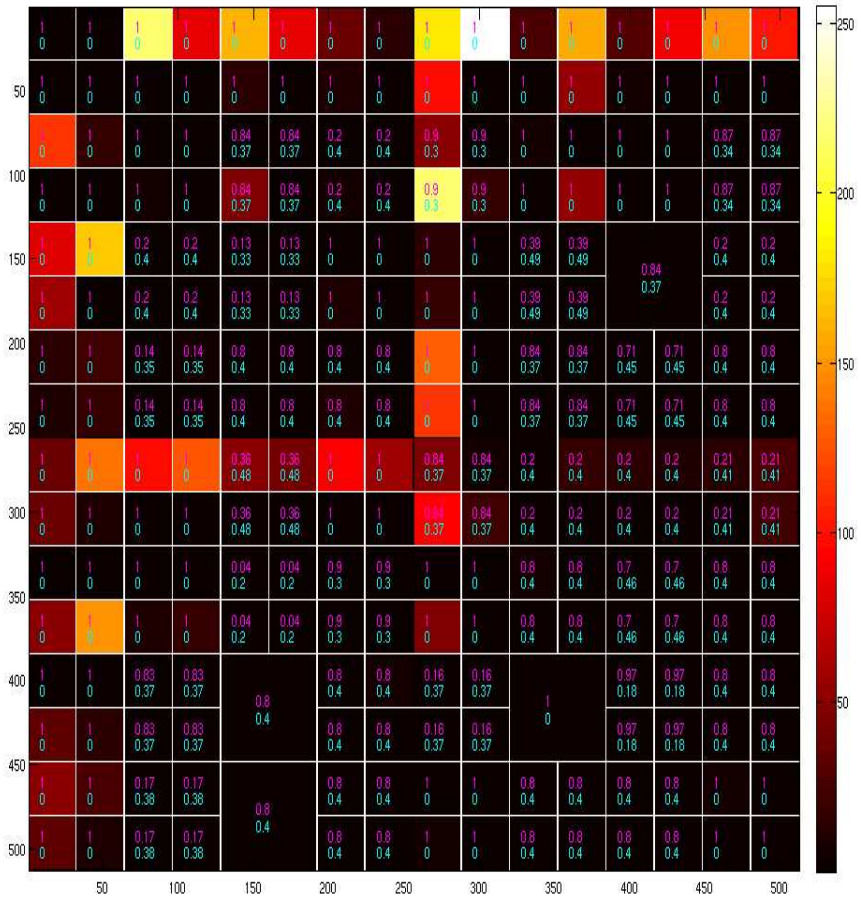


Figure B.46: Decomposition stability obtained using the Fishers Linear Discriminant over Pseudo-averaging based MAWTTs

B.4.2.3 Jensen-Shannon Distance

The Jensen Shannon distance also produces 16 different decompositions. The stability decomposition obtained using the Jensen Shannon distance is given in the Figure B.47.

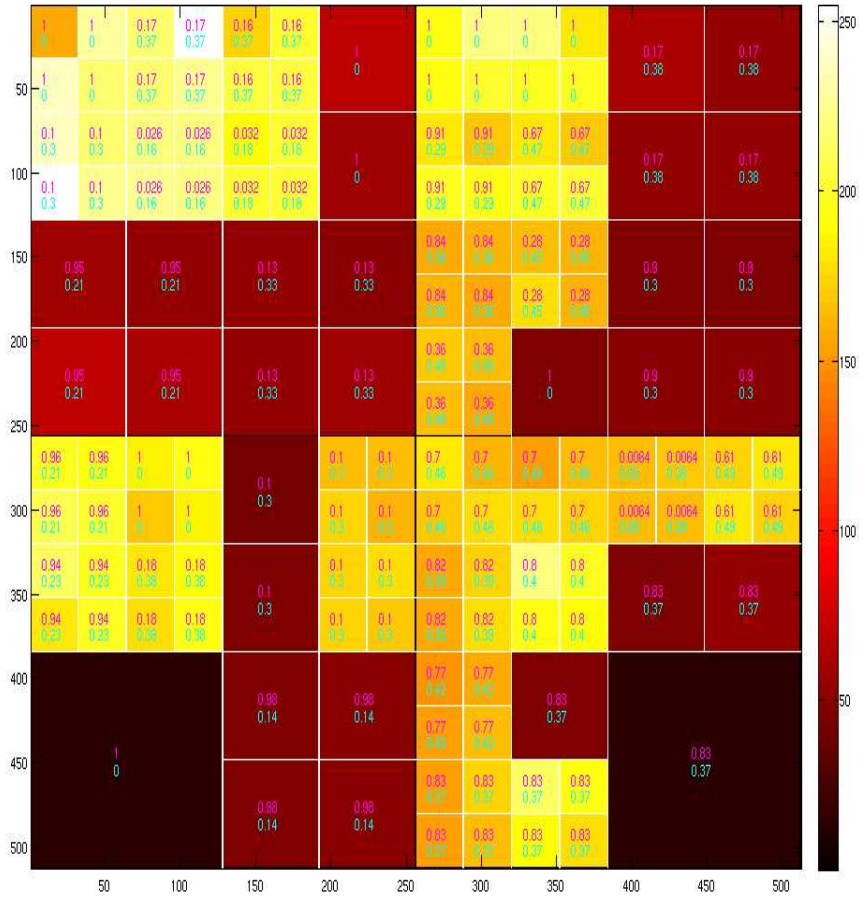


Figure B.47: Decomposition stability obtained using the Jensen Shannon distance over Pseudo-averaging based MAWTTs

B.4.2.4 Bhattacharya Distance

Bhattacharya distance produced 16 different decompositions. The stability decomposition obtained using the Bhattacharya distance is given in the Figure B.48.

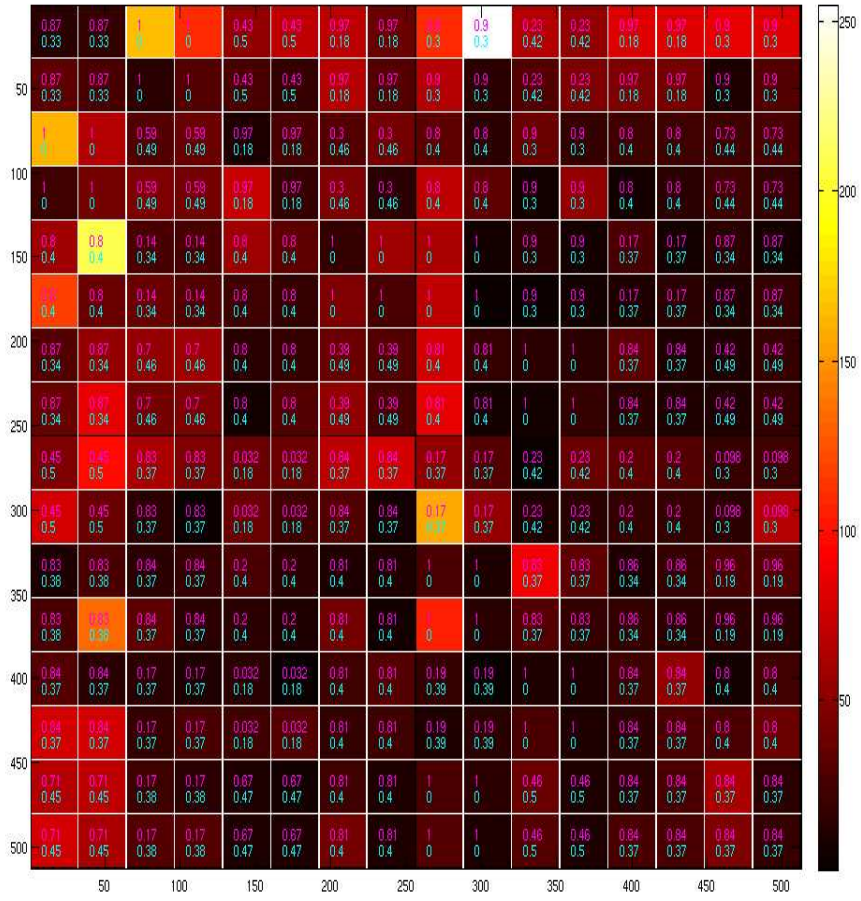


Figure B.48: Decomposition stability obtained using the Bhattacharya distance function over Pseudo-averaging based MAWTTs

B.4.2.5 Two Class Mahalanobis Distance

The Mahalanobis distance produces a total of 43 different decompositions. The stability of the decomposition obtained using the two class Mahalanobis distance is given in the Figure B.49.

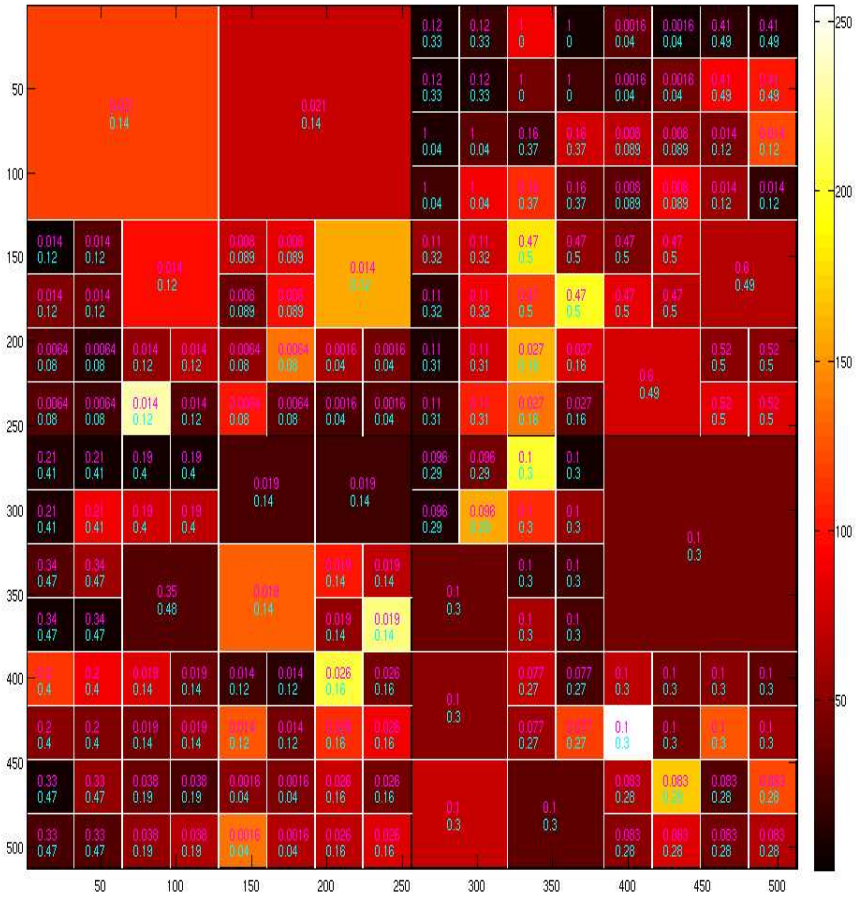


Figure B.49: Decomposition stability obtained using the Mahalanobis distance function over Pseudo-averaging based MAWTTs

The decompositions obtained are interesting in the sense that the low-level subband has not been decomposed to the fourth level in all instances. This would

have an impact on the clustering and classification accuracies produced using the distance function as we would see later.

B.4.2.6 Multi-class Mahalanobis Distance

As stated in the previous chapter, the multi-class Mahalanobis distance (MCMD) may be used to obtain a measure of differentiation between textures simultaneously and there is no need to add up the pairwise distances as in the case of all other distances. The MCMD produces a total of 81 different and unique decompositions. However, the decomposition showing the stability of the subbands obtained using the Multi-class Mahalanobis distance is given in the Figure B.50.

B.4.2.7 Energy Distance

The energy based distance function produces only one decomposition which is equivalent to the Full Wavelet Packet Transform(FWPT). The decomposition obtained using the energy-based distance function is presented in the figure B.51 and the stability information is also presented.

B.4.2.8 Relative Energy Distance

The Relative energy distance (RED) function produced 16 different decompositions. The stability decomposition obtained using the Relative Energy based Subbands distance is given in the Figure B.52.

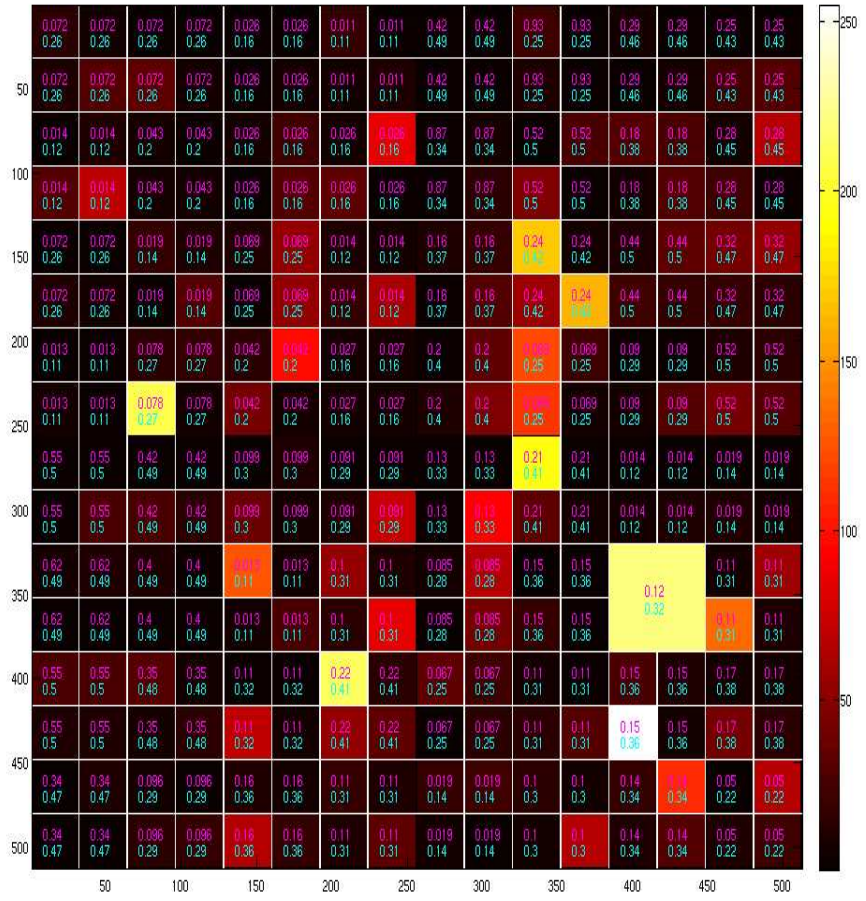


Figure B.50: Decomposition stability obtained using the Mahalanobis distance function in multi-class mode over Pseudo-averaging based MAWTTs

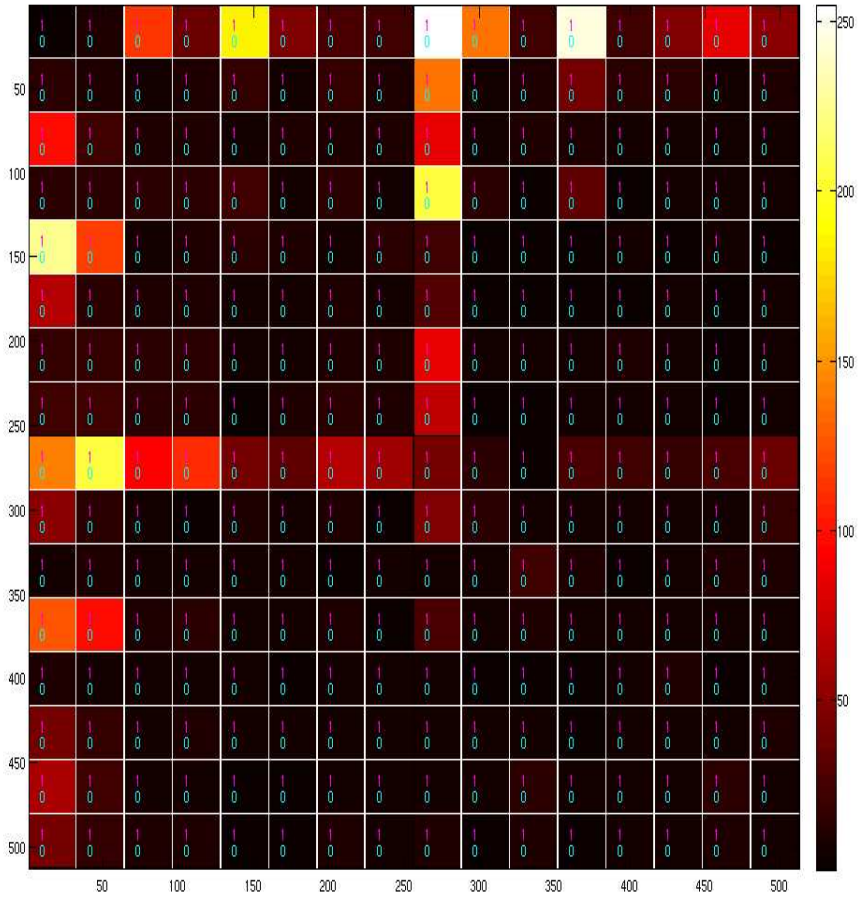


Figure B.51: Decomposition stability obtained using the Energy distance function over Pseudo-averaging based MAWTTs

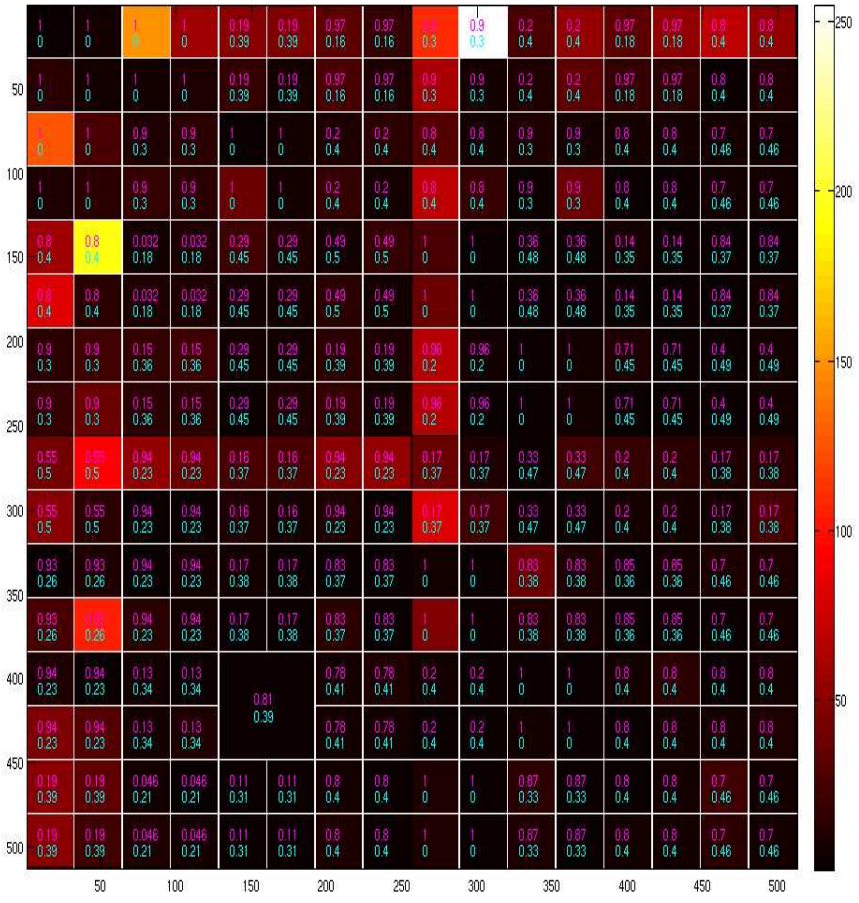


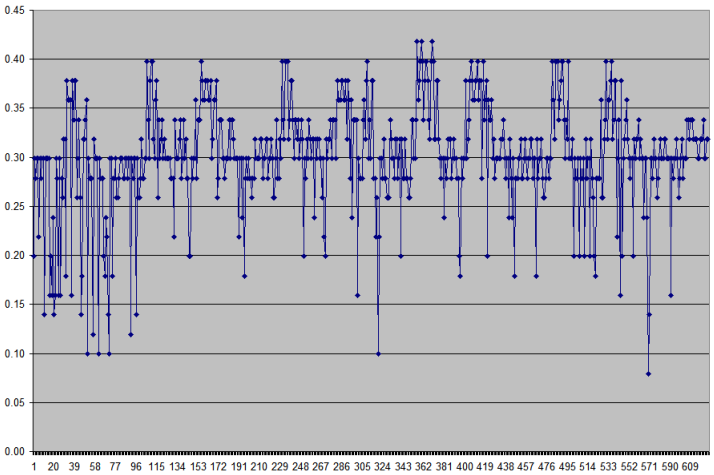
Figure B.52: Decomposition stability obtained using the Relative Energy distance function over Pseudo-averaging based MAWTTs

B.5 Decomposition Stability using Graph Matching

The graphs obtained for the graph distances obtained using the maximum common subgraph based Bunke's distance metric for different distance functions are given below.

B.5.1 Kullback-Leibler Distance

The Kullback-Leibler distance based decompositions are almost as unstable as the Hellinger distance. The various decompositions computed and the graph of the distances is given in the Figure B.53.



a.

b.

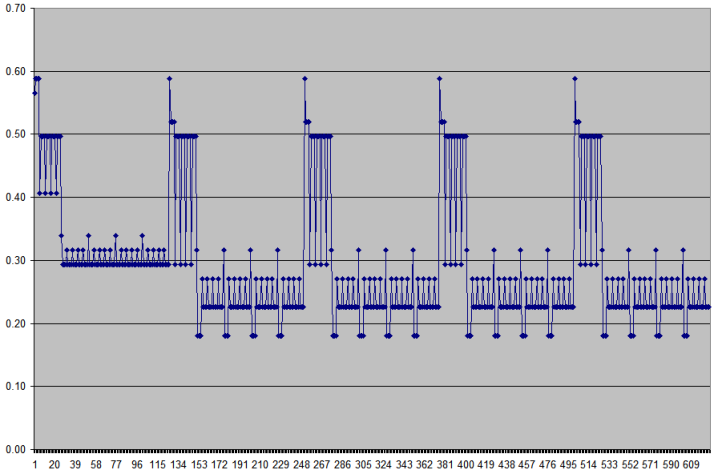


Figure B.53: Graph showing distances between various decompositions from the Union of all decompositions obtained using Kullback-Leibler distance over a. Standard Averaging, b. Pseudo-averaging based MAWTTs

B.5.2 Jensen Shannon Distance

Jensen-Shannon behaves similarly to Hellinger. This can be seen from the graphs in Figure B.54.

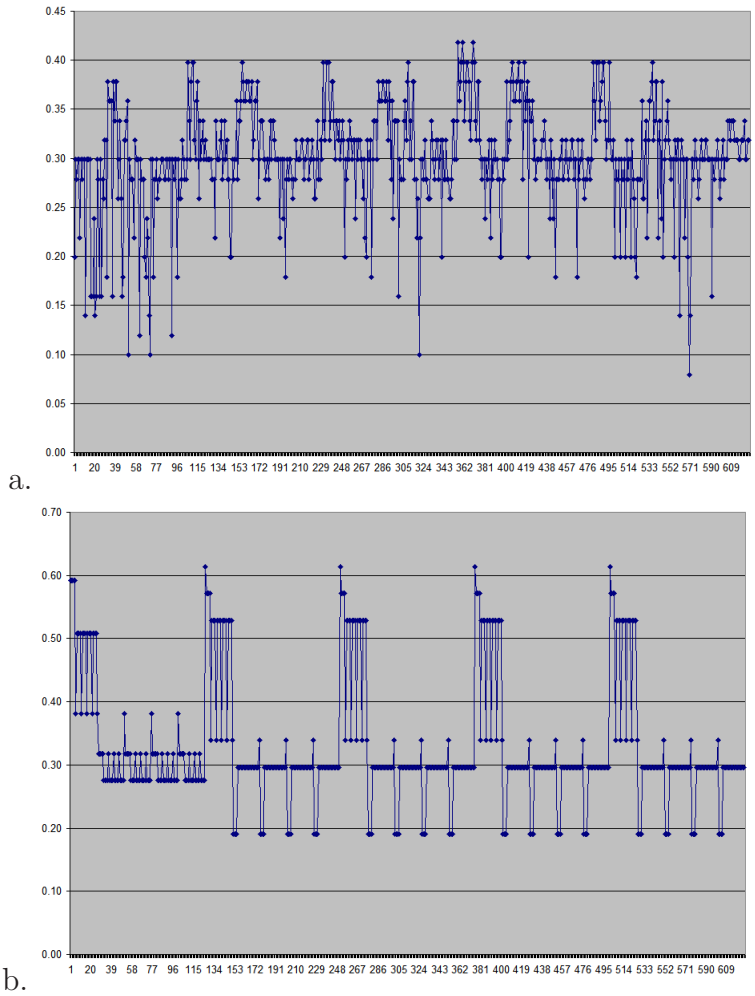


Figure B.54: Graph showing distances between various decompositions from the Union of all decompositions obtained using Jensen-Shannon distance over a. Standard Averaging, b. Pseudo-averaging based MAWTTs

B.5.3 Fishers Linear Discriminant

It can be seen from the distance function curves produced by the Fishers Linear Discriminant that some decompositions are repeated more frequently than others (Figure B.55). There is some variability in the beginning i.e. decompositions different from the later more frequently repeated ones are found but later on the same decompositions are repeated. Hence, the distance function is stable.

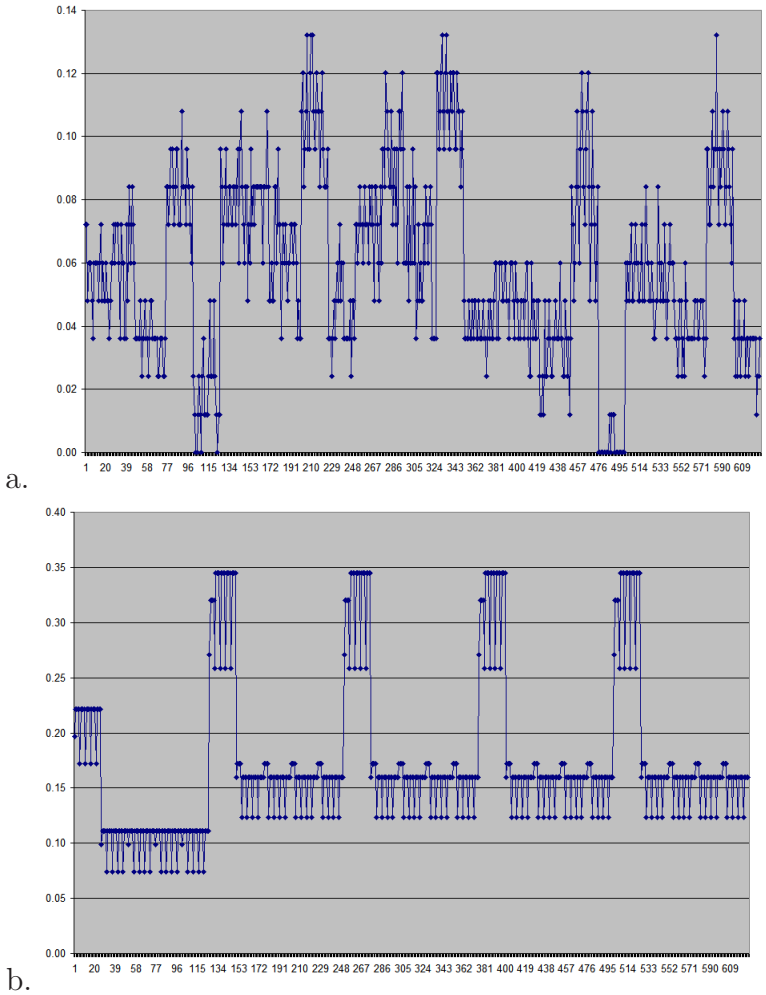


Figure B.55: Graph showing distances between various decompositions from the Union of all decompositions obtained using Fishers Linear distance over a. Standard Averaging, b. Pseudo-averaging based MAWTTs

B.5.4 Bhattacharya Distance

Behaviour of the Bhattacharya distance is similar to Fishers linear discriminant. This can seen in the Figure B.56.

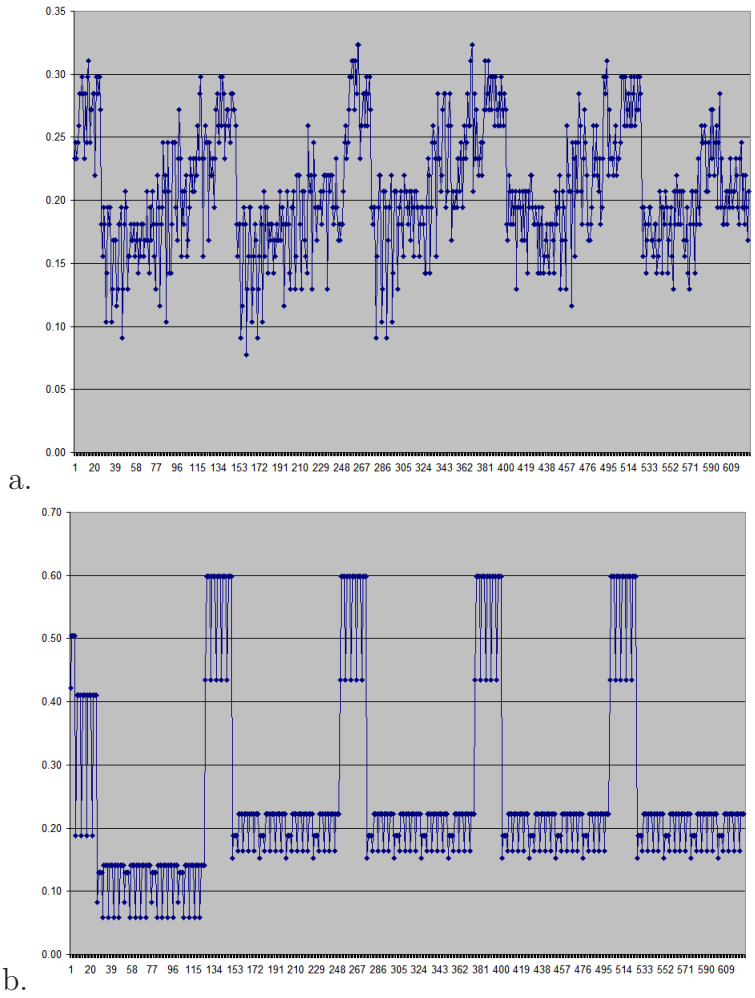


Figure B.56: Graph showing distances between various decompositions from the Union of all decompositions obtained using Bhattacharya distance over a. Standard Averaging, b. Pseudo-averaging based MAWTTs

B.5.5 Relative Energy Distance

The relative energy distance is similar to Bhattacharya and Fishers linear discriminant when it comes to decomposition stability. This is quite interesting as it does not rely on statistical measures such as mean and variance for computation but relies on subband energy. The Figure B.57 shows the graph distances obtained using Relative energy as a distance measure.

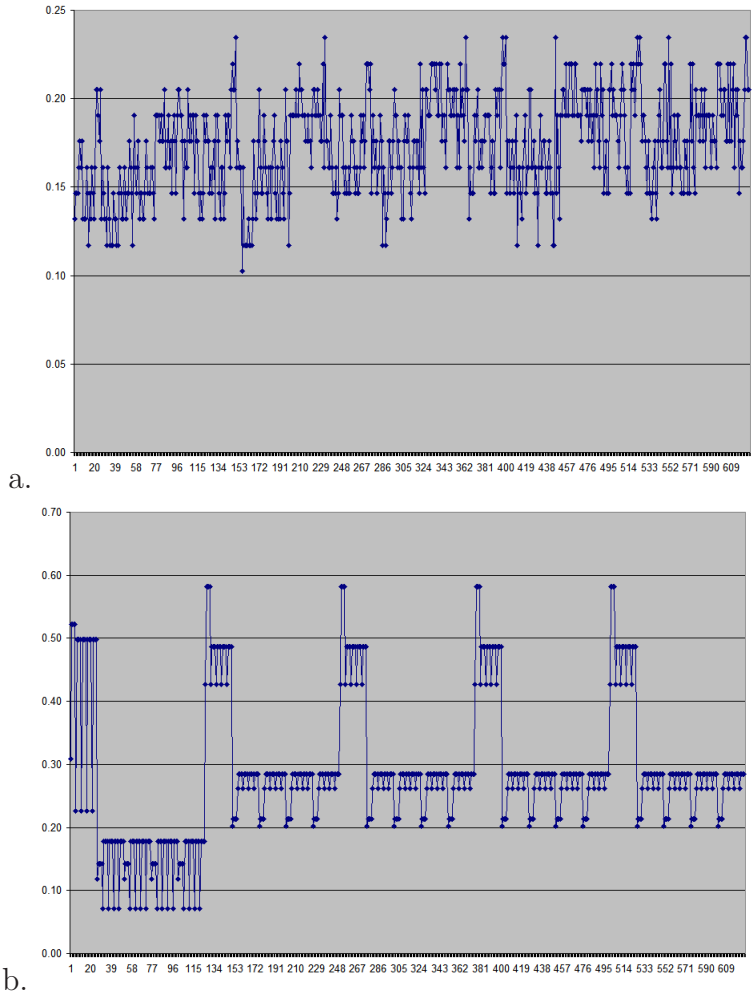


Figure B.57: Graph showing distances between various decompositions from the Union of all decompositions obtained using Relative Energy distance over a. Standard Averaging, b. Pseudo-averaging based MAWTTs

Bibliography

[1] Brigham, Hospital, W.: Neurosurgery: Facts about meningiomas (March 2009)

[2] Kleihues, P., Cavenee, W.K.: World Health Organization Classification of Tumours. Pathology and Genetics. Tumours of the Nervous System. IARC Press (2000)

[3] neurosurgerytoday.org: Meningiomas (2007)

[4] Burger, P.: What is an oligodendroglioma? Brain Pathol (2002) **12** (2002) 257–259

[5] Perry, A., Gutmann, D., Reifenberger, G.: Molecular pathogenesis of meningiomas. Journal of neuro-oncology **70**(2) (2004) 183–202

[6] Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N., Yener, B.: Histopathological image analysis: A review. Submitted to Medical Image Analysis (2009)

[7] Zbaren, P., Becker, M., Lang, H.: Staging of laryngeal cancer: Endoscopy, computed tomography and magnetic resonance versus histopathology. Europeans Archives of Oto-Rhino-Laryngology **254** (1997) S117–S122

[8] Adams, S., Baum, R.P., Stuckensen, T., Bitter, K., Hor, G.: Prospective comparison of 18f-fdg pet with conventional imaging modalities (ct, mri, us) in lymph node staging of head and neck cancer. Europeans Journal of Nuclear Medicine and Molecular Imaging **25** (2004) 1255–1260

[9] Loukas, C.G., Linney, A.: A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia and vasculature. Computer Methods and Programs in Biomedicine **74** (2004) 183–199

[10] Steel, G.G.: Basic Clinical Radiobiology. Arnold, London (1997)

[11] Green, R.: Parallel processing in a pattern recognition based image processing systems: the abbott adc-500 differential counter. In: Proc. IEEE Conf. Pattern Recognit. Image Process. (1978) 492–498

[12] Bartels, H., Wied, G.: High resolution prescreening systems for cervical cancer, the automation of uterine cancer cytology. Tutorials of Cytology (1976) 144

[13] Jelonek, Stefanowski, J.: Feature subset selection for classification of histological images. Artificial Intelligence in Medicine **9** (1997) 227–239

[14] Weaner, J., Au, J.: Application of automatic thresholding in image analysis scoring of cells in human solid tumours labeled for proliferation markers. Cytometry **29** (1997) 128–135

[15] Sahoo, P., Soltani, S., Wong, A.: Survey: a survey of thresholding techniques. Computer Vision Graphics Image Processing **41** (1988) 233–260

[16] Colley, M., Kommoss, F., Bibbo, M., Dytch, H., Frnklin, W., Holt, J., Wied, G.: Assessment of hormone receptors in breast carcinoma by immunocytochemistry and image analysis. Anal. Quant. Cytol. Histol. **11** (1989) 307–314

[17] Lamaziere, J., Lavallee, J., Zunino, C., Larrue, J.: Semi-quantitative study of the distribution of 2 cellular antigens by computer-directed color analysis. Lab. Invest. **68** (1993) 248–252

[18] Goldlust, E., Paczynski, R., He, Y., Hsu, C., Goldberg, M.: Automated measurement of infract size with scanned images of triphenyltetrazolium chloride-stained rat brains. Stroke **27** (1996) 1657–1662

[19] Uitto, J., Paul, J., Brockley, K., Pearce, R., Clark, J.: Elastic fibers in human skin: quantitation of elastic fibers by computerized digital analysis and determination of elastin by radioimmunoassay of desmosine. Lab. Invest. **49** (1983) 499–505

[20] Beier, K., Fahimi, H.: Application of automated image analysis for quantitative morphological studies of peroxisomes in rat liver in conjuction with cytochemical staining with 3-3?-diaminobenzidine and immunocytochemistry. Microscopic Research Techniques **21** (1992) 271–282

[21] Francis, I., Adeyanju, M., George, S., Junaid, T., Luthra, U.: Manual versus image analysis estimation of pcna in breast carcinoma. Anal. Quant. Cytol. Histol. **22** (2000) 11–16

[22] Ridler, T., Calvard, S.: Picture thresholding using an iterative selection method. Man Cybern. **8** (1978) 630–632

[23] Bond, S., Bradyaw, M., Gleesonb, F., Mortensen, N.: Image analysis for patient management in colorectal cancer. Computer Assisted Radiology and Surgery **1281** (2005) 278–283

[24] Ornberg, R., Woerner, B., Edwards, D.: Analysis of stained objects in histological sections by spectral imaging and differential absorption. J. Histochem. Cytochem. **47** (1999) 1307–1313

[25] Sharipo, E., Hartanto, V., Lepor, H.: Quantifying the smoothmuscle content of the prostate using double-immunoenzymatic staining and color assisted image analysis. Journal of Urology **147** (1990) 1167–1170

[26] Dobrinski, I., Ogawa, T., Avarbock, M.R., Brinster, R.L.: Computer assisted image analysis to assess colonization of recipient seminiferous tubules by spermatogonial stem cells from transgenic donor mice. Molecular reproduction and development (1999) 142–148

[27] Lehr, H., der Loos, C.V., Teeling, P., Gown, A.: Complete cromogen separation and analysis in double immunohistochemical stains using photoshop-based image analysis. Journal Histochem Cytochem. **47** (1999) 199–225

[28] Deverell, M., J.R.Salisbury, Whimster, W.: Comparisons of stains for image segmentation and measurement of nuclear parameters by computerised image analysis using ibas 2000. Pathol. Res. Pract. **185** (1989) 555–557

[29] Kohlberger, P., Obermair, A., Sliutz, G., Heinzl, H., Koelbl, H., Breitenecker, G., Gitsch, G., Kainz, G.: Quantitative immunohistochemistry of factor viii-related antigen in breast carcinoma. American Journal Clinical Pathology **105** (1996) 705–710

[30] Gonzales, R.C., Woods, R.E.: Digital Image Processing, 2nd Edition. Prentice Hall (2002)

[31] Garbay, C., Brugal, G., Choquet, C.: Application of colored image analysis to bone marrow cell recognition. Anal. Quant. Cytol. **4** (1986) 272–280

[32] van der Laak, J., Pahlplatz, M., Hanselaar, A., de Wilde, P.: Hue-saturation-density (hsd) model for stain recognition in digital images from transmitted light microscopy. Cytometry **39** (2000) 275–284

[33] Goto, M., Nagatomo, Y., Hasui, K., Amanaka, H., Murashima, S., Sato, E.: Chromaticity analysis of immunostained tumor specimens. Pathol. Res. Pract. **188** (1992) 433–437

[34] Ruifrok, A.: Quantification of immunohistochemical staining by color translation and automated thresholding. Anal. Quant. Cytol. Histol. **19** (1997) 107–113

[35] Zhou, R., Parker, D., Hammond, E.: Quantitative peroxidase-antiperoxidase complex-substrate mass deterimation in tissue sections by a dual wavelength method. Anal. Quant. Cytol. Histol. **14** (1992) 73–80

[36] Zhou, R., Hammond, E., Parker, D.: A multiple wavelength algorithm in color image analysis and its applications in stain decomposition in microscopy images. Medical Physics **23** (1996) 1977–1986

[37] Ong, S., Giam, S., Sinniah, R.: Adaptive window-based tracking for the detection of membrane structures in kindney electron micrographs. Machine Vision Applications **6** (1993) 215–223

[38] Kate, T., Belien, J., Smeulders, A., Baak, J.: Method for counting mitoses by image processing in feulgen stained breast cancer sections. Cytometry **14** (1993) 241–250

[39] Adiga, P.U., Chaudhri, B.: Region based techniques for segmentation of volumetric histo-pathological images. Comput. Methods Programs Biomed. **61** (2000) 23–47

[40] Law, A., Lam, K., Lam, F., Wong, T., Poon, J., Chan, F.: Image analysis system for assessment of immunohistochemically stained proliferative marker (mib-1) in oesophageal squamous cell carcinoma. Comput. Methods Programs Biomed **70** (2003) 37–45

[41] O'Gorman, L., Sanderson, A., Preston, K.: A system for automated liver tissue image analysis: methods and results. IEEE Trans. Biomed. Eng. **32** (1985) 696–706

[42] Jain, J., Smith, S., Backer, E.: Segmentation of muscle cell pictures: a preliminary study. IEEE Trans. Pattern Anal. Mach. Intell. **2** (1980) 232–242

[43] Adiga, P., Chaudhuri, B., Rodenacker, K.: Semi-automatic segmentation of tissue cells from confocal microscope images. In: Proceedings of 13th International Conference on Pattern Recognition. (1996)

[44] Mouroutis, T., Roberts, A., Bharath, A.: Robust cell nuclei segmentation using statistical modelling. Bioimaging **6** (1998) 79–91

[45] Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. In: Proceedings of 1st International Conference on Computer Vision,. (1987) 259–268

[46] Fok, Y.L., Chan, J., Chin, R.: Automated analysis of nerve-cell images using active contour models. IEEE Trans. Med. Imaging **15** (1996) 353–368

[47] Zhou, P., Pycock, D.: Robust statistical models for cell image interpretation. Image Vision Computing **15** (1997) 307–316

[48] Yamada, H., Merritt, C., Kasvand, T.: Recognition of kidney glomerulus by dynamic programming matching method. IEEE Trans. Pattern Anal. Mach. Intell. **10** (1988) 731–737

[49] Wu, H., Gil, J., Barba, J.: Optimal segmentation of cell images. Image Signal Process. **145** (1998) 50–56

[50] Parker, J.: Algorithms for Image Processing and Computer Vision,. Wiley, New York (1997)

[51] Murata, S.I., Herman, P., Lakowicz, J.: Texture analysis of fluorescence lifetime images of nuclear dna with effect of fluorescence resonance energy transfer. Cytometry **43** (2001) 94–100

[52] Qureshi, H., Rajpoot, N., Nattkemper, T., Hans, V.: Comparison: Meningioma classification using wavelet packets and normal texture based classification. In: Proceedings of British Machine Vision Association Student Chapters Meeting. (March 2007)

[53] Mutter, G.L., Baak, J.P.A., Crum, C.P., Ralph M. Richart, A.F., Faquin, W.C.: Endometrial precancer diagnosis by histopathology, clonal analysis, and computerized morphometry. The Journal of Pathology **190** (2000) 462–469

[54] Orbo, A., Baak, J.P.A., Kleivan, I., Lysne, S., Prytz, P.S., Broeckaert, M.A.M., Slappendel, A., Tichelaar, H.J.: Computerised morphometrical analysis in endometrial hyperplasia for the prediction of cancer development. a long term retrospective study from northern norway. The Journal of Clinical Pathology **53** (2000) 697–703

[55] Inman, C., Rees, L., Barker, E., Haverson, K., Stokes, C., Bailey, M.: Validation of computer-assisted, pixel-based analysis of multiple-colour immunofluorescence histology. Journal of Immunological Methods **302**(1-2) (2005) 156 – 167

[56] Angelini, L., Antonaci, A.R., Angelis, R.D., Maceratini, R., Daniele, A., Neroni, M.: Statistical analysis of clinical breast cancer by computer. Informatics for Health and Social Care 3.2 **3** (1978) 131 – 135

[57] Hamilton, P., Bartels, P., Thompson, D., Anderson, N., Montironi, R., Sloan, J.: Automated location of dysplastic fields in colorectal histology using image texture analysis. J. Pathol. **182** (1997) 68–75

[58] Esgiar, A., Naguib, R., Sharif, B., Bennett, M., Murray, A.: Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa. IEEE Trans. Inf. Technol. Biomed. **2** (1998) 197–203

[59] Gilles, F., Gentile, A., Doussal, V.L., Bertrand, F., Kahn, E.: Grading of cystosarcoma phyllodes by texture analysis of tissue architecture. Anal. Quant. Cytol. Histol. **16** (1994) 95–100

[60] Gilles, F., Gentile, A., Doussal, V.L., Kahn, E.: Use of texture parameters in the classification of soft tissue tumours. Anal. Quant. Cytol. Histol. **16** (1994) 315–320

[61] Cross, S.: Fractals in pathology. J. Pathol. **182** (1997) 1–8

[62] Cross, S., Bury, J., Silcocks, P., Stephenson, T., Cotton, D.: Fractal geometric analysis of colorectal polyps. J. Pathol. **172** (1994) 248–262

[63] Cross, S., Howat, A., Stephenson, T., Cotton, D., Underwood, J.: Fractal geometric analysis of material from molar and non-molar pregnacies. J. Pathol. **173** (1994) 115–118

[64] Irinopoulou, T., Rigaut, J., Benson, M.: Toward objective prognostic grading of prostatic carcinoma using image analysis. Anal. Quant. Cytol. Histol. **15** (1993) 341–344

[65] Gibson, D., Gaydecki, P.: Definition and application of a fourier domain texture measure: applications to histological image analysis. Comput. Biol. Med. **25** (1995) 551–557

[66] Haralick, R., Shanmugan, K., Dinstein, J.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **3** (1973) 610–621

[67] Lessmann, B., Hans, V., Degenhard, A., Nattkemper, T.W.: Feature space exploration of pathology images using content-based database visualization. In: Proceedings SPIE Medical Imaging. (2006)

[68] Qureshi, H., Rajpoot, N., Wilson, R., Nattkemper, T., Hans, V.: Comparative analysis of discriminant wavelet packet features and raw image features for classification of meningioma subtypes. In: Proceedings of Medical Image Understanding and Analysis. (2007)

[69] Qureshi, H., Wilson, R., Rajpoot, N.: Optimal wavelet basis for wavelet packets based meningioma subtype classification. In: Proceedings 12th Medical Image Understanding and Analysis (MIUA'2008). (2008)

[70] Katouzian, A., Baseri, B., Konofagou, E.E., Laine, A.F.: Texture-driven coronary artery plaque characterization using wavelet packet signatures. In: ISBI, IEEE (2008) 197–200

[71] Jafari-Khouzani, K., Soltanian-Zadeh, H.: Multiwavelet grading of pathological images of prostate. IEEE Transactions on Biomedical Engineering **50**(6) (2003) 697–704

[72] Ong, S., Jin, X., Jayasooriah, Sinniah, R.: Image analysis of tissue sections. Comput. Med. Biol. **26** (1996) 269–279

[73] Duda, R., Hart, P.: Pattern Classification and Scene Analysis. Wiley, New York (1973)

[74] Firestone, L.: Automated microscopy for lymph node cancer diagnosis. In: Proceedings of SPIE 1894. (1993) 15–20

[75] Hibbard, L., McKeel, D.: Multiscale detection and analysis of the senile plaques of alzheimer's disease. IEEE Trans. Biomed. Eng. **42** (1995) 1218–1225

[76] Hittelet, A., Yeaton, P., Decaestecker, C., Remmelink, M., Nagy, N., Cremer, M., Salmon, I., Kiss, R., Bourgeois, N.: Discrimination between dysplastic and malignant epithelium of the ampulla of vater based on quantitative image cytometric data. Anal. Quant. Cytol. Histol. **22** (2000) 98–106

[77] Amaral, T., McKenna, S.J., Robertson, K., Thompson, A.: Classification of breast-tissue microarray spots using colour and local invariants. In: ISBI. (2008) 999–1002

[78] Hufnagl, P., Guski, H., Wolf, G., Wenzelides, K., Martin, H., Roth, K.: The particle expert system for tumour grading by automated image analysis. Anal. Quant. Cytol. Histol. **11** (1989) 440–446

[79] Dawson, A., Austin, R., Weinberg, D.: Nuclear grading of breast carcinoma by image analysis. Am. J. Clin. Pathol. **95** (1991) S29–S37

[80] Bartels, P., Gahm, T., Thompson, D.: Automated microscopy in diagnostic histopathology from image processing to automated reasoning. Int. J. Imaging Syst. Technol. **8** (1997) 214–223

[81] Thompson, D., Bartels, P., Bartels, H., Hamilton, P., Sloan, J.: Knowledge-guided segmentation of colorectal histopathologic imagery. Anal. Quant. Cytol. Histol. **4** (1993) 236–246

[82] Thompson, D., Bartels, P., Bartel, H., Montironi, R.: Image segmentation of cribriform gland tissue. Anal. Quant. Cytol. Histol. **17** (1995) 314–322

[83] Bartels, P., Thompson, D., Weber, J.: Expert system in histopathology. iv. the management of uncertainty. Anal. Quant. Cytol. Histol. **14** (1992) 1–13

[84] Looney, C.: Pattern Recognition Using Neural Networks. Oxford University Press, Oxford, UK (1997)

[85] Dytch, H., Wied, G.: Artificial neural networks and their use in quantitative pathology. Anal. Quant. Cytol. Histol. **12** (1990) 379–393

[86] Klencki, M., Slowinska-Klencka, D., Lewinski, A.: Multifarious system for quantitative analysis of histologic compartments. Comput. Biomed. Res. **30** (1997) 165–169

[87] Refenes, A., Jain, N., Alsulaiman, M.: An integrated neural network system for histological image understanding. In: Proceedings of SPIE 1386. (1990) 62–74

[88] Karakitsos, P.: A comperative study of three variations of the learning vector quantizer in the discrimination of benign from malignant gastric cells. Cytopathology **9** (1998) 114–125

[89] Sjostrom, P., Frydel, B., Wahlberg, L.: Artificial neural network-aided image analysis system for cell counting. Cytometry **36** (1999) 18–26

[90] Amaral, T., McKenna, S.J., Robertson, K., Thompson, A.: Scoring of breast tissue microarray spots through ordinal regression. In: VISSAPP (2). (2009) 243–248

[91] McKenna, S., Ricketts, I., Cairns, A., Hussein, K.: A comparison of neural network architectures for cervical cell classification. In: Artificial Neural Networks, 1993., Third International Conference on. (May 1993) 105–109

[92] Lessmann, B., Nattkemper, T.W., Hans, V., Degenhard, A.: A method for linking computed image features to histological semantics in neuropathology. Journal of Biomedical Informatics - Special Issue: Intelligent Data Analysis **40**(6) (Dec 2007) 631–41

[93] Wirjadi, O., Breuel, T., Feiden, W., Kim, Y.J.: Automated feature selection for the classification of meningioma cell nuclei. In Handels, H., Ehrhardt, J., Horsch, A., Meinzer, H.P., Tolxdorff, T., eds.: Bildverarbeitung fr die Medizin. Informatik Aktuell, Springer (2006) 76–80

[94] Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M., Tomaszewski, J.: Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In: 5th IEEE International Symposium on

Biomedical Imaging: From Nano to Macro (ISBI 2008). (May 2008) 284–287

[95] Petushi, S., Garcia, F.U., Haber, M.M., Katsinis, C., Tozeren, A.: Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. BMC Medical Imaging **6** (2006) 14

[96] Ji, Q., Engel, J., Craine, E.: Texture analysis for classification of cervix lesions. IEEE Transactions on Medical Imaging **19**(11) (Nov. 2000) 1144–1149

[97] Schupp, S., Elmoataz, A., Fadili, J., Herlin, P., Bloyet, D.: Image segmentation via multiple active contour models and fuzzy clustering with biomedical applications. In: Proceedings. 15th International Conference on Pattern Recognition. Volume 1. (2000) 622–625 vol.1

[98] Karvelis, P.S., Fotiadis, D.I., Georgiou, I., Syrrou, M.: A watershed based segmentation method for multispectral chromosome images classification. In: Proc IEEE Eng Med Biol Soc. Volume 1. (2006) 3009–12

[99] Elmoataz, A., Bloyet, D.: Image segmentation via multiple active contour models and fuzzy clustering with biomedical applications. In: ICPR '00: Proceedings of the International Conference on Pattern Recognition, Washington, DC, USA, IEEE Computer Society (2000) 1622

[100] Wilson, R.: Bmvc93 tutorial notes on wavelet transforms. In: British Machine Vision Conference. (1993)

[101] Wilson, R., Spann, M.: Image Segmentation and Uncertainty. Wiley (1988)

[102] Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(7) (1989) 674–693

[103] Mallat, S.: A compact multiresolution representation: The wavelet model. In: IEEE Computer Society Workshop on Computer Vision. (1987)

[104] Peters, T.M., Williams, J.C., Bates, J.H.T., Pike, G.B., Munger, P., eds.: The Fourier Transform in Biomedical Engineering. Springer (1998)

[105] Thibos, L.N., Wheeler, W., Horner, D.: Power vectors: An application of fourier analysis to the description and statistical analysis of refractive error. In: Visual Science and Its Applications. (March 1997)

[106] Garcia-Caurel, E., Drvillon, B., Martino, A.D., Schwartz, L.: Application of fourier transform infrared ellipsometry to assess the concentration of biological molecules. Applied Optics **41** (2002) 7339–7345

[107] Julesz, B., Caelli, T.: On the limits of fourier decompositions in visual texture perception,. Perception **8** (1979) 1979

[108] Helms, H.D.: Nonrecursive digital filters: design method for achieving specifications of frequency response. IEEE Transactions **AU-16** (1968) 336

[109] Campbell, F., Robson, J.: Application of fourier analysis to the visibility of gratings. Journal of Physiology (1968) 197

[110] Welch, P.: The use of the fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions **AU-15** (1967) 70

[111] Bingham C., Thinly, J.W., Godfrey, M.D.: Modern techniques of power spectrum estimation. IEEE Transactions **AU-15** (1967) 56

[112] Nawab, S.H., Quatieri, T.F.: Short-time Fourier transform. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1987)

[113] Graps, A.: An introduction to wavelets. In: IEEE Computational Science and Engineering. IEEE Computer Society (1995)

[114] Rioul, O., Vitterli, M.: Wavelets and signal processing. IEEE Signal Processing Magazine **8 (Issue 4)** (Oct. 1991) 14–38

[115] Gabor, D. Journal Institute of Electrical Engineers **93** (1946) 429

[116] Hilgevoord, J., Uffink, J.: The uncertainty principle. In Zalta, E.N., ed.: The Stanford Encyclopedia of Philosophy. Stanford (Fall 2008)

[117] Hubbard, B.B.: The World According to Wavelets: The Story of a Mathematical Technique in Making. A. K. Peters Ltd. (1998)

[118] Rajpoot, N.M.: Adaptive Wavelet Image Compression. Thesis submitted to University of Warwick (2001)

[119] Coifman, R.R., Wickerhauser, M.V.: Entropy-based algorithms for best basis selection. IEEE Transactions on Information Theory **38**(2) (1992) 713–718

[120] Zhang, J., Tan, T.: Brief review of invariant texture analysis methods. Pattern Recognition **35**(3) (2002) 735 – 747

[121] Chang, T., Kuo, C.J.: Texture analysis and classification with tree-structured wavelet transform. In: IEEE Transactions Image Processing. (1993)

[122] man Pun, C., Lee, M.C.: Log-polar wavelet energy signatures for rotation and scale invariant texture classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(5) (2003) 590–603

[123] Chen, J.L., Kundu, A.: Rotation and gray scale transform invariant texture identification using wavelet decomposition and hidden markov model. IEEE Transactions on Pattern Analysis and Machine Intelligence **16** (1994) 208–214

[124] Laine, A., Fan, J.: Texture classification by wavelet packet signatures. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. (1993)

[125] Sengur, A., Turkoglu, I., Ince, M.C.: Wavelet packet neural networks for texture classification. Expert Systems with Applications **32**(2) (2007) 527 – 533

[126] Unser, M.: Texture classification and segmentation using wavelet frames. IEEE Transactions on Image Processing **4** (1995) 1549–15604

[127] Kim, S.C., Kang, T.J.: Texture classification and segmentation using incomplete tree structured wavelet packet frame and gaussian mixture model. In: IEEE International Workshop on Imaging Systems and Techniques. (2005)

[128] Rajpoot, N.: Local discriminant wavelet packet basis for texture classification. In: Proceedings SPIE Wavelets X, San Diego, California. (2003)

[129] Bhalerao, A., Rajpoot, N.: Discriminant feature selection for texture classification. In: Proceedings British Machine Vision Conference (BMVC'2003). (2003)

[130] Huang, K., Aviyente, S.: Information-theoretic wavelet packet subband selection for texture classification. Signal Processing **86**(7) (2006) 1410 – 1420

[131] Huang, K., Aviyente, S.: Wavelet feature selection for image classification. Image Processing, IEEE Transactions on **17**(9) (Sept. 2008) 1709–1720

[132] Randen, T., Husoy, J.H.: Filtering for texture classification: a comparative study. IEEE Transactions on Pattern Analysis and Machine Intelligence **21**(4) (1999) 291–310

[133] Arivazhagan, S., Ganesan, L., Priyal, S.P.: Texture classification using gabor wavelets based rotation invariant features. Pattern Recognition Letters **27**(16) (2006) 1976 – 1982

[134] Livens, S., Scheunders, P., Van de Wouwer, G., van Dyck, D., Smets, H., Winkelmans, J., Bogaerts, W.: A texture analysis approach to corrosion image classification. Microscopy, Microanalysis, Microstructures **7**(2) (1996) 143–52

[135] Haralick, R.: Statistical and structural approaches to texture. In: Proceedings of the IEEE. (1979)

[136] Smith, L.: A tutorial on principal component analysis (2002)

[137] Qureshi, H., Sertel, O., Rajpoot, N., Wilson, R., Gurcan, M.: Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. In: Proceedings 11th Medical Image Computing and Computer-Assisted Intervention (MICCAI'2008). (2008)

[138] Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24**(7) (Jul 2002) 971–987

[139] Ma, W., Manjunath, B.: Texture features and learning similarity. In: Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on. (Jun 1996) 425–430

[140] Qureshi, H., Rajpoot, N., Masood, K., Hans, V.: Classification of meningiomas using discriminant wavelet packets and learning vector quantization. In: Proceedings of Medical Image Understanding and Analysis. (2006)

[141] Julesz, B., Bergen, J.R.: Textons, the fundamental elements in preattentive vision and perception of textures. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1987)

[142] Lu S.W., R.Y., C.Y., S.: Hierarchical attributed graph representation and recognition of handwritten chinese characters. Pattern Recognition **24** (1991) 617–632

[143] Christmas W.J., K.J., M., P.: Structural matching in computer vision using probabilistic relaxation. IEEE Transaction on Pattern Analysis and Machine Intelligence **17** (1995) 749–764

[144] Wong, E.: Model matching in robot vision by subgraph isomorphism. Pattern Recognition **25** (1992) 287–304

[145] Ullman, J.: An algorithm for subgraph isomorphism. J. ACM **23** (1976) 31–42

[146] Shapiro, L., Haralick, R.: Structural descriptions and inexact matching. IEEE Transactions on Pattern Analysis and Machine Intelligence **3** (1981) 504–519

[147] Horaud, R., Skordas, T.: Stereo correspondence through feature grouping and maximal cliques. IEEE Transactions on Pattern Analysis and Machine Intelligence **11** (1989) 1168–1180

[148] Levinson, R.: Pattern associativity and the retrieval of semantic networks. Comput. Math. Appl. **23** (1992) 573–600

[149] Bunke, H.: On a relation between graph edit distance and maximum common subgraph. Pattern Recognition Lett. **18 8** (1997) 689694

[150] Bunke, H., Shearer, K.: A graph distance metric based on the maximal common subgraph. Pattern Recognition Letters (1998)

[151] Materka, A., Strzelecki, M.: Texture analysis methods - a review. In: COST B11 report, Institute of Electronics, Technical University of Lodz. (1998)

[152] Hartigan, J., Wong, M.: A k-means clustering algorithm. JR Stat. Soc., Ser. C **28** (1979) 100–108

[153] Steinbach, M., Karypis, G., Kumar, V.: A comparison of document clustering techniques. In: KDD workshop on text mining. Volume 400., Citeseer (2000) 525–526

[154] Wilpon, J., Rabiner, L.: A modified K-means clustering algorithm for use in speaker-independent isolated word recognition. The Journal of the Acoustical Society of America **75** (1984) S93

[155] Huang, Z.: Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery **2**(3) (1998) 283–304

[156] Steinley, D.: Properties of the Hubert-Arabie adjusted Rand index. Psychological methods **9**(3) (2004) 386–396

[157] Hubert, L., Arabie, P.: Comparing partitions. Journal of classification **2**(1) (1985) 193–218

[158] Mazhelis, O.: One-class classifiers: a review and analysis of suitability in the context of mobile-masquerader detection. South African Computer Journal (2006) 29–48

[159] Bioch, J., Verbeke, W., van Dijk, M.: Neural networks: New tools for data analysis? In: Neural Network Applications and Tools. Workshop on (1993). (Sep 1993) 29–38

[160] Vapnik, V.: Statistical Learning Theory. Springer N.Y. (1998)

[161] Nattkemper, T.W., Arnrich, B., Lichte, O., Timm, W., Degenhard, A., Pointon, L., Hayes, C., Leach, M.O.: Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods. Artificial Intelligence in Medicine **34**(2) (2005) 129–139

[162] Rajpoot, K., Rajpoot, N.: Svm optimization for hyperspectral colon tissue cell classification. In: Proceedings 7th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'2004). Volume 3217., Springer (2004) 829–837

[163] Drucker, H., Wu, D., Vapnik, V.: Support vector machines for spam categorization. IEEE Transactions on Neural networks **10**(5) (1999) 1048–1054

[164] Guo, G., Li, S., Chan, K.: Face recognition by support vector machines. In: Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000. Proceedings. (2000) 196–201

[165] Tong, S., Koller, D.: Support vector machine active learning with applications to text classification. The Journal of Machine Learning Research **2** (2002) 45–66

[166] Muller, K., Smola, A., Ratsch, G., Scholkopf, B., Kohlmorgen, J., Vapnik, V.: Predicting time series with support vector machines. Lecture notes in computer science (1997) 999–1004

[167] Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001) Software available at urlhttp://www.csie.ntu.edu.tw/ cjlin/libsvm.

[168] Jain, A., Zongker, D.: Feature selection: Evaluation, application, and small sample performance. IEEE Trans. Pattern Anal. Mach. Intell. **19**(2) (1997) 153–158

[169] Yu, L., Liu, H.: Efficient feature selection via analysis of relevance and redundancy. J. Mach. Learn. Res. **5** (2004) 1205–1224

[170] Julesz, B.: Experiments in the visual perception of texture,. Scientific American (1975) 34–43

[171] Pollard, D.E.: A user's guide to measure theoretic probability. Cambridge University Press. (2002)

[172] Lin, J.: Divergence measures based on the shannon entropy. Information Theory, IEEE Transactions on **37**(1) (1991) 145–151

[173] Mahalanobis, P.C.: On the generalised distance in statistics. In: Proceedings National Institute of Science. India. (1936)