TECHNICAL UNIVERSITY OF CRETE
ELECTRICAL AND COMPUTER ENGINEERING DEPARTMENT

APPLICATION OF ARTIFICIAL INTELLIGENT METHODOLOGY IN MEDICAL DIAGNOSIS

by

Georgios Spiliotis

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DIPLOMA OF
ELECTRICAL AND COMPUTER ENGINEERING

2018

THESIS COMMITTEE

Professor Michael Zervakis
Professor Konstantinos Kalaitzakis
Dr. Eleftheria Sergaki

ABSTRACT

Magnetic Resonance Imaging (MRI) is a widely used medical imaging modality that provides accurate information about the human tissue, anatomy and pathology, of a non-invasive form. If used to scan a human brain it can provide images of high contrast, therefore distinguishing between the three major brain tissues: Cerebro-Spinal Fluid (CSF), Grey Matter (GM) and White Matter (WM). In this way MRI can greatly assist radiologists and doctors in providing a more precise diagnosis and therapy. Because of their unpredictable appearance and shape, segmenting brain tumors from multi-modal imaging data is one of the most challenging tasks in medical image analysis. Manual detection and classification of brain tumor by an expert is still considered the most acceptable method, but it is too time-consuming, especially because of the large amount of data that have to be analysed manually.

In this thesis we examine, optimize and finally combine specific state-of-the-art methods comprising of four consistent methods for Computer-Aided Diagnosis (CAD) processes for detection of a brain tumor from MRI, of T2 weighted modality, from the axial plane (T2 MRI). We denote the four proposed methodologies as “Method 1” to “Method 4”. These methodologies are based on image pre-processing and classification by utilizing neural networks (NN) or a hybrid combination of neural networks and fuzzy logic (ANFIS).

In order to gauge the current innovation status in automated brain tumor segmentation and to compare various proposed methods in bibliography, we use a large dataset of brain tumor MR scans, in which the relevant tumor structures have been delineated. These are provided freely from Multimodal Brain Tumor Image Segmentation (BRATS) MICCAI 2015.

Our dataset of the training and testing data set, referred to male and female adult persons, includes 24 non-tumorous cases and 202 tumorous cases that all have been segmented visually by our experienced neurosurgeon partner, Dr. A. Krasoudakis. The healthy MRI scans are from “St. George” general hospital of Chania, Crete and from Harvard General

Hospital database. Our data set contains about 5% high-grade, 82% low-grade glioma cases, 3% unhealthy but not recognizable cases and 10% healthy cases.

At the pre-processing stage we apply a skull-stripping algorithm to isolate the brain region. Subsequently, we use a high-pass Gaussian filter for sharpening and a median filter for noise reduction. In post-processing stage, for image segmentation we use Otsu's threshold as well as we implement morphological operators for region of interest (ROI) definition.

In our proposed CAD Method 1, feature extraction is made using Grey-level co-occurrence matrix (GLCM) and 13 statistical features are calculated. In Method 2, feature extraction is made by using Discrete Wavelet Transform (DWT) and dimensionality reduction is implemented using Principal Components Analysis (PCA). In Method 3, all the above methods are combined and GLCM matrix is applied after the DWT and PCA stages, so to provide the necessary statistical features. In Method 4, the Mean-Shift algorithm is implemented at the post-processing stage for better segmentation results and features extraction is made according to Method 3. The features extracted from every proposed CAD method are processed at first with a feed-forward artificial neural network (ANN) with back-propagation training algorithm and then with an adaptive neuro-fuzzy inference system (ANFIS) for the methods with GLCM matrix.

The experimental results of the proposed methods have been validated and evaluated for performance over a testing set of images based on sensitivity, specificity and accuracy with the best results reaching 98.8% sensitivity, 62.5% specificity and 95.6% accuracy.

ΠΕΡΙΛΗΨΗ

Η Μαγνητική Τομογραφία (MRI) αποτελεί έναν ευρέως διαδεδομένο, μη-επεμβατικό, τρόπο ιατρικής απεικόνισης που παρέχει ακριβή πληροφορία για τους ανθρώπινους ιστούς, την ανατομία και την παθογένεια. Όταν χρησιμοποιείται για την απεικόνιση του ανθρώπινου εγκεφάλου παρέχει εικόνες υψηλής αντίθεσης, κάτι που βοηθάει στο να ξεχωρίσουμε τους τρεις βασικούς εγκεφαλικούς ιστούς: το εγκεφαλονωτιαίο υγρό (CSF), τη Φαιά ουσία (GM) και την Λευκή ουσία (WM). Αυτή η πληροφορία βοηθάει σημαντικά τους ιατρούς προκειμένου να παρέχουν ακριβέστερη διάγνωση και θεραπεία. Λόγω του απρόβλεπτου σχήματός τους, η κατάτμηση των εγκεφαλικών όγκων με βάση δεδομένα από διαφορετικού τύπου απεικονίσεις, συνιστά ένα από τα πιο απαιτητικά προβλήματα ανάλυσης ιατρικών εικόνων. Η χειροκίνητη εύρεση και κατηγοριοποίηση εγκεφαλικών όγκων από κάποιον ειδικό εξακολουθεί να θεωρείται ως η πλέον αποδεκτή μέθοδος, παρόλο που είναι αρκετά χρονοβόρα, κυρίως λόγω του μεγάλου αριθμού των δεδομένων που περιλαμβάνει και τα οποία πρέπει να αναλυθούν από το χρήστη.Σε αυτήν την διπλωματική εργασία εξετάζουμε, βελτιώνουμε και συνδυάζουμε συγκεκριμένες βιβλιογραφικές μεθόδους, παρουσιάζοντας τέσσερις νέες μεθοδολογίες Computer-Aided Diagnosis (CAD) για την εύρεση εγκεφαλικών όγκων από MRI, τύπου T2, του αξονικού επιπέδου (T2 MRI). Οι μεθοδολογίες αυτές βασίζονται στην προ-επεξεργασία εικόνων και στην κατηγοριοποίησή τους, χρησιμοποιώντας νευρωνικά δίκτυα (NN) ή έναν υβριδικό συνδυασμό νευρωνικών δικτύων και ασαφούς λογικής (ANFIS). Με σκοπό να συγκρίνουμε τις υπάρχουσες προτεινόμενες μεθόδους της βιβλιογραφίας ως προς την αυτοματοποιημένη κατηγοριοποίηση εγκεφαλικών όγκων, χρησιμοποιούμε ένα μεγάλο αριθμό από Μαγνητικές Τομογραφίες εγκεφαλικών όγκων, στις οποίες η περιοχή του

όγκου είναι οριοθετημένη. Αυτές οι εικόνες παρέχονται ελεύθερα από τη βάση δεδομένων Multimodal Brain Tumor Image Segmentation (BRATS) MICCAI 2015.

Οι εικόνες που χρησιμοποιήσαμε για εκπαίδευση και αξιολόγηση αναφέρονται σε ενήλικους, άντρες και γυναίκες, περιλαμβάνοντας 24 υγιείς περιπτώσεις και 202 περιπτώσεις στις οποίες έχει παρατηρηθεί εγκεφαλικός όγκος. Όλες οι εικόνες έχουν εποπτευτεί οπτικά από τον έμπειρο νευροχειρούργο συνεργάτη μας. Οι εικόνες για τις υγιείς περιπτώσεις προμηθεύτηκαν από το Γενικό Νοσοκομείο Χανίων «Αγ. Γεώργιος» και από την βάση δεδομένων του Harvard General Hospital. Τα δεδομένα μας αποτελούνται από 5% υψηλού βαθμού, 82% χαμηλού βαθμού γλοιώματα, 3% μη-υγιείς αλλά μη αναγνωρίσιμες περιπτώσεις και 10% υγιείς περιπτώσεις.Στο στάδιο της προ-επεξεργασίας εφαρμόζουμε έναν αλγόριθμο απαλοιφής για να απομονώσουμε την περιοχή του εγκεφάλου. Στην συνέχεια, χρησιμοποιούμε ένα υψιπερατό Gaussian φίλτρο για καλύτερη ευκρίνεια και ένα median φίλτρο για απαλοιφή του θορύβου. Στο στάδιο της μετεπεξεργασίας χρησιμοποιούμε το Otsu κατώφλι για κατάτμηση της εικόνας καθώς και μορφολογικούς συντελεστές για τον ορισμό της περιοχής ενδιαφέροντος (ROI).Στην πρώτη προτεινόμενη μέθοδο, η εξαγωγή χαρακτηριστικών έγινε με τη χρήση Grey-level co-occurrence πίνακα (GLCM) και υπολογίστηκαν 13 στατιστικά χαρακτηριστικά. Στην δεύτερη μέθοδο η εξαγωγή χαρακτηριστικών έγινε με την χρήση Discrete Wavelet Transform (DWT) και ελαχιστοποίηση διαστάσεων με την μέθοδο Principal Components Analysis (PCA). Στην τρίτη μέθοδο οι δύο παραπάνω μέθοδοι συνδυάστηκαν και ο GLCM πίνακας υπολογίστηκε μετά την χρήση DWT, PCA, για την εξαγωγή των στατιστικών χαρακτηριστικών. Στην τελευταία μέθοδο ο Mean-Shift αλγόριθμος εφαρμόζεται για καλύτερα αποτελέσματα κατάτμησης της εικόνας και το στάδιο της εξαγωγής χαρακτηριστικών γίνεται σύμφωνα με την Τρίτη μέθοδο. Τα εξαγόμενα χαρακτηριστικά για κάθε μια από τις τέσσερις μεθόδους επεξεργάζονται από ένα νευρωνικό δίκτυο απλής τροφοδότησης με αλγόριθμο οπισθοδιάδοσης για την εκπαίδευσή του και για τις μεθόδους που περιέχουν GLCM πίνακα με ένα προσαρμοστικό νευρο-ασαφές σύστημα συμπερασμού (ANFIS).

Τα πειραματικά αποτελέσματα των προτεινόμενων μεθόδων επαληθεύτηκαν και αξιολογήθηκαν με βάση ένα σετ εικόνων για αξιολόγηση συγκριτικά με τις τιμές της ευαισθησίας, της ακρίβειας και της συνολικής ορθότητας, με τα καλύτερα αποτελέσματα να πετυχαίνουν 98.8%, 62.5% και 95.6% αντίστοιχα.

ACKNOWLEDGEMENTS

First of all I would like to thank my advisor from Technical University of Crete, Dr. Eleftheria Sergaki, for her guidance and support throughout this work and beyond that.

I also wish to extend my gratitude to my teacher Prof. Michael Zervakis for introducing me to the theory of Image analysis and supporting me in my Thesis and I would like to express gratitude to my committee member Prof. Konstantinos Kalaitzakis. Both of these professors are excellent in their respective classes.

Furthermore special thanks are given to the candidate PhD student Theodoros Kaptenakis and of course the neurosurgery Dr. Antonios Krasoudakis for all the assistance they game me to this Thesis.

Additionally I would like to thank all my close friends who made my time studying in Chania easier and smoother.

Last but more importantly I would like to thank my family and Maria Papathanasiou who supported me in every way possible through out these past years.

TABLE OF CONTENTS

ABSTRACT **i**

ΠΕΡΙΛΗΨΗ **iii**

ACKNOWLEDGEMENTS **vi**

TABLE OF CONTENTS **vii**

LIST OF TABLES **x**

LIST OF FIGURES **xi**

LIST OF SYMBOLS AND ABBREVIATIONS **xvii**

CHAPTER 1. INTRODUCTION **1**
1.1 Motivation and objectives of the study **1**
1.2 Outline of the dissertation **2**

CHAPTER 2. CLINICAL BACKGROUND **3**
2.1 Human Brain Anatomy **3**
2.1.1 Introduction 3
2.1.2 Basic brain tissues 3
2.2 Brain Tumour **5**
2.2.1 Introduction 5
2.2.2 Types of Brain Tumours 6
2.3 Magnetic Resonance Imaging **7**
2.3.1 Introduction 7
2.3.2 Basic MRI Sequences 7

CHAPTER 3. COMPUTER AIDED DIAGNOSIS AND RELATED STUDIES **9**
3.1 Computer Aided Diagnosis (CAD) Process **9**
3.2 Texture Analysis in CAD Process **9**
3.3 Pre-processing Stage of CAD Process **10**
3.4 Image Segmentation of CAD Process **11**
3.4.1 Thresholding segmentation method 12
3.4.2 Region-based segmentation methods 13
3.4.3 Classification methods 14

3.4.4 Clustering algorithms 14
3.5 Morphological Operations of the CAD Process 15
3.6 Feature Extraction and Dimensionality Reduction of CAD Process 16
3.6.1 Feature Extraction using Grey Level Co-Occurrence Matrix (GLCM) 17
3.6.2 Statistics features formulas 18
3.6.3 Future Extraction based on Discrete Wavelet Transform (DWT) 21
3.6.4 Dimensionality Reduction using PCA 24
3.7 Brain Regions Classification based on Texture Analysis using Artificial Intelligence (AI) 26
3.7.1 Artificial Neural Networks 26
3.7.2 Adaptive Neuro – Fuzzy Inference System (ANFIS) 28
3.8 Literature Review 33

CHAPTER 4. PROPOSED METHODS, MATERIALS AND IMPLEMENTATION WITH MATLAB 36
4.1 An Overview of the proposed Methods 36
4.2 Thesis Database Selection 38
4.3 Thesis Database Implementation: Pre-processing stage 39
4.3.1 Preparation of data used 39
4.3.2 Skull Stripping algorithm used 39
4.3.3 Sharpening using high pass filter 43
4.3.4 Reducing noise using Median filter 44
4.4 Thesis Database Implementation: Post-processing stage 45
4.4.1 Thesis proposed CAD Process Method 1: Otsu's Thresholding, Morphological Operators, GLCM 46
4.4.2 Thesis proposed CAD Process Method 2: Otsu's thresholding, Morphological Operators, DWT, PCA (3-level 2D decomposition via Daubechies 4 and SVD decomposition) 51
4.4.3 Thesis proposed CAD Process Method 3: Otsu's thresholding, Morphological Operators, DWT, PCA, GLCM 52
4.4.4 Thesis proposed CAD Process Method 4: Mean-Shift segmentation, Otsu thresholding, Morphological Operators, DWT, PCA, GLCM 53

CHAPTER 5. EXPERIMENTAL CLASSIFICATION RESULTS USING ANN AND ANFIS 55
5.1 Classification 55
5.1.1 Artificial Neural Network 55
5.1.2 ANFIS 56
5.2 Classification results of Method 1 57
5.2.1 Results of ANN 57
5.2.2 Results of ANFIS 64
5.3 Classification results of Method 2 68
5.3.1 Results of ANN 68
5.4 Classification results of Method 3 77
5.4.1 Results of ANN 77
5.4.2 Results of ANFIS 85
5.5 Classification results of Method 4 89

5.5.1 Results of ANN 89
5.5.2 Results of ANFIS 97
5.6 Classification results without pre-processing and ROI segmentation 101
5.6.1 Results based on GLCM 101
5.6.2 Results based on DWT and PCA 113

CHAPTER 6. CONCLUSION, DISCUSSION AND RECOMMENDATIONS 122
6.1 Conclusions 122
6.2 Recommendations 125

REFERENCES 127

LIST OF TABLES

Table 1: Hybrid algorithm procedure which combines the least square method and the gradient descent 33
Table 2: Classification results of related studies. [11] and [12] results are referred to a combination of training and testing dataset, while in [13] the results are referred to testing dataset only 35
Table 3: Thesis four Proposed CAD Process Methods 38
Table 4: Thesis results for case Method 1. Texture features of the used database images 50
Table 5: Thesis results. Comparison of different classification results for Method 1 using different amount of database cases 67
Table 6: Thesis results. Comparison of different classification for Method 2 for different number of database cases 76
Table 7: Thesis results. Comparison of different classification results for Method 3 for different number of database cases 88
Table 8: Thesis results. Comparison of classification results for Method 4 using different number of database cases 100
Table 9: Thesis results. Comparison of classification results from GLCM feature extraction without pre-processing and ROI segmentation 112
Table 10: Thesis results. Comparison of ANN classification results from DWT and PCA feature extraction without pre-processing and ROI segmentation 121
Table 11: Comparison of different classification results from the presented proposed Methods using the whole dataset 123
Table 12: Comparison of our proposed algorithm and the algorithms presented in [13]. 125

LIST OF FIGURES

Figure 1: Basic types of MRI sequences: T1, T2, Flair 8

Figure 2: Block diagramm of a typical methodology of CAD process 10

Figure 3: Example of computation of co-occurrence matrix for a given $\mathbf{4 \times 4}$ pixel image (a) with three grey levels (b). In this example, the matrix is computed in horizontal direction with one pixel separation. The number o transitions of grey-levels is counted and allocated in the co-occurrence matrix (c). 18

Figure 4: Block diagram of a 3-level wavelet decomposition tree 23

Figure 5: Schematic diagram of 3-level 2D DWT 24

Figure 6: Example of PCA transformation of 3-dimensional data to 2-dimensional data 25

Figure 7: Model of artificial neuron 26

Figure 8: Model of neural network 27

Figure 9: ANFIS Architecture 29

Figure 10: Thesis Flowchart of CAD process proposed Methods 37

Figure 11: Thesis steps used for skull stripping algorithm based on thresholding segmentation and morphological operations 40

Figure 12: .Thesis results. Example of the application of skull stripping algorithm used. (a) initial MRI image, (b) MRI image after the application of skull stripping algorithm. 42

Figure 13: Thesis results. Example of the application of sharpening filter. (a) high-pass filter, (b) original MRI image, (c) enhanced details of original image, (d) sharpened MRI image 44

Figure 14: Thesis results. Example of the application of median filter. (a) image after high pass filter, (b) image after median filter 45

Figure 15: Thesis results. Examples of MRI images segmented by Otsu's Thresholding. (a) Median filtered image with tumor, (b) segmented image with tumor, (c) Median filtered image without tumor, (d) segmented image without tumor 47

Figure 16: Thesis results. Examples of application of erosion and opening by reconstruction morphological operations. (a) is a segmented MRI image, (b) is the MRI image after morphological operations 49

Figure17: Thesis result of Method 2. Variances against No. of principle components 52

Figure 18: Thesis results of Method 4. Example of application of Mean-Shift algorithm on the pre-processing image, (a) pre-processed image, (b) image after Mean-Shift algorithm 54

Figure 19: Thesis results. Architecture of Backpropagation Neural Network 57

Figure 20: Thesis results of training ANN on Method 1 using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM) 58

Figure 21: Thesis results of ANN on Method 1 using a subset of the database. Plot performance with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM) 59

Figure 22: Thesis results of testing of ANN on Method 1 case of subset data. Confusion matrix of testing data with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM)... 60

Figure 23: Thesis results of training ANN on Method 1 using the whole dataset. Confusion matrix of training data with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM) 61

Figure 24: Thesis results of ANN on Method 1 using the whole dataset. Plot performance with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM) 62

Figure 25: Thesis results of testing ANN on Method 1 using the whole dataset. Confusion matrix of testing data with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM)... 63

Figure 26: Thesis results of training ANFIS on Method 1 using a subset of the database. Training plot of ANFIS for Method 1(Otsu's Thresholding, Morphological Operators, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness..... 64

Figure 27: Thesis results of testing ANFIS on Method 1 using a subset of the database. Testing plot of ANFIS for Method 1(Otsu's Thresholding, Morphological Operators, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness..... 65

Figure 28: Thesis results of testing ANFIS on Method 1 using a subset of the database. Confusion matrix of testing data with ANFIS for Method 1(Otsu's Thresholding, Morphological Operators, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness 66

Figure 29: Thesis results of Method 2. Architecture of Backpropagation Neural Network 68

Figure 30: Thesis results of training ANN on Method 2 using a subset of the database. Confusion matrix of training data with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA)..... 69

Figure 31: Thesis results of testing ANN on Method 2 using a subset of the database. Plot performance with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA) 70

Figure 32: Thesis results of testing ANN on Method 2 using a subset of the database. Confusion matrix of testing data with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA)..... 71

Figure 33: Thesis results of Method 2. Architecture of backpropagation neural network 72

Figure 34: Thesis results of training ANN on Method 2 using the whole dataset. Confusion matrix of training data with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA)..... 73

Figure 35: Thesis results of testing ANN on Method 1 using the whole dataset. Plot performance with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA) 74

Figure 36: Thesis results of testing ANN on Method 2 using the whole dataset. Confusion matrix of testing data with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA)..... 75

Figure 37: Thesis results for Method 3. Architecture of backpropagation neural network 77

Figure 38: Thesis results of training ANN on Method 3 using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity and kurtosis based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)..... 78

Figure 39: Thesis results of testing ANN on Method 3 using a subset of the database. Plot performance with inputs the features of energy, homogeneity and kurtosis based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)..... 79

Figure 40: Thesis results of testing ANN on Method 3 using a subset of the database. Confusion matrix of testing data with inputs the features of energy, homogeneity and kurtosis based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)..... 80

Figure 41: Thesis results for Method 3. Architecture of backpropagation neural network 81

Figure 42: Thesis results of training ANN on Method 3 using the whole dataset. Confusion matrix of training data with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)..... 82

Figure 43: Thesis results of testing ANN on Method 3 using the whole dataset. Plot performance with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) 83

Figure 44: Thesis results of testing ANN on Method 3 using the whole dataset. Confusion matrix of testing data with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 3 (Otsus Thresholding, Morphological Operators, DWT, PCA, GLCM) 84

Figure 45: Thesis training results of ANFIS on Method 3 using the whole dataset. Training plot of ANFIS for Method 3(Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness 85

Figure 46: Thesis results of testing ANFIS on Method 3 using the whole dataset. Testing plot of ANFIS for Method 3(Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness..... 86

Figure 47: Thesis results of testing ANFIS on Method 3 using the whole dataset. Confusion matrix of testing data with ANFIS for Method 3(Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness 87

Figure 48: Thesis results of Method 4. Architecture of backpropagation neural network 89

Figure 49: Thesis results of training ANN on Method 4 using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity and kurtosis based on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) 90

Figure 50: Thesis results of testing ANN on Method 4 using a subset of the database. Plot performance with inputs the features of energy, homogeneity and kurtosis based on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)..... 91

Figure 51: Thesis results of testing ANN on Method 4 using a subset of the database. Confusion matrix of testing data with inputs the features of energy, homogeneity and kurtosis based on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) 92

Figure 52: Thesis rsults for Method 4. Architecture of backpropagation neural network 93

Figure 53: Thesis results of training ANN on Method 4 using the whole dataset. Confusion matrix of training data with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) 94

Figure 54: Thesis results of testing ANN on Method 4 using the whole dataset. Plot performance with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) 95

Figure 55: Thesis results of testing ANN on Method 4 using the whole dataset. Confusion matrix of testing data with inputs the features of energy, homogeneity, kurtosis and skewnessbased on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) 96

Figure 56: Thesis results of training ANFIS on Method 4 using the whole dataset. Training plot of ANFIS for Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)with inputs the features of energy, homogeneity and kurtosis 97

Figure 57: Thesis results of testing ANFIS on Method 4 using the whole dataset. Testing plot of ANFIS for Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity and kurtosis 98

Figure 58: Thesis results of testing ANFIS on Method 4 using the whole dataset. Confusion matrix of testing data with ANFIS for Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity and kurtosis 99

Figure 59: Thesis results based on GLCM. Architecture of backpropagation neural network 101

Figure 60: Thesis results of ANN based on GLCM using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity, skewness and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM 102

Figure 61: Thesis results of ANN based on GLCM using a subset of the database. Plot performance of neural network with inputs the features of energy, homogeneity, skewness and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM 103

Figure 62: Thesis results of ANN based on GLCM using a subset of the database. Confusion matrix of testing data with inputs the features of energy, homogeneity, skewness and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM 104

Figure 63: Thesis results of ANN based on GLCM using the whole dataset. Architecture of backpropagation neural network 105

Figure 64: Thesis results of ANN based on GLCM using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM..... 106

Figure 65: Thesis results of ANN based on GLCM using a subset of the database. Plot performance of neural network with inputs the features of energy, homogeneity and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM..... 107

Figure 66: Thesis results of ANN based on GLCM using the whole dataset. Confusion matrix of testing data with inputs the features of energy, homogeneity and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM 108

Figure 67: Thesis results of ANFIS based on GLCM using the whole dataset. Training plot of ANFIS with inputs the features of energy, homogeneity and kurtosis that were

extracted with GLCM matrix from MRI without pre-processing and ROI segmentation 109

Figure 68: Thesis results of ANFIS based on GLCM using the whole dataset. Testing plot of ANFIS with inputs the features of energy, homogeneity and kurtosis that were extracted with GLCM matrix from MRI without pre-processing and ROI segmentation 110

Figure 69: Thesis results of ANFIS based on GLCM using the whole dataset. Confusion matrix of testing data with ANFIS, with inputs the features of energy, homogeneity and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM..... 111

Figure 70: Theis rsults. Architecture of backpropagation neural network 113

Figure 71: Thesis results of ANN based on PCA using a subset of the database. Confusion matrix of training data with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation 114

Figure 72: Thesis results of ANN based on PCA using a subset of the dataset. Plot performance of neural network with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation 115

Figure 73: Thesis results of ANN based on PCA using a subset of the database. Confusion matrix of testing data with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation 116

Figure 74: Thesis results. Architecture of backpropagation neural network 117

Figure 75: Thesis results of ANN based on PCA using the whole dataset. Confusion matrix of training data with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation..... 118

Figure 76: Thesis results of ANN based on PCA using the whole dataset. Plot performance of neural network with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation 119

Figure 77: Thesis results of ANN based on PCA using the whole dataset. Confusion matrix of testing data with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation..... 120

Figure 78: Thesis results. ROC curve of testing data of the proposed algorithm for classification of human brain MRI images using ANN clasifier..... 124

LIST OF SYMBOLS AND ABBREVIATIONS

CAD	Computer Aided Diagnosis
MRI	Magnetic Resonance Imaging
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
GM	Grey Matter
WM	White Matter
CSF	Cerebrospinal Fluid
CNS	Central Nervous System
CT	Computed Tomography
FLAIR	Fluid Attenuated Inversion Recovery
ROI	Region of Interest
FCM	Fuzzy C-means
PDF	Probability Density Function
LBP	Local Binary Patterns
DWT	Discrete Wavelet Transform
PCA	Principal Components Analysis
ICA	Independent Components Analysis
LDA	Linear Discriminant Analysis
GLCM	Grey Level Co-Occurrence Matrix
IDM	Inverse Difference Moment

CHAPTER 1. INTRODUCTION

1.1 Motivation and objectives of the study

In this thesis we examine the problem of Computer Aided Diagnosis (CAD) of brain tumors. The classification and detection of brain tumors from different medical images demands high accuracy since it is crucial for human life. Even now, the most accurate classification and detection technique for detecting and recognizing a brain tumor is the manual segmentation from an expert. Although manual segmentation and classification provides the best accuracy results it is a time consuming procedure for the expert who is burdened to process a great amount of such cases. In addition, manual segmentation and classification has high inter and intra-observer variability. Automated classification aims to reduce human error by assisting experts with some software that could lead to better results.

The main objective of this thesis is to examine different classification techniques for the problem of recognizing a healthy person from a patient that contains a brain tumor, based on a magnetic resonance imaging (MRI) of T2 weighted modality from the axial plane. For the completion of this task specific methods are examined, optimized and finally combined in four consistent methodologies which are presented in this Thesis. The proposed methodologies are named as “Method 1 to 4”. These methodologies are based on image pre-processing, classification with neural networks (NN) and a hybrid combination of neural network and fuzzy logic (ANFIS). Hence, the study of several pre-processing techniques, segmentation algorithms and classification algorithms had to be done.

1.2 Outline of the dissertation

Chapter 1 gives an introduction to the problem statement and the objectives of the study. **Chapter 2** provides a clinical background of human brain, types of brain tumors and magnetic resonance imaging (MRI). **Chapter 3** describes the structure of a Computer Aided Diagnosis (CAD) system as well as the theoretical background of the techniques that are used. **Chapter 4** demonstrates analytically the four proposed methodologies that are implemented. **Chapter 5** contains the experimental results of our proposed methodologies. **Chapter 6** summarizes, concludes the present Thesis and proposes recommendations for future work.

CHAPTER 2. CLINICAL BACKGROUND

2.1 Human Brain Anatomy

2.1.1 Introduction

As mentioned in [1] the human brain is the central organ of the human nervous system. It controls most of the activities of the body. By interpreting and managing the information that receives by the sense organs, it sends instructions to the whole body. The human brain is contained and protected by the skull bones and three layers of tissue, called meninges. It consists of three major parts, the cerebrum, the brain stem and the cerebellum. The cerebrum is the largest part of the brain and has two cerebral hemispheres. It covers the whole area of cerebral cortex which is an outer layer of gray matter, covering the core of white matter. The cerebral is responsible for the procedures of abstract thought, self-control and planning. The brainstem controls breathing, heart rate and other autonomic processes, while cerebellum is responsible for the body's balance, posture and the coordination of movement.

2.1.2 Basic brain tissues

The basic brain tissues that human brain contains are the grey matter (GM), the white matter (WM) and the cerebrospinal fluid (CSF).

The grey matter (GM) as mentioned in [1] is the major component of central nervous system (CNS), consisting of neuronal cell bodies, neuropil (dendrites and myelinated as well as unmyelinated axons), glial cells, synapses and capillaries. We can

separate grey matter from white matter due to the large quantity of cell bodies and relatively few myelinated axons that grey matter contains. Unlike the grey matter the white matter contains many myelinated axons and relatively few cell bodies. The grey matter is distributed at the surface of the cerebral cortex and the cerebellum, as well as the depths of cerebrum and the brainstem.

The white matter (WM) as mentioned in [1] refers to areas of the central nervous system (CVS) that are made up of myelinated axons. White matter constitutes the deep part of the brain and the superficial parts of the spinal cord. It affects learning and brain functions by coordinating communication between brain regions. White matter named for its white colour derived from the lipid content of myelin which surrounds the nerve fibres (axons). It is composed by bundles which connects grey matter areas and carry nerve impulses between neurons.

Cerebrospinal fluid (CSF) as mentioned in [1] is a clear, colourless body fluid found in the brain and spinal cord. It acts as a 'cushion' or a buffer for the brain providing basic mechanical and immunological protection to the brain inside the skull as well as auto regulation of cerebral blood flow. There is about 125-150 mL of CSF at any time in the brain. This CSF circulates within the ventricular system of the brain. It is been produced mainly from the two lateral ventricles and then passes through the interventricular foramina to the third ventricle, the cerebral aqueduct, the fourth ventricle and from there to the subarachnoid space. CSF contains around 0.3% protein or approximately 15 to 40 mg/mL, depending on sampling site.

CSF is important mainly for five reasons:

1. Buoyancy: The actual mass of human brain is about 1400-1500 grams; however the net brain suspended in the CSF is equivalent to a mass of 25-50 grams. The brain therefore exists in neutral buoyancy, which allows the brain to maintain its density without being impaired by its own weight, which cut off blood supply and kill neurons in the lower sections without CSF.
2. Protection: CSF protects the brain tissue from injury when jolted or hit, by providing a fluid buffer that acts as a shock absorber from some forms of mechanical injury.
3. Prevention of brain ischemia: The prevention of brain ischemia is made by decreasing the amount of CSF in the limited space inside the skull. This decreases total intracranial pressure and facilitates blood perfusion.
4. Homeostasis: CSF allows for regulation of the distribution of substances between cells of the brain, and neuroendocrine factors, to which slight changes can cause problems or damage the nervous system.
5. Clearing waste: CSF allows for the removal of waste products of the brain, and in the brain's lymphatic system. Metabolic products diffuse rapidly into the CSF and are removed into the bloodstream as CSF is absorbed.

2.2 Brain Tumour

2.2.1 Introduction

A brain tumour is an abnormal growth of tissue that creates a mass (neoplasm) within the brain or the central spinal canal [2]. Brain tumours can be divided in two main

types, malignant or cancerous and benign brain tumours. Malignant tumours have two subcategories, primary tumours that are created in the brain and metastatic or secondary tumours that have been started from somewhere else in the body and spread to the brain. Brain tumours can also be categorized by a grading system.

2.2.2 Types of Brain Tumours

1. Benign or Malignant Tumours: The main characteristic of a benign tumour is that it comprises slow growing cells and has homogeneity in structure. Benign tumours can be dangerous due to their location where they push soft tissues of the brain. Malignant tumours considered life threating because of their aggressive and invasive nature and consist of active (cancerous) cells that spread quickly.
2. Primary or Secondary Tumours: As mentioned before primary tumors are created in the brain and may be benign or malignant. Secondary or metastatic tumors are malignant tumors that caused by a primary tumor somewhere in the body and invaded brain via lymphatic system and blood vessels, circulating through the bloodstream.
3. Grading System: Tumours are categorized by low-grade and high-grade. Low-grade tumours like glioma considered benign tumours while high-grade tumours like glioblastoma considered malignant tumours.

2.3 Magnetic Resonance Imaging

2.3.1 Introduction

As mentioned in [3] brain tumour examination can be done by CT tomography although this examination may cause small or low grade tumors to be missed. Magnetic resonance imaging (MRI) is significantly more sensitive to the presence of tumor. MRI has a high degree of confidence in the diagnosis of high grade tumors and is widely used for identifying location and size of brain tumors. MRI examination has several advantages as it is a non-invasive method of diagnosis, it does not cause any radiation damage to the patient's tissue and it is preferred for better contrast imaging of cancerous tissues [4].

An MRI machine uses a powerful magnetic field to align the magnetization of proteins in the body, and radio frequency fields to systematically alter the alignment of this magnetization. This causes the protons to produce a rotation magnetic field of larger frequency detectable by the scanner and this information is recorded to construct an image of the scanned area of the body [5].

2.3.2 Basic MRI Sequences

The basic types of MRI are T1-weighted, T2-weighted and Fluid Attenuated Inversion Recovery (FLAIR) MRI. In T1-weighted MRI we can observe better anatomical details, the CSF fluid appears dark, the grey matter grey and the white matter in white colour. The T1-weigheted images are produced by using short TE and TR times. In contrast, T2-weigheted images are produced by using longer TE and TR times. In T2-weigheted scans the CSF fluid appears white, the grey matter light grey and the white

matter dark. In the third basic sequence, Flair, the TE and TR times are very long. By doing this, abnormalities remain bright but CSF fluid is made dark. The Flair sequence is the best scan for a quick look for pathology [6].

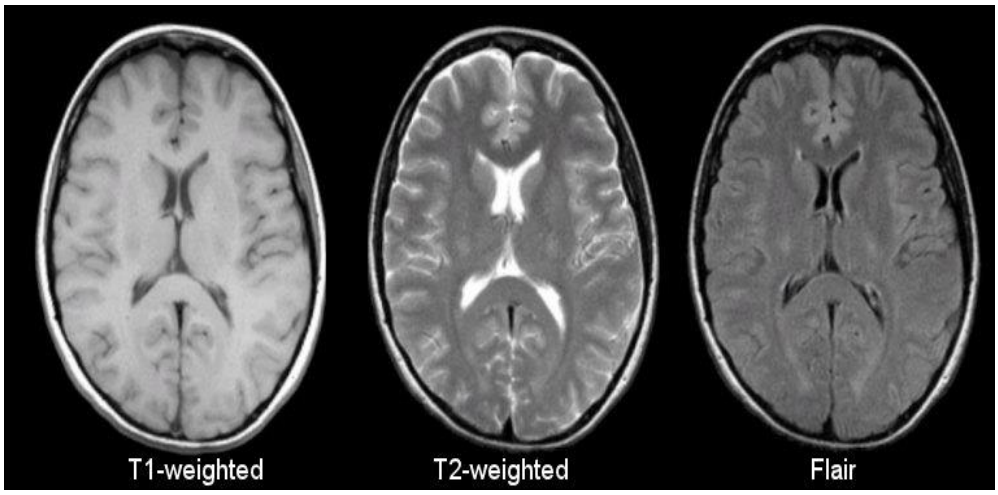

Figure 1: Basic types of MRI sequences: T1, T2, Flair

CHAPTER 3. COMPUTER AIDED DIAGNOSIS AND RELATED STUDIES

3.1 Computer Aided Diagnosis (CAD) Process

One of the major aspects for encountering a brain tumor is the early detection of the cancerous cells. As mentioned before MRI is the most common sequence for detecting a tumor area and the task of radiologists is very critical. Computer Aided Diagnosis systems can provide a great help in analyzing the tumor area and assist radiologists in interpreting medical images by providing a "second opinion". In this chapter we present and describe the basic theory of each stage of a CAD process. The CAD system can improve the diagnostic accuracy of radiologists as they lighten the burden of increased workload and reduce cancer missed due to fatigue [20]. CAD process is based in machine learning and pattern recognition techniques for learning and recognizing brain diseases. To create CAD systems several image processing techniques under the spectrum of texture analysis are necessary.

3.2 Texture Analysis in CAD Process

After applying images from MRI sequence it is necessary to process these images with several image processing techniques that are suitable for automated diagnosis. The figure below shows the basic structure of a CAD process that is composed of:

1. a preprocessing module
2. region of interest (ROI) segmentation
3. feature extraction

4. dimensionality reduction
5. classification and evaluation

Diagnosis is always based on the opinion of an expert radiologist or a neurologist or a neurosurgeon doctor.

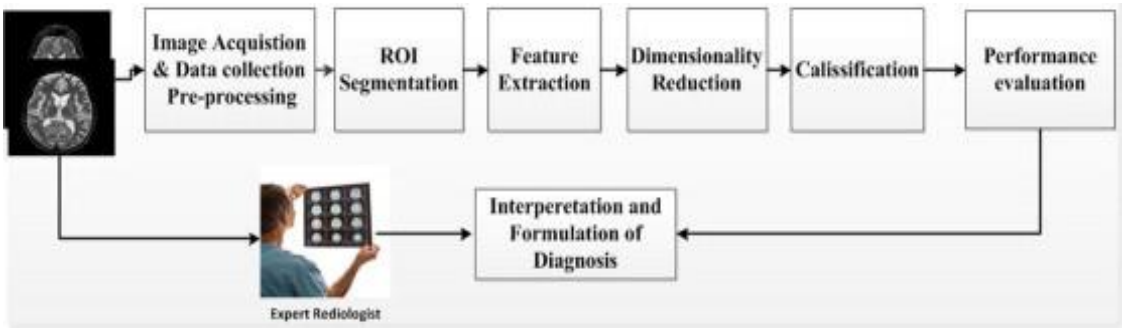

Figure 2: Block diagramm of a typical methodology of CAD process

3.3 Pre-processing Stage of CAD Process

In the pre-processing stage we use methods in order to reduce the noise and improve resolution contrast of the image. It is supposed that MRI images are affected by Gaussian noise. For de-noising these images there is a variety of proposed methods such as linear and non-linear filters, anisotropic diffusion algorithm and wavelet analysis. Although these methods exhibit similarities in the quality of de-noising, the computational time and the preservation of edges.

In linear filters output values are linear combinations of the original's image pixels. With this process linear filters reduce noise but influence negatively on the edges and the details of the image and lead to blur images. To avoid this phenomenon non-linear filters

are used which update the value of a pixel based on the transformation applied. The new value would be based on the pixel values that are nearer to the local neighborhood and a centralized window on that pixel [21]. Median filter is a common method for de-noising by smoothing the image and improve its quality. It preserves the edges of the processed image. The main idea of median filter is to run through the image pixel by pixel, replacing each value with the median value of neighbouring pixels. The pattern of neighbours is called the “window”.

Intensity normalization of pixels is an important method of pre-processing an MRI image when we use classification algorithms and image segmentation techniques and it is based on histogram equalization methods. To improve contrast resolution, techniques for sharpening features like edges and textures are used. Un-sharp masking is a technique where the fraction of a high-pass filtered image is added to the original image. With this method we enhance the local difference between intensity values by subtracting the smoothed image from the original image. Another well-known method for contrast enhancing is Contrast Enhanced Adaptive Histogram Equalization [22].

3.4 Image Segmentation of CAD Process

Image segmentation refers to the process of partitioning a digital image into multiple regions which have different features between them such as intensity, colour, texture or other statistical properties. Every region is a set of pixels with similar features. Segmentation methods is used for locate and identify objects and boundaries in an image. The union of all the segments – regions results in the whole image. There is many segmentation techniques, however there is no standard segmentation technique that works

well for all images. Segmentation techniques can differ by detecting discontinuities like edge detection or detecting similarities like thresholding, region growing, region splitting and merging [23]. Image segmentation algorithms are categorised in three basic categories below:

- Thresholding algorithms
- Region based algorithms
- Clustering and classification algorithms

3.4.1 Thresholding segmentation method

*A*s mentioned in [10], is a simple and effective region segmentation method in which the objects of the image are classified by comparing their intensities with one or more thresholds. If we make the assumption that the objects and the background of the image have a bimodal distribution then the objects can be separated from the background by a single threshold called global thresholding. However, if the image contains multiple regions corresponding to different objects then it is better to use local thresholding. Thresholding segmentation creates a binary image from a gray-level one. Global thresholding turns all pixels below the threshold to zero and all pixels above that threshold in one, so if $g(x, y)$ is a thresholded version of $f(x, y)$ at some global threshold T then,

$$g(x,y) = \begin{cases} 1, if\ f(x,y) \geq T \\ 0, \quad otherwise \end{cases} \qquad \text{(Eq.1)}$$

in which pixels with the value of 1 corresponds to the region of interest (ROI) and the values of 0 to the background. The major problem of thresholding segmentation is that only

the intensity is considered, not any relationship between the pixels. This may lead to pixels that are not in the desired region to be included and pixels within the desired region to be excluded. This effect is relevant to the noise.

3.4.2 Region-based segmentation methods

Region-based segmentation methods examine pixels in an image and form regions by merging neighbourhood pixels with homogeneity properties based on a predefined similarity criterion. The region growing and the watershed segmentation methods are the most commonly region-based methods used for brain tumour segmentation.

Region growing starts with at least one seed that belongs to the structure of interest [7]. Seeds can be chosen manually or with an automated seed-finding procedure. Neighbours of the seed are checked and those who satisfy the similarity criteria are added to the region. The similarity criteria is based on intensity values or other features and the whole procedure stops when no more neighbour pixels can be added to the region. The main disadvantage of region growing method is the partial volume effect which blurs the intensity distinction between regions at the border of two regions, because the voxel may represent more than one kind of tissue types, in brain segmentation.

Watershed segmentation method [7] can be explained by a metaphor based on the behaviour of water in landscape. When it rains, drops of water falling in different regions will follow the landscape downhill. The water will end up at the bottom of valleys. Each valley is associated with a catchments basin. At points where water coming from different basins meets, dams will be built. When the water level has reached the highest peak in the landscape, the process ends. As a result, the landscape is partitioned into region separated

by damns, called watershed lines or watersheds. It produces a complete contour of the images.

3.4.3 Classification methods

Image Segmentation based on classification methods in brain tumour segmentation are used to determine the regions of an image. Classification methods are *pattern recognition methods* with labelled data, where based on several features we can separate regions. The main drawback of this method is that training data are required and acquiring such medical data is a difficult task. *Artificial Neural Networks* (ANN) are commonly used for this classification approach.

3.4.4 Clustering algorithms

Image Segmentation based on clustering algorithms are similar with classification algorithms however clustering techniques are unsupervised. This means that there is no labelled data and a training phase. Clustering is the process of grouping objects with similar characteristics in one cluster, while other objects with different similarities in other clusters. The similarity criterion is based on a distance measure such as Euclidean distance. Each cluster is formed around a centroid iteratively based on their distance from this centroid and when a cost function is been optimised the algorithm stops and the clusters take their final form. Such unsupervised algorithms are *K-means* and *Fuzzy C-means (FCM)*.

At the stage of the CAD post-processing process, one of the proposed methodologies of the present thesis is the clustering algorithm named *Mean-Shift* [8].

Mean-Shift is a non-parametric iterative algorithm that can be used for finding modes, clustering, etc. The main idea of this algorithm is that *Mean-shift* considers feature space as an empirical probability density function (pdf). If dense regions, clusters, are present in the feature space, then they correspond to the mode which is considered the local maxima of the probability density function. *Mean-Shift* associates each data point with the nearest peak of the pdf. For each data point the algorithm defines a window around it and computes the mean for this defined dataset. Then it shifts the centre of the window to this mean and repeats the algorithm till it converges. The goal of this iterative procedure is that after each iteration the algorithm shifts to more dense regions.

So if we present the algorithm in steps, these should be like:

1. Define a window around each data point
2. Compute the mean of data belonging to this window
3. Shift the window to the mean and repeat until it converges

3.5 Morphological Operations of the CAD Process

Morphological operations are a collection of non-linear operations related to the shape or morphology of features in an image. Morphological operators probe an image with a small shape or template called a structuring element. The structuring element is positioned at all possible locations in the image and it is compared with the corresponding neighborhood of pixels. The basic operations of morphological processing are opening, closing, erosion and dilation.

With erosion, pixels on the boundaries of the objects are removed and this corresponds to the deletion of small objects. The number of pixels removed from the image depends on the size and shape of the structuring element used to erode the image. The process of erosion can be translated as the logical operator *AND* between the structuring element and the pixels of the binary image.

Opening by reconstruction can be thought of conceptually as repeated dilations of an image, called the *marker image,* until the contour of the marker image fits under a second image, called the *mask image.* It differs with dilation because this process is based on the connectivity of pixels rather than a structuring element and is based on two images.

3.6 Feature Extraction and Dimensionality Reduction of CAD Process

Feature extraction is the transformation of an image into a set of useful for classification features which have been extracted from this image. Feature extraction and selection are important stages for classification and detection of brain tumours. An optimum feature set should have features that sustain the information and reduce the dimensionality of the primitive feature space. Feature extraction methods are divided in:

1. *statistical methods*
2. *model-based methods*
3. *transform methods*

Co-occurrence matrix and *local binary patterns (LBP)* are examples of statistical methods, *fractal models* are a common model-based technique and *wavelet transform methods* are used as transform methods algorithms.

Many CAD systems perform *discrete wavelet transform (DWT)* decomposition to obtain wavelet coefficients at different levels. It is a common technique to use a *Principal components analysis (PCA), independent components analysis (ICA) and linear discriminant analysis (LDA)* after DWT for dimensionality reduction of data and feature selection.

3.6.1 Feature Extraction using Grey Level Co-Occurrence Matrix (GLCM)

For the collection of higher-level information of an image such as shape, texture and contrast, the Grey Level Co-occurrence Matrix (GLCM) was used. GLCM is one of the most widely used image analysis applications introduced by Haralick et al. [28]. The co-occurrence matrix allows extraction of statistical information regarding the distribution of the pixels pairs. The technique follows two steps, first it computes the GLCM matrix and then the texture features based on the GLCM are calculated. Pairs of pixels are separated by a predefined distance based on certain directions in the image and the resulting values are allocated in the co-occurrence matrix. The count is based on the number of pairs of pixels that have the same distribution of grey level values as shown in Figure 3. For 2D images, co-occurrence matrices are computed in the following directions: horizontal, vertical, 45^o, 135^o.

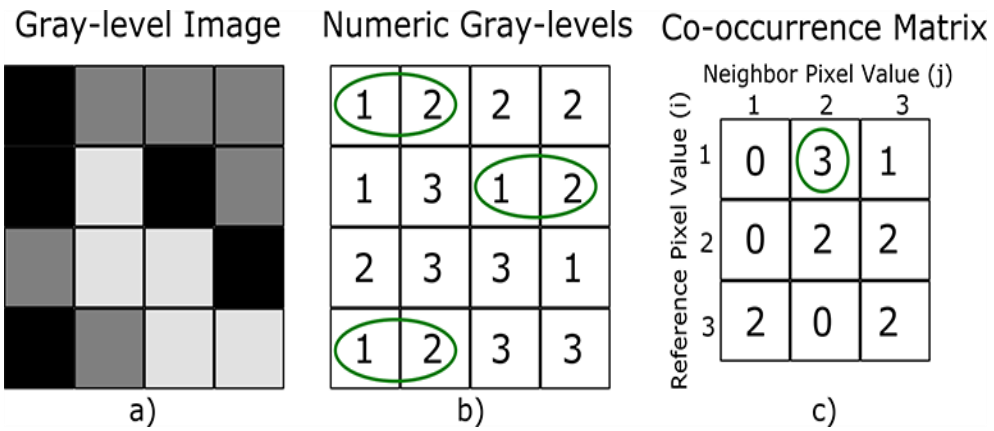

Figure 3: Example of computation of co-occurrence matrix for a given 4×4 pixel image (a) with three grey levels (b). In this example, the matrix is computed in horizontal direction with one pixel separation. The number o transitions of grey-levels is counted and allocated in the co-occurrence matrix (c).

3.6.2 Statistics features formulas

Textural findings and analysis could improve the diagnosis, tumor staging and therapy response assessment. The statistics features formula for the useful features are listed below.

1. *Mean (M).* The mean of an image is calculated by adding all the pixel values of an image divided by the total number of pixels in an image.

$$\mathbf{M} = \left(\frac{1}{m \times n}\right) \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) \qquad \text{(Eq.2)}$$

2. *Variance (V)* is a measure of variation. A small value of variance means that the data points tend to be very close to the mean while high variance means that data point are spread out from the mean and hence from each other.

$$\mathbf{V} = \left(\frac{1}{m \times n}\right) \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (f(x,y) - M) \quad \text{(Eq.3)}$$

3. *Standard Deviation (SD)* is the second central moment describing probability distribution of an observed population and can serve as a measure of inhomogeneity. A higher value indicates better intensity level and high contrast of edges in an image.

$$\mathbf{SD} = \sqrt{\left(\frac{1}{m \times n}\right) \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (f(x,y) - M)^2} \quad \text{(Eq.4)}$$

4. *Entropy* (*E*) is calculated to characterize the randomness of the textural image and is defined as

$$\mathbf{E} = -\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f(x,y)\, log_2 f(x,y) \quad \text{(Eq.5)}$$

5. *Skewness* (S_k) is a measure of symmetry or the lack of symmetry.

$$S_k(\mathrm{X}) = \left(\frac{1}{m \times n}\right) \frac{\sum (f(x,y) - M)^3}{SD^3} \quad \text{(Eq.6)}$$

6. *Kurtosis* (K_{urt}) describes the shape of a random variable's probability distribution.

$$K_{urt} = \left(\frac{1}{m \times n}\right) \frac{\sum (f(x,y) - M)^4}{SD^4} \quad \text{(Eq.7)}$$

7. *Energy* (E_n) can be defined as the quantifiable amount of the extent of pixel pair repetitions. Energy is a parameter to measure the similarity of an image.

$$\boldsymbol{E_n = \sqrt{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} f^2(x,y)}} \quad \text{(Eq.8)}$$

8. *Contrast* (C_{on}) is a measure of intensity of a pixel and its neighbor over the image, and is defined as

$$\boldsymbol{C_{on} = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(x-y)^2 f(x,y)} \quad \text{(Eq.9)}$$

9. *Inverse Difference Moment (IDM) or Homogeneity* is a measure of local homogeneity of an image. IDM may have a single or a range of values so as to determine whether the image is textured or not.

$$\mathbf{H} = \boldsymbol{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}\frac{f(x,y)}{1+(x-y)}} \quad \text{(Eq.10)}$$

10. *Correlation (C_{orr})* describes the spatial dependencies between the pixels and it is defined as

$$\boldsymbol{C_{orr} = \frac{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(x,y)f(x,y) - M_x M_y}{\sigma_x \sigma_y}} \quad \text{(Eq.11)}$$

11. *Smoothness (S_m)*

$$\mathbf{Sm = 1 - \left(\frac{1}{1+\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} f(x,y)}\right)} \quad \text{(Eq.12)}$$

12. *Root Mean Square (RMS)*

$$\mathbf{RMS = \frac{1}{N}\sum_{i=1}^{N} X_{RMS}} \quad \text{(Eq.13)}$$

with M_x and σ_χ are the mean and standard deviation in the horizontal spatial domain, M_y and σ_y are the mean and the standard deviation in the vertical domain and X_{RMS} $= \sqrt{\frac{1}{N}\sum_{i=1}^{N}|x_n|^2}$.

3.6.3 Future Extraction based on Discrete Wavelet Transform (DWT)

Discrete wavelet transform has become the method of choice for many image analysis and classification problems because it gives information about the signal, both in frequency and in time domains. DWT performs the function of transforming images from the spatial domain into the frequency domain. By applying DWT, we are able to decompose an image into the corresponding sub-bands with their relative DWT coefficients [26]. The DWT is implemented by using cascaded filter banks in which a *low and the high pass filter* satisfy certain criteria. As a result of this decomposition there are *four sub-band (LL, LH, HH, HL)* images at each scale. The LL sub-band can be regarded as the approximation component of the image, while the other three sub-bands can be regarded as the detailed components of the image. For every step of the decomposition only the LL sub-band is used for producing the next level. The fundamentals of DWT can

be described as follows [27]. Suppose $x(t)$ is a square integrable function, then the continuous wavelet transform of $x(t)$ relative to a given $\psi(t)$ is defined as

$$\boldsymbol{W_{\psi}(a,b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)dt} \quad \text{(Eq.14)}$$

where

$$\boldsymbol{\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-a}{b}\right)} \quad \text{(Eq.15)}$$

Here, the wavelet $\psi_{a,b}(t)$ is calculated from the mother wavelet $\psi(t)$ by the translation and dilation factors: *a* is the dilation factor and *b* the translation parameter (both real positive numbers). Eq.14 can be discretized by restraining *a* and *b* to a discrete lattice ($a = 2^b \; and \; a > 0$) to give the DWT, which can be expressed as follows.

$$\boldsymbol{ca_{j,k}(n) = DS\left[\sum_{n} x(n)g_j^*(n - 2^j k)\right]} \quad \text{(Eq.16)}$$

$$\boldsymbol{cd_{j,k}(n) = DS\left[\sum_{n} x(n)h_j^*(n - 2^j k)\right]} \quad \text{(Eq.17)}$$

where $ca_{j,k}$ and $cd_{j,k}$ refer to the coefficients of the approximation components and the detail components, respectively. The low-pass and the high-pass filter are denoted by $g(n)$ and $h(n)$. The wavelet scale and the translation factor are denoted by *j* and *k*, while *DS* operator means the down sampling. Eq.16 and Eq.17 are the fundamental equations for wavelet decomposition. They decompose $x(n)$ into two signals, the *approximation*

coefficients $ca(n)$ and the *detail components* $cd(n)$. This procedure is called *one level decompose*. The above decomposition process can be iterated with successive approximations being decomposed in turn, so that the one signal is broken down into various levels of resolution. The whole process is called *wavelet decomposition tree*, show in Figure 4.

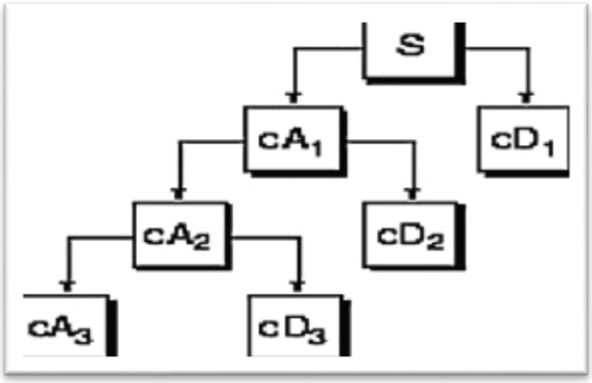

Figure 4: Block diagram of a 3-level wavelet decomposition tree

In case of 2D images, the DWT is applied to each dimension separately. As a result there are four sub-bands as they described before and Figure 5 demonstrates their schematic diagram.

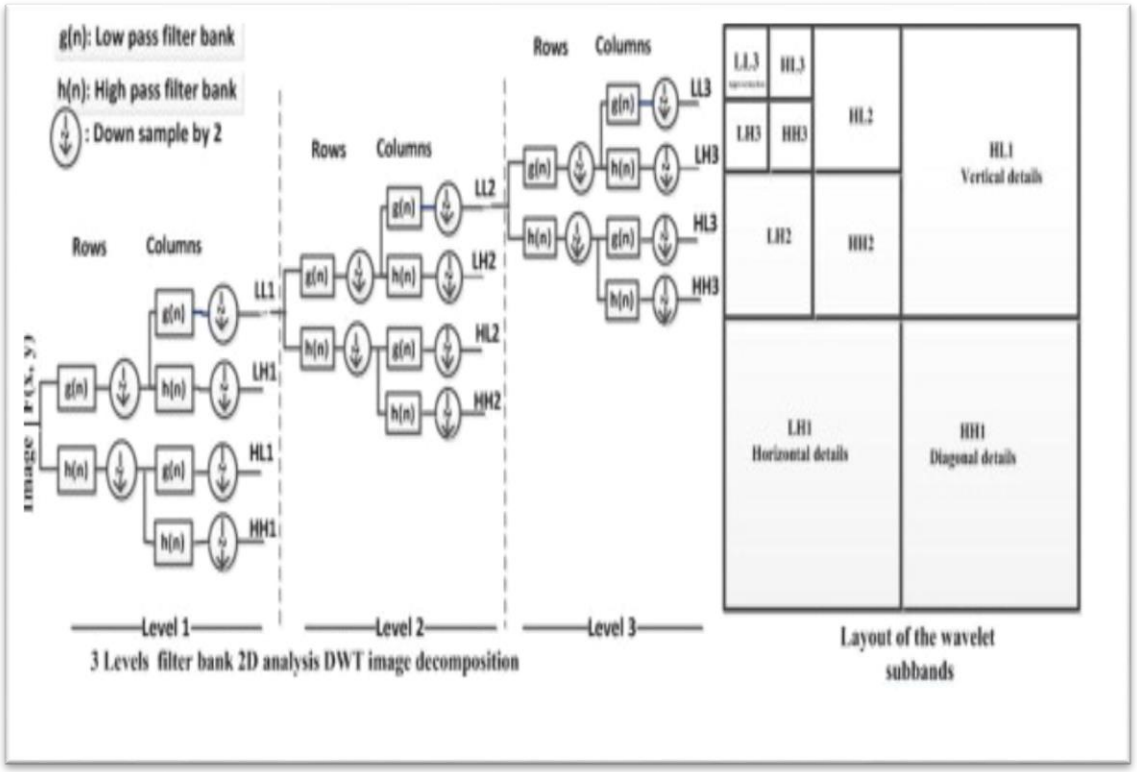

Figure 5: Schematic diagram of 3-level 2D DWT

3.6.4 Dimensionality Reduction using PCA

Excessive features increase the computation time and memory storage and make classification more complicated. This is called the curse of dimensionality. It is required to reduce the number of features for the classification process. The principal components analysis (PCA) is a well-known technique of identifying the most meaningful basis to re-express a data set. In other words, the goal of PCA analysis is to transform the existing input features of a data set consisting of a large number of interrelated variables into a new lower-dimension feature space while retaining most of the variation. The input feature space is transformed into a lower feature space using the largest eigenvectors and forms a

new set of ordered variables according to their variances or the importance. This technique has three effects: first, it orthogonalizes the components of the input vectors so that uncorrelated with each other, then it orders the resulting orthogonal components so that those with the largest variations come first, and finally it eliminates those components which contributes least to the variation in the data set. The purpose of PCA is to reduce the dimensionality of the wavelet coefficients which results in a more efficient and accurate classifier.

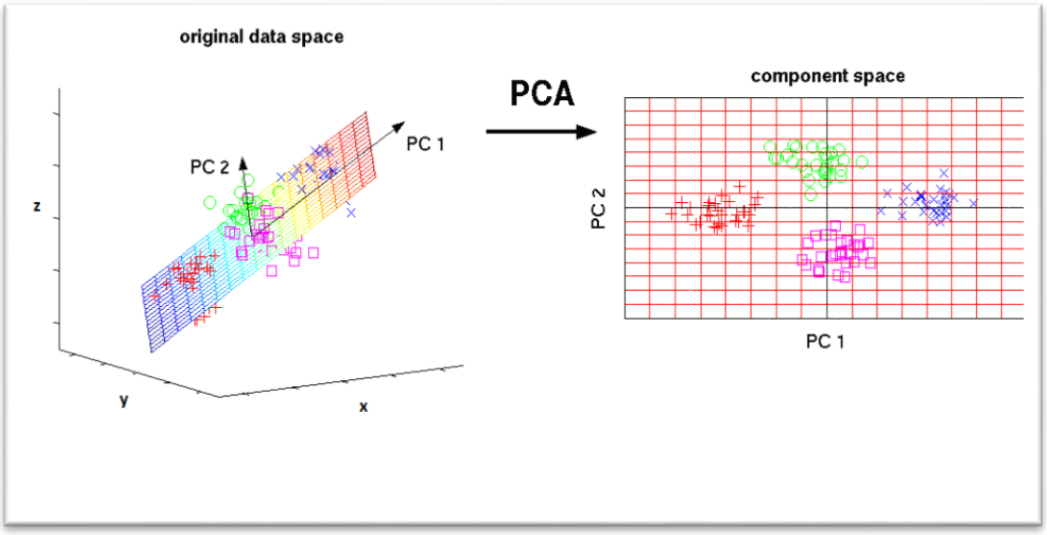

Figure 6: Example of PCA transformation of 3-dimensional data to 2-dimensional data

3.7 Brain Regions Classification based on Texture Analysis using Artificial Intelligence (AI)

The main goal of texture analysis is the classification of different tissues of the brain in several regions. Following the feature selection step there are many algorithms that aim in highest discrimination by using a combination of the extracted features. As in the present thesis we use ANN and ANFIS, we briefly describe these two classifiers for MRI texture analysis applications, *artificial neural networks (ANN) and adaptive neuro fuzzy inference system (ANFIS).*

3.7.1 Artificial Neural Networks

Artificial neural networks are networks that consist of a number of artificial neurons which are connected and act as a computational machine.

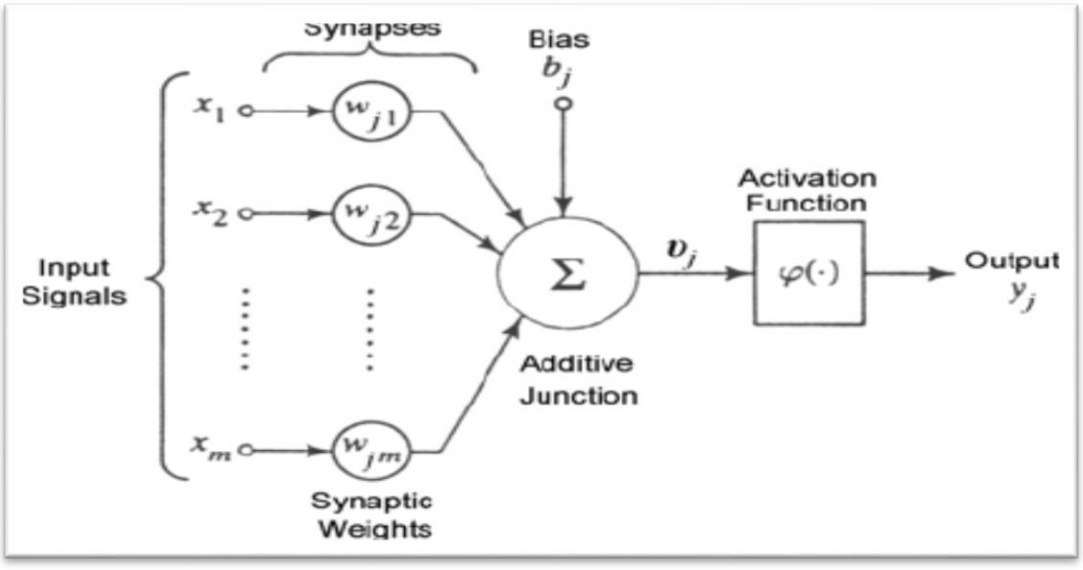

Figure 7: Model of artificial neuron

An artificial neuron as in figure above receive some input signals x_i. Every x_i of the neuron is mapped with a weight w_j which amplifies or reduces the input signal. The modified by the weight inputs become inputs of the neuron where they are been added by a sum unit to produce the net input u_j, which stimulate an activation function for the production of the final output of the neuron, *y = f(net)*, where

$$net = x_1w_1 + x_2w_2 + \ldots + x_nw_n + b = \sum_{1}^{n} x_jw_j + b \qquad \text{(Eq.18)}$$

The factor b is called *bias* and acts just like a weight factor w_j and its activation function is always 1. *Bias* is used for signal amplification and improved convergence.

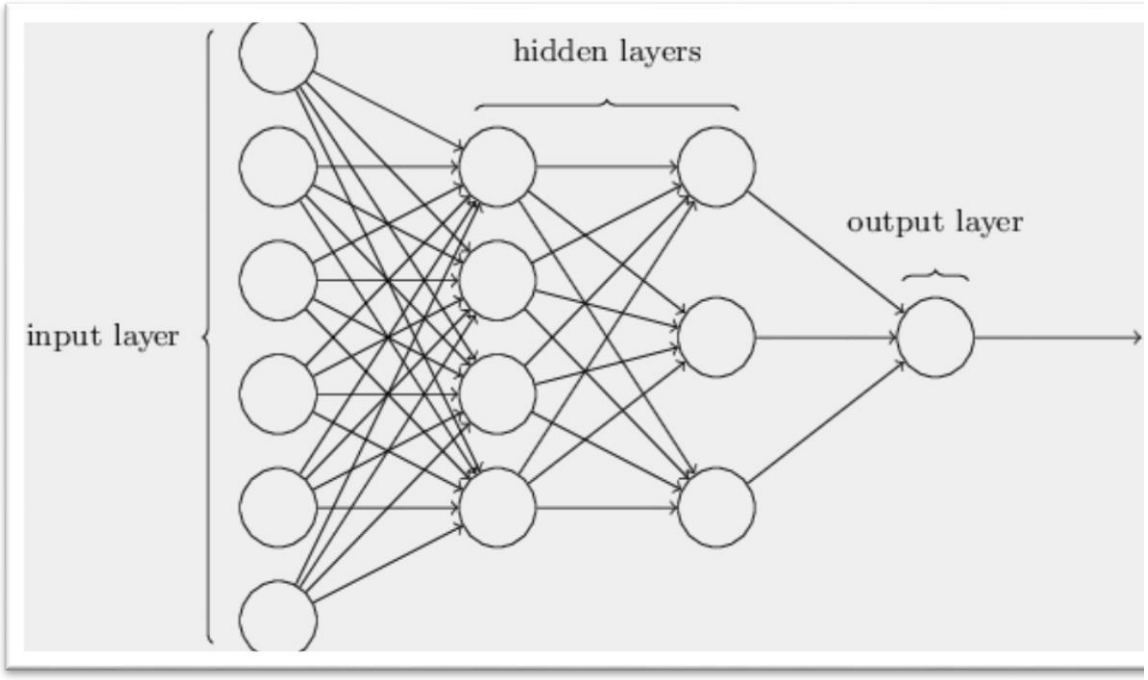

Figure 8: Model of neural network

The structure of artificial neural networks is made up by layers. Input layer has inputs that are signals from the outside environment. The outputs of this layer become inputs for hidden layer and after this stage, at the end there is the output layer where the final output is shaped. The knowledge that has been acquired from the network, is translated into the values of the weights. This configuration of weights is done during the training phase. In this stage, the output of the neural network is compared with the desired output and an error is computed. The error is fed back as input to the artificial neural network which adjust the parameters for minimizing this error function. This procedure is called training phase and it ends when an acceptable error value is calculated or when the network cannot resolve the problem.

3.7.2 Adaptive Neuro – Fuzzy Inference System (ANFIS)

ANFIS network [5] is a fuzzy Takagi Sugeno model and consists of five layers. For describing ANFIS architecture we consider two fuzzy *IF/THEN* rules of a first order Sugeno system:

$$\text{Rule 1: } IF\ (x \text{ is } A_1)\ AND\ (y \text{ is } B_1)\ THEN\ (f_1 = p_1 x + q_1 y + r_1)$$
$$\text{Rule 2: } IF\ (x \text{ is } A_2)\ AND\ (y \text{ is } B_2)\ THEN\ (f_2 = p_2 x + q_2 y + r_2) \quad \text{(Eq.19)}$$

where x and y are inputs, A_i and B_i are the fuzzy sets and f_i are the outputs. The outputs f_i are defined by the inputs and the parameters of the system, p_i, q_i,και r_i which are calculated in the learning phase.

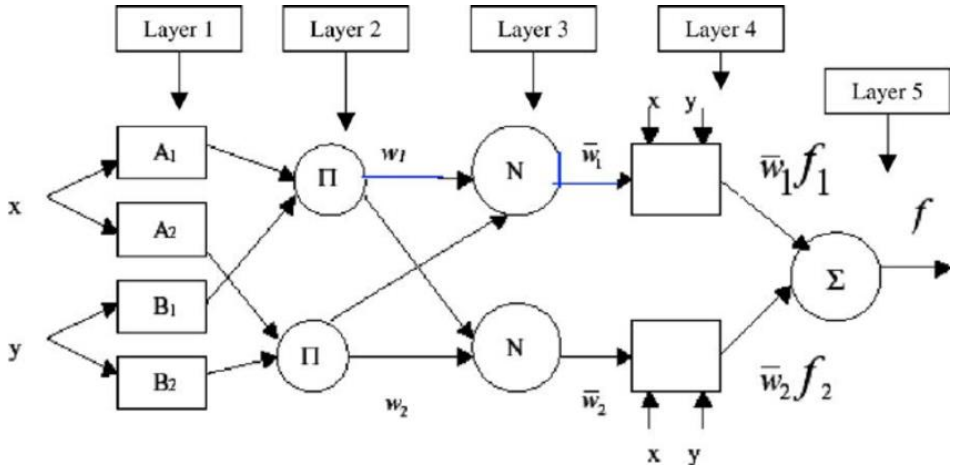

Figure 9: ANFIS Architecture

- 1st layer:

All the nodes in this layer are adaptive nodes. The outputs of this layer are the fuzzy membership grade of each input in a fuzzy set, which are given by:

$$O_1^1 = \mu_{A_i}(x), \text{ for } i = 1, 2$$
$$O_2^1 = \mu_{B_{i-2}}(x), \text{ for } i = 3, 4 \qquad \text{(Eq.20)}$$

where $\mu_{A_i}(x)$ can adopt any membership function. For example, if the Bell shaped membership function is employed $\mu_{A_i}(x)$ is given by:

$$\boldsymbol{\mu_A(x) = \frac{1}{1 + \left|\frac{x - c_i}{a_i}\right|^{2b}}} \qquad \text{(Eq.21)}$$

where a_i, b_i and c_i are the parameters of the Bell shaped membership function. Parameters in this layer will be referred to as premise parameters.

- 2nd layer:

  In this layer the nodes are fixed. The outputs give the product of the inputs and represents the firing strengths of a rule. The outputs are given by:

$$O_i^2 = w_i = \mu_{A_i}(x)\ \mu_{B_i}(x), \text{for } i = 1, 2 \qquad \text{(Eq.22)}$$

- 3rd layer:

  In this layer the nodes are also fixed. The outputs of this layer play a normalization role to the firing strength of the previous layer. The normalized firing strengths are represented by:

$$O_i^3 = \overline{w_i} = \frac{w_i}{w_1 + w_2}, \text{for } i = 1, 2 \qquad \text{(Eq.23)}$$

- 4th layer:

  In this layer the nodes are adaptive. Each node of this layer gives an output that can be described by:

$$O_i^4 = \overline{w_i} f_i = w_l(p_i x + q_i y + r_i) \qquad \text{(Eq.24)}$$

which calculates the product of the normalized firing strengths and a first order polynomial (for a first order Sugeno model). Parameters in this layer will be referred to as consequent parameters.

- 5th layer:

  In this layer there is only one fixed node which gives the overall output of the system:

$$y = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad \text{(Eq.25)}$$

After discussing the architecture of ANFIS we now describe the learning algorithm of the system. The main goal of the learning phase of ANFIS is to tune all the modifiable parameters, namely premise parameters $\{a_i, b_i, c_i\}$ and consequent parameters $\{p_i, q_i, r_i\}$ to fit the training data. When the premise parameters $\{a_i, b_i, c_i\}$ are fixed, then the output of ANFIS model is calculated by Eq.25,

$$f = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \quad \text{(Eq.26)}$$

where according to Eq.23 we can write the above relation as,

$$f = \overline{w_1} f_1 + \overline{w_2} f_2 \quad \text{(Eq.27)}$$

Substituting the fuzzy *if-then* rules into Eq.27, it becomes,

$$f = \overline{w_1}(p_1 x + q_1 y + r_1) + \overline{w_2}(p_2 x + q_2 y + r_2) \quad \text{(Eq.28)}$$

$$\Leftrightarrow \quad f = (\overline{w_1} x)p_1 + (\overline{w_1} y)q_1 + (\overline{w_1})r_1 + (\overline{w_2} x)p_2 + (\overline{w_2} y)q_2 + (\overline{w_2})r_2 \quad \text{(Eq.29)}$$

which is a linear combination of consequence parameters. When the premise parameters are not fixed, then the search space becomes larger and the convergence of the space slower. For solving this problem a *hybrid algorithm* which combines the *least square method and the gradient descent* is adopted. This hybrid algorithm is composed of a *forward and a backward pass*. In the forward pass, the least squares method is used to optimize the consequence parameters, with the premise parameters fixed. When the optimal parameters are found, the backward pass starts. Now, the *gradient descent method* is used to adjust optimally the premise parameters, based on an output error and the forward pass starts again. It has been proven that this hybrid algorithm is very efficient in training the ANFIS [9].

Parameters	Forward pass	Backward pass
Premise parameters $\{a_i, b_i, c_i\}$	fixed	Gradient descent
Consequent parameters $\{p_i, q_i, r_i\}$	Least squares	fixed
Signals	Node outputs	Error signals

Table 1: Hybrid algorithm procedure which combines the least square method and the gradient descent

3.8 Literature Review

In recent years, researchers have proposed several approaches for automated and accurate classification methods in MRI images. This section will present some approaches that have been developed and used in brain MRI images.

R. C. Patil et al, 2012, [10], present a brain tumor extraction method based on two stages. Image enhancement is performed using high-pass and median filter in the image and then a segmentation algorithm based on watershed segmentation and morphological operations is used for the extraction of the tumorous area.

E. El-Dahsahan, T. Hosny and A. Salem, 2014, [11] proposed a hybrid technique which uses a 3-level DWT to extract the coefficients and a PCA algorithm for

dimensionality reduction. Then two classifiers, feed forward backpropagation artificial neural network (FP-ANN) and K-nearest (KNN), have been developed.

Y. Zhang and L. Wu, 2012, [12] used also DWT transform for the extraction of the details and the approximations and PCA algorithm for the selection. This approach proposed a kernel support vector machine (KSVM) for classification and a K-fold cross validation algorithm for generalization of KSVM. For this method four kernels were used, linear kernel (LIN), homogeneous polynomial (HPOL), inhomogeneous polynomial (IPOL) and Gaussian radial basis (GRB) which provides the most accurate classification results.

N. Bahadure et al, 2017, [13] proposed an approach based on Berkeley wavelet transform (BWT) and SVM classification. For the classification, the study uses the features that have been extracted by the Grey Level Co-Occurrence matrix (GLCM) after a skull stripping process.

X. Zhou et al, 2016, [14] presents a computer-assisted diagnosis method based on wavelet entropy (WE) of the feature space approach and a feed-forward neural network (FNN) classification method. The decomposition of the image is made by the use of DWT and for every coefficient matrix that has been created an entropy feature is computed. This study uses a cross validation algorithm for generalization and claims that with two features provides a 100% accuracy.

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
PCNN+DWT+PCA+BPNN [11]	100	92.3	99
DWT+PCA+k-NN [11]	96	97	98
DWT+PCA+KSVM (LIN) [12]	96.4	85	95
DWT+PCA+KSVM (HPOL) [12]	97.1	95	96.88
DWT+PCA+KSVM (IPOL) [12]	99.2	90	98.12
DWT+PCA+KSVM (GRB) [12]	99.2	100	99.38
BWT+GLCM+ANFIS [13]	96.72	79.74	90.04
BWT+GLCM+ANN [13]	97.5	76.54	85.57
BWT+GLCM+SVM [13]	97.72	94.2	96.51
BWT+GLCM+k-NN [13]	93.33	77.77	87.06

Table 2: Classification results of related studies. [11] and [12] results are referred to a combination of training and testing dataset, while in [13] the results are referred to testing dataset only

CHAPTER 4. PROPOSED METHODS, MATERIALS AND IMPLEMENTATION WITH MATLAB

4.1 An Overview of the proposed Methods

In this chapter, we present and describe the finally four new consistent methodologies for CAD process that materialized after careful examination, optimization and combination of the existing tools, in order to compare them with the existing in the literature CAD process for MRI classification and detection of brain tumors. Our proposed four methods are based on the stages of:

1. *preprocessing the images,* including skull stripping of the original MRI and filtering
2. *post-processing* for detecting the region of interest (ROI)
3. *feature extraction, feature selection and classification*

In every stage, several scenarios and different techniques are implemented for better results. The flow stages of the proposed methods are presented in figure 10.

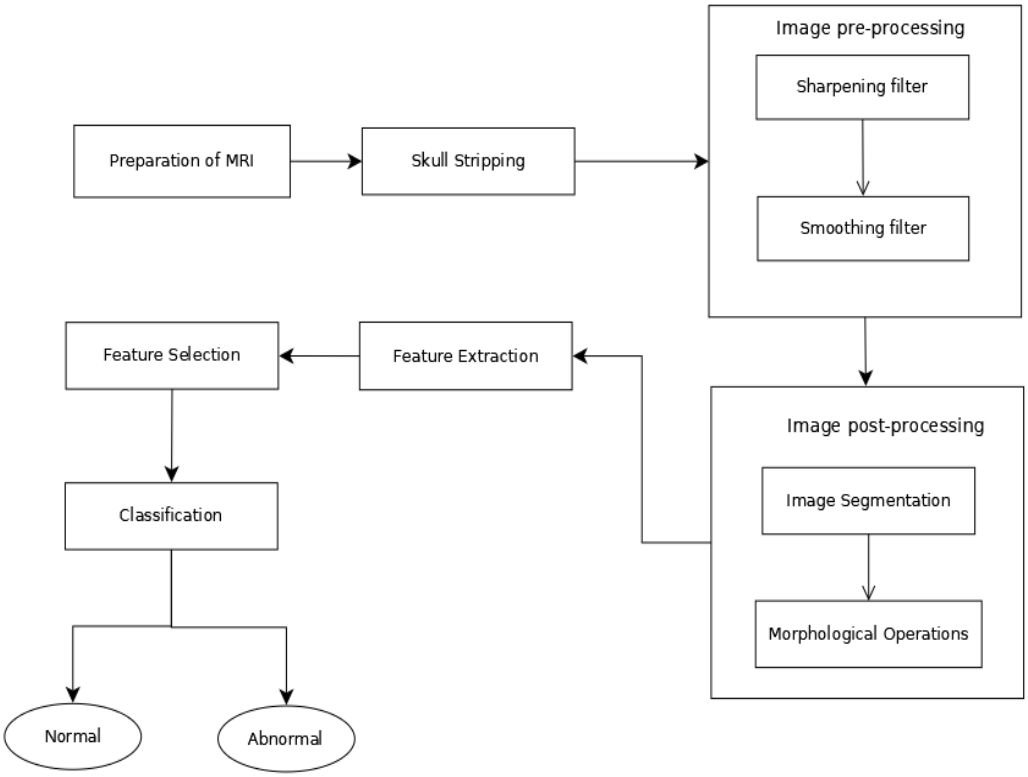

Figure 10: Thesis Flowchart of CAD process proposed Methods

In all four methods the preprocessing part maintain the same. In other words, in every method the images that have to be post-processed are prepared properly and are skull stripped so that they do not include any unnecessary information. Our proposed methods differ in image segmentation and feature extraction techniques. The methods that are implemented in this study can be concluded in Table 3 below.

Thesis Proposed CAD Process	Flow Stages of the proposed method
Method 1	Preprocessing ⟶ Otsu thresholding ⟶ Morphological Operators ⟶ GLCM ⟶ ANN / ANFIS
Method 2	Preprocessing ⟶ Otsu thresholding ⟶ Morphological Operators ⟶ DWT ⟶ PCA ⟶ ANN / ANFIS
Method 3	Preprocessing ⟶ Otsu thresholding ⟶ Morphological Operators ⟶ DWT ⟶ PCA ⟶ GLCM ⟶ ANN / ANFIS
Method 4	Preprocessing ⟶ Mean-Shift segmentation ⟶ Otsu thresholding ⟶ Morphological Operators ⟶ DWT ⟶ PCA ⟶ GLCM ⟶ ANN / ANFIS

Table 3: Thesis four Proposed CAD Process Methods

4.2 Thesis Database Selection

In this thesis all the MRI images that are used are referred to T2 MRI images of the axial plane. As mentioned before, the acquisition of a proper dataset is a challenging task. We managed to gather data from

1. “MICCAI BraTS 2015” database,

2. “The Whole Brain Atlas” from Harvard database
3. General Hospital “St. George” of Chania.

The database of MRI brain images that are used is referred to male and female adult persons and it consists of 24 non tumorous cases and 202 cases that was diagnosed with brain tumor. Because of their unpredictable appearance and shape, segmenting brain tumors from multi-modal imaging data is one of the most challenging tasks in medical image analysis. Although many different segmentation strategies have been proposed in the literature, it is hard to compare existing methods because the validation datasets that are used differ widely in terms of input data (structural MR contrasts; perfusion or diffusion data; ...), the type of lesion (primary or secondary tumors; solid or infiltrative growing), and the state of the disease (pre- or post-treatment).

4.3 Thesis Database Implementation: Pre-processing stage

4.3.1 Preparation of data used

Because of the different sources of the collected data it was necessary to become a certain preparation of them. The data from the BraTS database are in “*mha*” file format, which is referred to three dimensional images, while images from Harvard database and from the hospital are in “*gif*” file format. All our data are modified in two dimensional images in “*jpeg*” file format. The new images consist of 240*240 pixels in grayscale type with 8-bit grey level rate.

4.3.2 Skull Stripping algorithm used

Skull stripping is an important process in magnetic resonance imaging analysis and it is required for the effective examination of brain tumour. With this process all non-brain tissues like skull, fat and skin can be separated from the brain region. There are several techniques available for skull stripping using image contours, *thresholding* and *histogram analysis.*

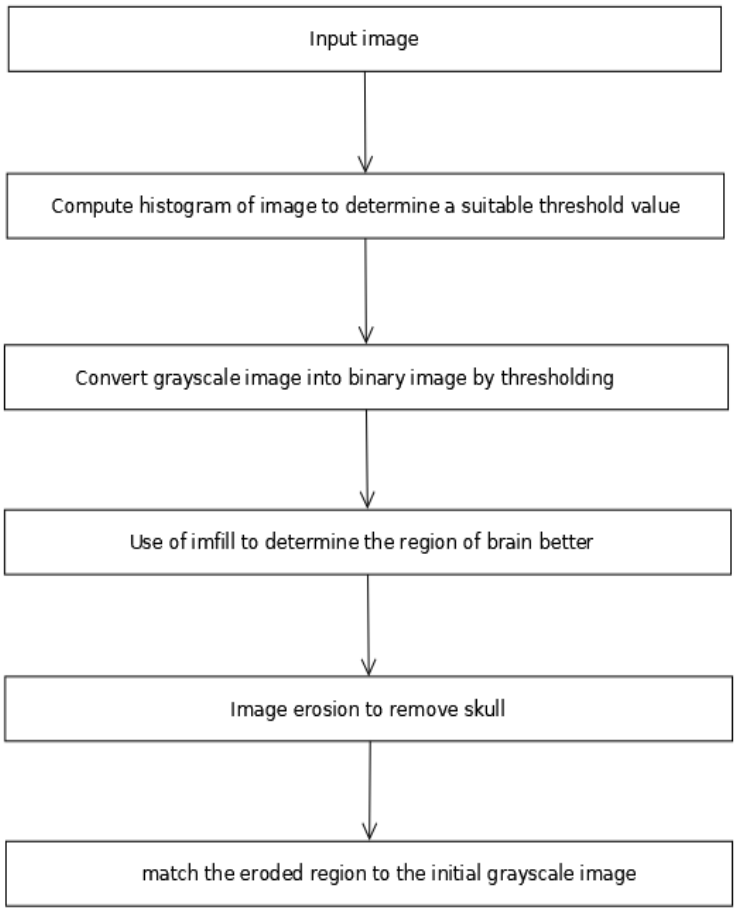

Figure 11: Thesis steps used for skull stripping algorithm based on thresholding segmentation and morphological operations

In this thesis, skull stripping process is implemented by an algorithm based on *thresholding segmentation and morphological operations*. The steps of this algorithm are presented in Figure 11 above as follows.

1. After the image is read, the histogram of the input image is computed to determine a proper threshold value for segmentation. This threshold value is applied on the image and is binarized.
2. In addition, *imfill* function is used to fill the holes of brain region that were developed and describe the brain region better.
3. After the image is eroded by a disk structuring element and the region of skull was removed. Then, the binarized image is matched with the initial grayscale image and the region of interest is finally determined.

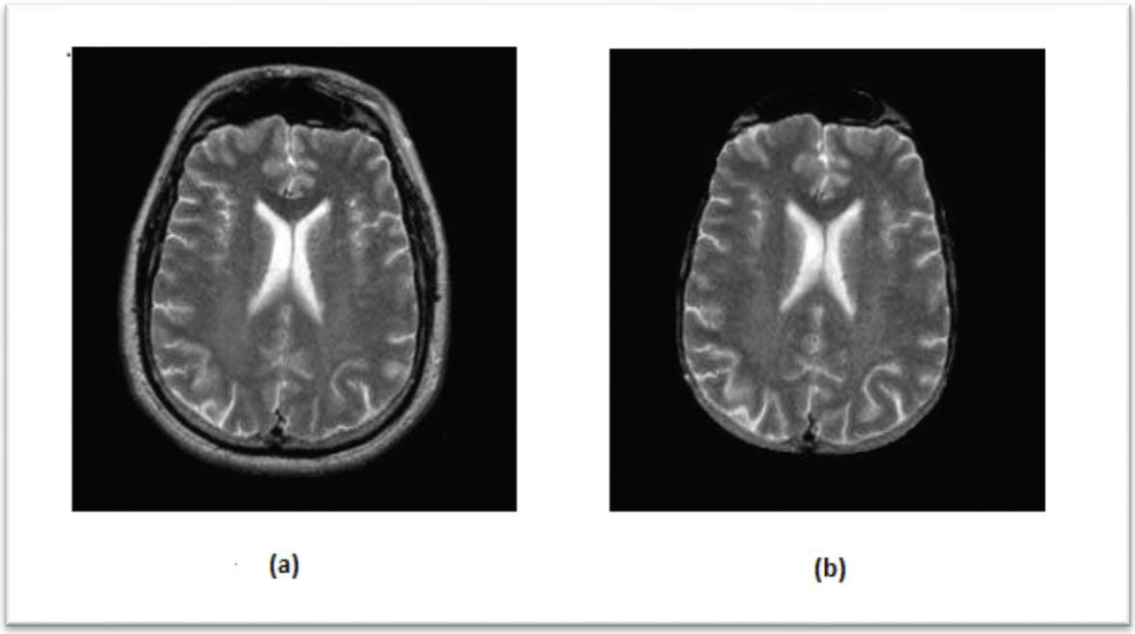

Figure 12: .Thesis results. Example of the application of skull stripping algorithm used. (a) initial MRI image, (b) MRI image after the application of skull stripping algorithm

4.3.3 Sharpening using high pass filter

After having extracting the skull pixels of the image, a high pass Gaussian filter is applied in order to sharpen the edges of the original MRI image. An image is sharpened when contrast is enhanced between neighbourhood regions with variation in darkness or brightness. For applying a high pass Gaussian filter first we computed the *low pass filter* with kernel function,

$$g(x, y) = \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad \text{(Eq.30)}$$

Then, by subtracting the kernel function we compute the *high pass filter*. In figure below we show the images of the application of the high pass Gaussian filter.

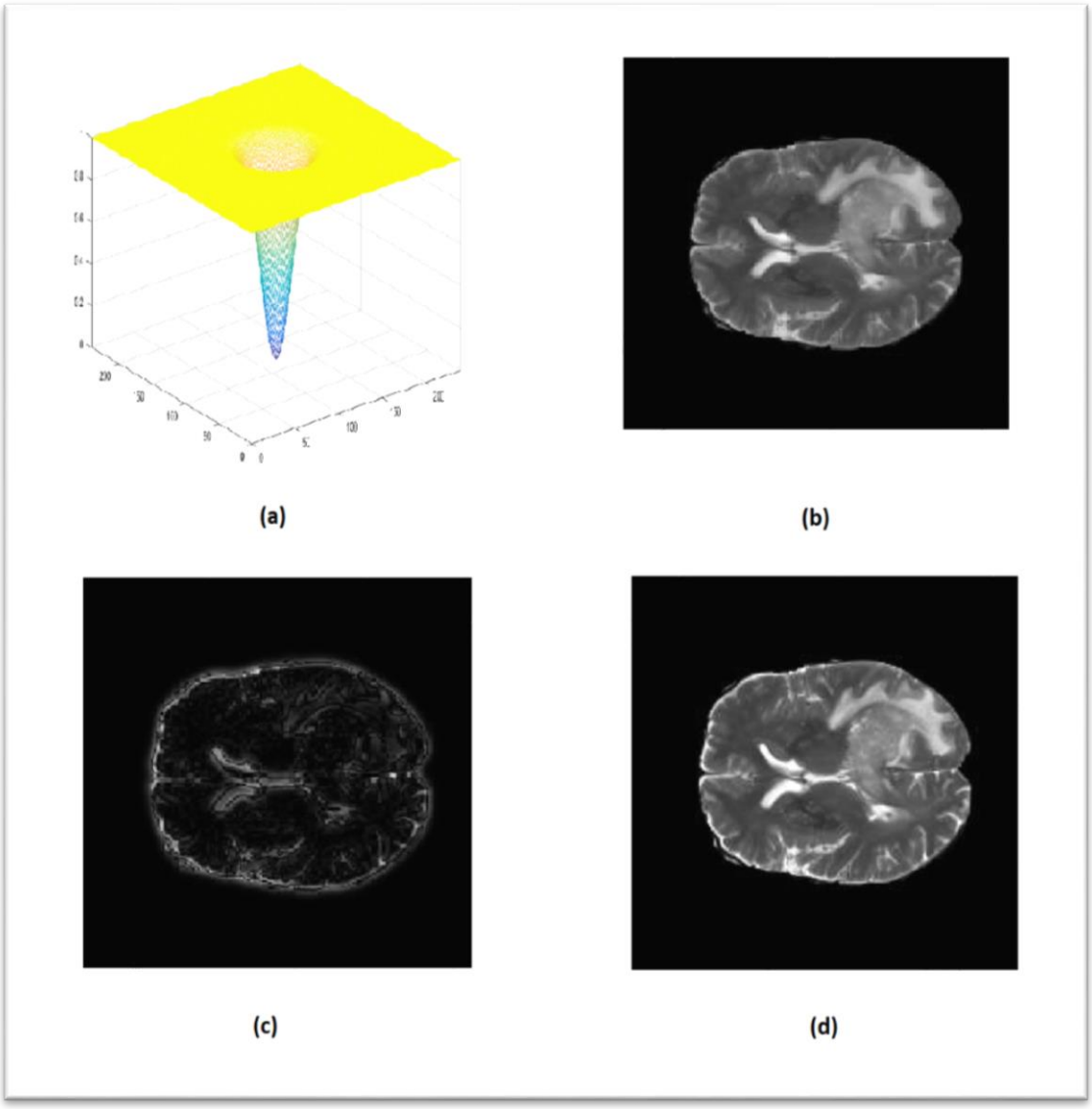

Figure 13: Thesis results. Example of the application of sharpening filter. (a) high-pass filter, (b) original MRI image, (c) enhanced details of original image, (d) sharpened MRI image

4.3.4 Reducing noise using Median filter

The high pass Gaussian filter sharpened all final details of image including the noise. For reducing the noise a *median filter* has been used. Median filter is a non-linear method used

to remove noise while preserving the edges. For our implementation a 3×3 window around the corresponding pixel is used.

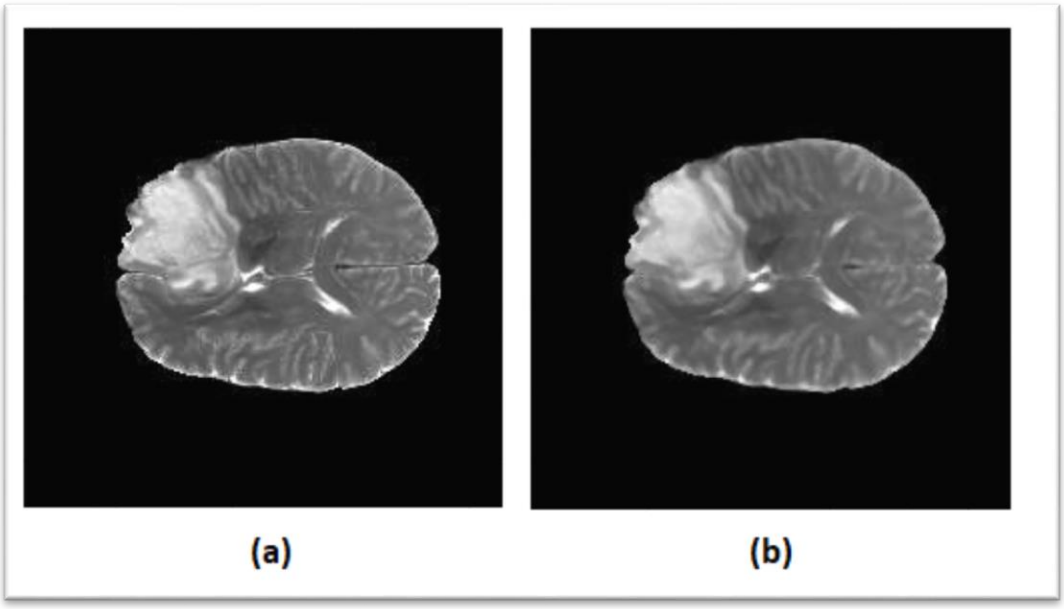

Figure 14: Thesis results. Example of the application of median filter. (a) image after high pass filter, (b) image after median filter

4.4 Thesis Database Implementation: Post-processing stage

The post-processing stage in this thesis aims:

i) to detect the region of tumour from the images

ii) to extract features for textural analysis of the images

iii) to reduce the dimensionality of the extracted features

Several techniques and different scenarios implemented for better results. In the following paragraphs we describe the post processing stage of our proposed methods.

4.4.1 Thesis proposed CAD Process Method 1: Otsu's Thresholding, Morphological Operators, GLCM

In this method, after the filtering of the images we implement global thresholding segmentation. Global thresholding segmentation is used in order to achieve a binirized image with gray level one (1), representing the tumor in the tumorous images and the brain ventricles in the non-tumorous images, and gray level zero (0) to represent the background. The binarization is based in a threshold value *T* that is calculated with *Otsu's thresholding method*. Otsu's method assumes that the image contains two classes of pixels, foreground and background, following bi-modal histogram and then chooses the threshold to minimize the infraclass variance of the black and white pixels [24]. The binarized image can be described by Eq.1.

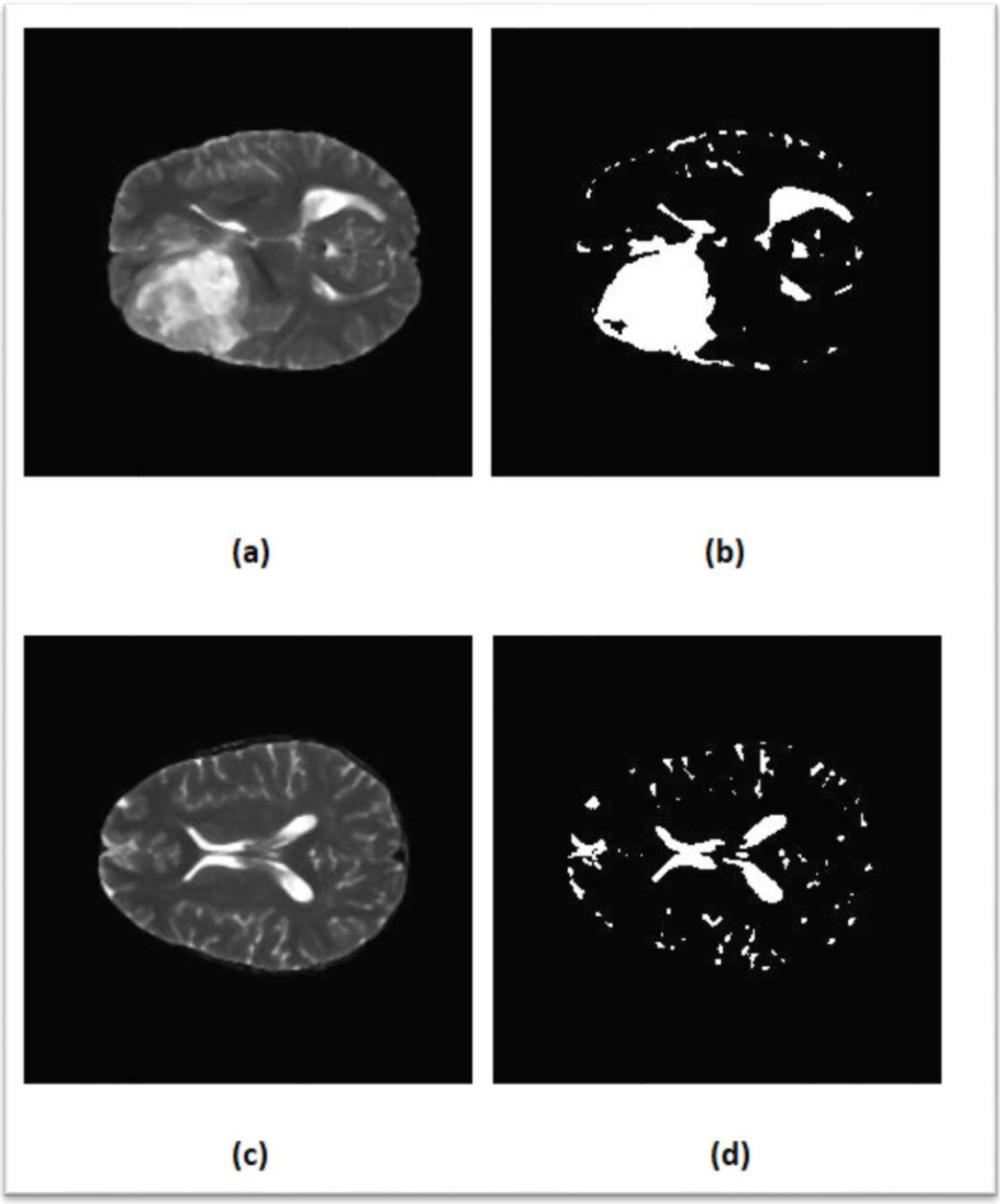

Figure 15: Thesis results. Examples of MRI images segmented by Otsu's Thresholding. (a) Median filtered image with tumor, (b) segmented image with tumor, (c) Median filtered image without tumor, (d) segmented image without tumor

After the Otsu's Thresholding segmentation of the images, morphological operations are used for the purpose of isolating the tumor in an image and removing small objects that are not included in the region of interest. Erosion and opening by reconstruction are applied. In our method 1, for the erosion operation a disk structuring element with radius equal to 4 pixels is chosen and the erosion is implemented by Matlab® functions shown by these expressions:

```
se = strel('disk', 4);
Ie = imerode(I, se);
```

where *se* is the structuring element, *Ie* is the eroded image and *I* is the original image.

For implementing the opening by reconstruction operation the Matlab® function below is used:

```
Iobr = imreconstruct(Ie, I);
```

where *Iobr* is the reconstructed image, *Ie* is the marker and *I* is the mask. The results of these operations are shown below in Figure 16.

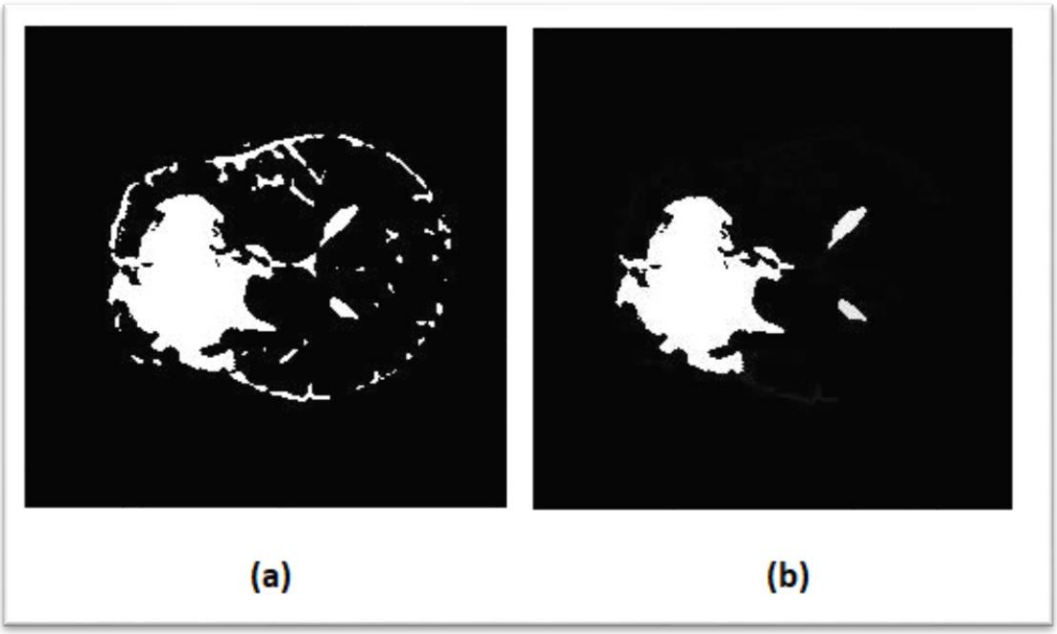

Figure 16: Thesis results. Examples of application of erosion and opening by reconstruction morphological operations. (a) is a segmented MRI image, (b) is the MRI image after morphological operations

In this method the feature extraction stage is implemented in two steps.

i) First the GLCM matrix is calculated for the horizontal orientation and then,

ii) the statistical features that describe the texture of the images are calculated.

The 13 statistical features as mentioned in the previous chapter are *mean, variance, standard deviation, entropy, skewness, kurtosis, energy, contrast, homogeneity, correlation, smoothness and rms*. In **Table 4** a small set of features is presented.

“,” is used for decimal values						
No.	**Contrast**	**Correlation**	**Energy**	**Kurtosis**	**Mean**	**Diagnosis**
1	0,354532	0,146769	0,896522	42,28021	0,003009	1
2	0,334064	0,180214	0,868543	24,76191	0,004019	1
3	0,364035	0,181498	0,879327	31,69369	0,00586	1
10	0,288377	0,237865	0,880061	28,58626	0,003297	0
11	0,32383	0,155316	0,853933	24,97971	0,003976	0
12	0,328582	0,127542	0,842785	23,18412	0,004225	0

Table 4: Thesis results for case Method 1. Texture features of the used database images

In our proposed CAD process Method 1, a matrix of 226 rows and 15 columns is been created. The rows of that matrix represent the patients while the first column describes the number of the patient, the next 13 columns describe the textural features and the last column describes the diagnosis of an expert for every patient. This matrix is going to be the input for our classification method.

The proposed Method 1, is also implemented without pre-processing of the original images and the results will be discussed in the next chapter.

4.4.2 Thesis proposed CAD Process Method 2: Otsu's thresholding, Morphological Operators, DWT, PCA (3-level 2D decomposition via Daubechies 4 and SVD decomposition)

Our proposed Method 2 for CAD process, the stages of preprocessing, image segmentation and morphological operations have been maintained the same as that of our Thesis Method 1. The difference of this method is shown in the feature extraction process where DWT is used for transforming the image into a new domain. A 3-level 2D decomposition via Daubechies 4 wavelet is utilized to extract features. After the decomposition with approximation and detail operators, describing by Eq.16 and Eq.17, an image can be characterized by its wavelet coefficients. Despite this, feature reduction is made as the number of extracted features that describing an image are now reduced from 57600 to 1296, which resulting from 36×36, the size of approximation coefficient of the 3-level decomposition. However this number should be further reduced by the application of a PCA algorithm. As mentioned in the previous chapter, excessive features increase the computation time and memory storage. In our Thesis Method 2, PCA algorithm is implemented using SVD (singular-value decomposition). The curve of cumulative sum of variance versus the number of principal components is shown in Figure17. It shows that the first 5 principal components of 1296 in total, which is 0.39% of the original features, could preserve 96% of total variance.

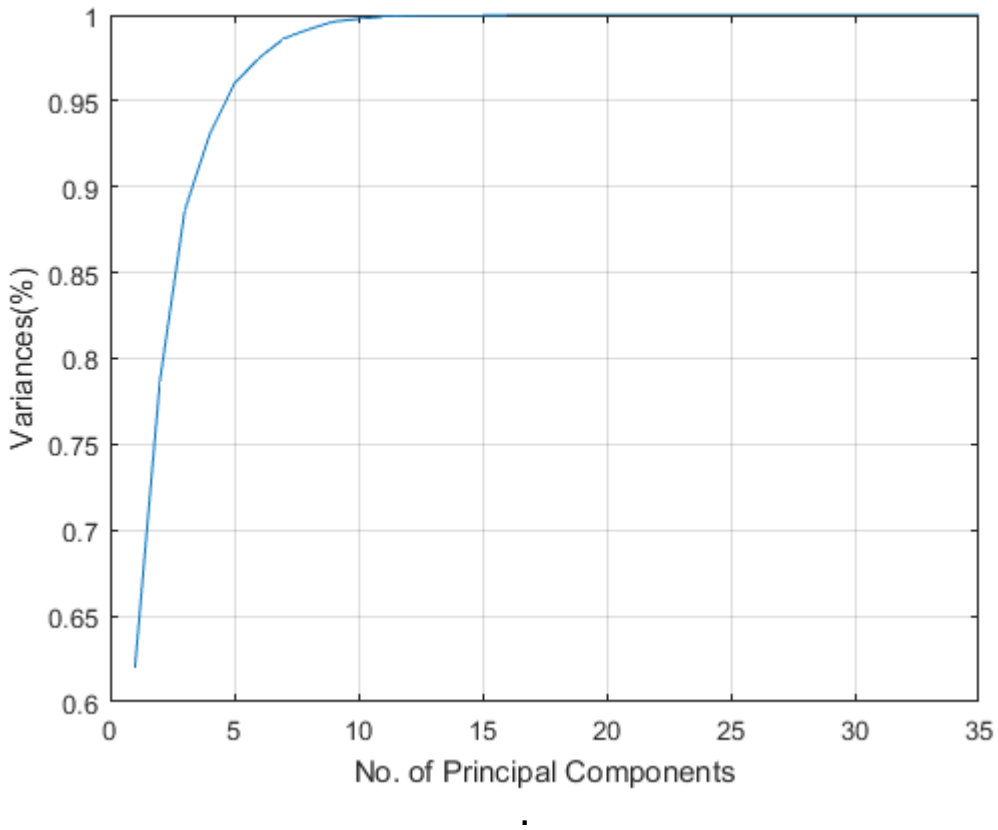

Figure17: Thesis result of Method 2. Variances against No. of principle components

Thesis Method 2 is also implemented without pre-processing of the original images and the results will be discussed in the next chapter.

4.4.3 Thesis proposed CAD Process Method 3: Otsu's thresholding, Morphological Operators, DWT, PCA, GLCM

Method 3 of our CAD process proposed methods combines Method 1 and Method 2. As in Methods 1 and 2, after the pre-processing stage, global thresholding with Otsu's threshold is implemented and the morphological operators of erosion and opening by reconstruction are used. With this process the ROI of every image is isolated. Then, discrete

wavelet transform is implemented for feature extraction and reduction. As in Method 2, the approximation coefficients are given by the top left sub-band LL of the 3-level decomposition. The Daubechies 4 and Haar wavelets are tested and the Daubechies 4 wavelet is chosen. After this step the PCA algorithm is applied for further feature reduction, and as in Method 2, the first 5 principal components could preserve 96% of the information. Based on these features, for each image the 13 statistical values of GLCM matrix are calculated. As a result a matrix is formed with each row representing an image. The first column of this matrix is the number of the image, the next 13 column represents the statistical values of every image and the last column contains the diagnosis for each case. This matrix is used as input in our classification algorithm.

4.4.4 Thesis proposed CAD Process Method 4: Mean-Shift segmentation, Otsu thresholding, Morphological Operators, DWT, PCA, GLCM

The proposed Method 4 is based in the concept of the previous one following the same procedures for feature extraction and dimensionality reduction, as well as the image pre-processing. The main difference of Method 4 is observed in the segmentation stage, where the algorithm of Mean-Shift is implemented for better clustering results. Although the main idea is not to segment the image in several different clusters like [15], we apply Mean-Shift algorithm for smoothing the image, further reduce the noise and improve contrast between the segmented regions with different intensity values. The implementation of the Mean-Shift algorithm is made based on a flat kernel function for estimating the p.d.f and a bandwidth window equal to 0.1. In Figure 18, the effect of Mean-Shift algorithm in the pre-processing image is presented.

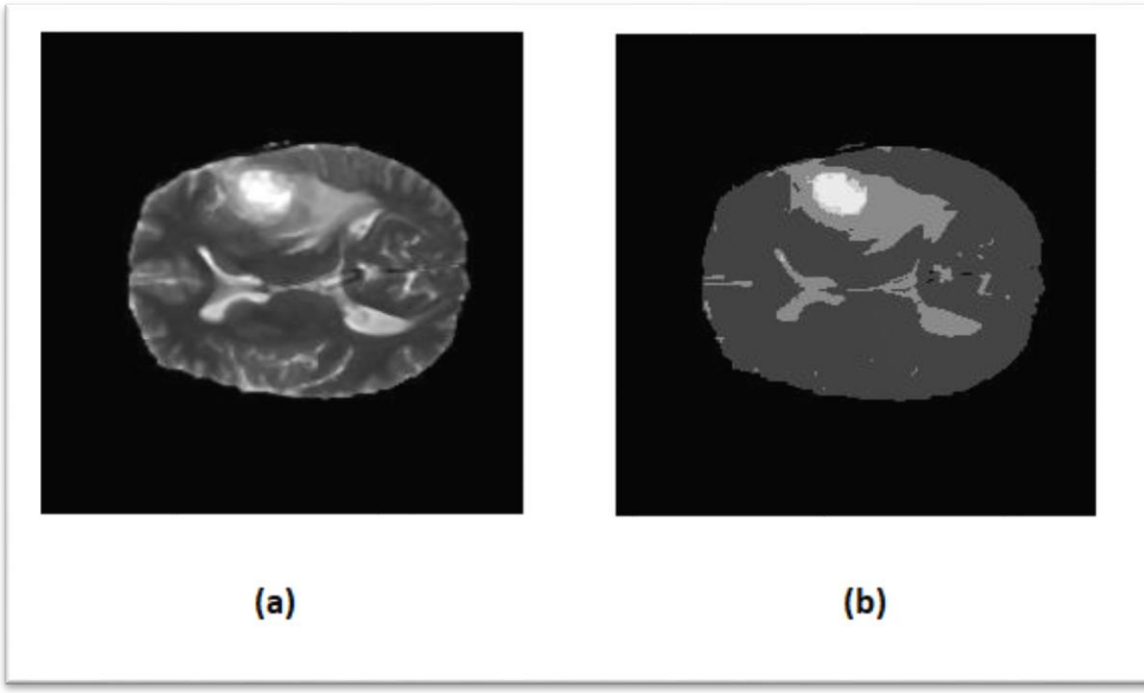

Figure 18: Thesis results of Method 4. Example of application of Mean-Shift algorithm on the pre-processing image, (a) pre-processed image, (b) image after Mean-Shift algorithm

After the Mean-Shift application we segment the image with global thresholding via Otsu's threshold as we do in previous Methods. Then just like Method 3, morphological operators are implemented and feature extraction and dimensionality reduction is applied via DWT and PCA algorithms. The last step of Method 4 is the feature extraction with the GLCM matrix.

CHAPTER 5. EXPERIMENTAL CLASSIFICATION RESULTS USING ANN AND ANFIS

5.1 Classification

5.1.1 Artificial Neural Network

The features extracted from each case of our proposed Methods are classified separately using *feed-forward backpropagation neural network*. The *patternnet* function of Matlab® is used for creating the network. The *patternnet* function is called with input the hidden layers. Below we present an example with two hidden layers with the first layer including ten neurons and the second layer including five.

```
Hiddenlayers = [10 5]
net = patternnet(hiddenlayers);
```

After the creation of the network the train process follows.

For training the network we use the *train* function of Matlab® setting our training parameters. We set the maximum number of iterations, the error tolerance and the learning rate. For training algorithm the *Scaled Conjugate Gradient* algorithm in some cases and the *Levenberg-Marquardt algorithm in others* are used, and for the performance function we use the *Mean Squared Error*. Below we present an example of setting the specified parameters of the neural network.

```
net.trainParam.lr = 0.05;
net.trainParam.epochs = 100;
net.trainParam.goal = 1e-6;
net.trainFcn = 'trainscg';
net.performFcn='mse';

net1 = train(net, input, target);
```

In a *feed-forward backpropagation neural network* the *weights* and the *biases* contain the information of how "significant" is every input, while it travels through the network. The *weights* and *biases* of the network are updated automatically by the training function.

5.1.2 ANFIS

For the features extracted from Methods 1, 3 and 4, as well as for the scenarios without pre-processing the images, we implemented classification using ANFIS classifier for improving our results. In all cases the inference system is generated with *grid partition* and 3 layers with 5 membership functions on each layer. The membership function that is chosen is the *Gaussian Combination membership function* for the inputs and the *Linear membership function* for the outputs, to better describe the model. The error tolerance is set equal to 10^{-4} and the number of epochs is set to 100. For all our ANFIS models the number of rules is 125.

In the next section we present the results of our proposed Methods in detail. Because of the limited number of our healthy cases we implement our Methods not only using the whole Thesis dataset, but also using a subset of the database which contains less number of non-healthy cases. In every proposed Method we present at first the classification results using the features extracted from the smaller dataset and then the results using the features extracted from the whole dataset.

5.2 Classification results of Method 1

5.2.1 Results of ANN

First we test the network using a subset of the thesis dataset containing only some of the patients with ratio 30% healthy cases and 70% non-healthy cases. This network contains two hidden layers with 8 neurons each and has 4 inputs and one output. The inputs are the features of energy, homogeneity, skewness and kurtosis.

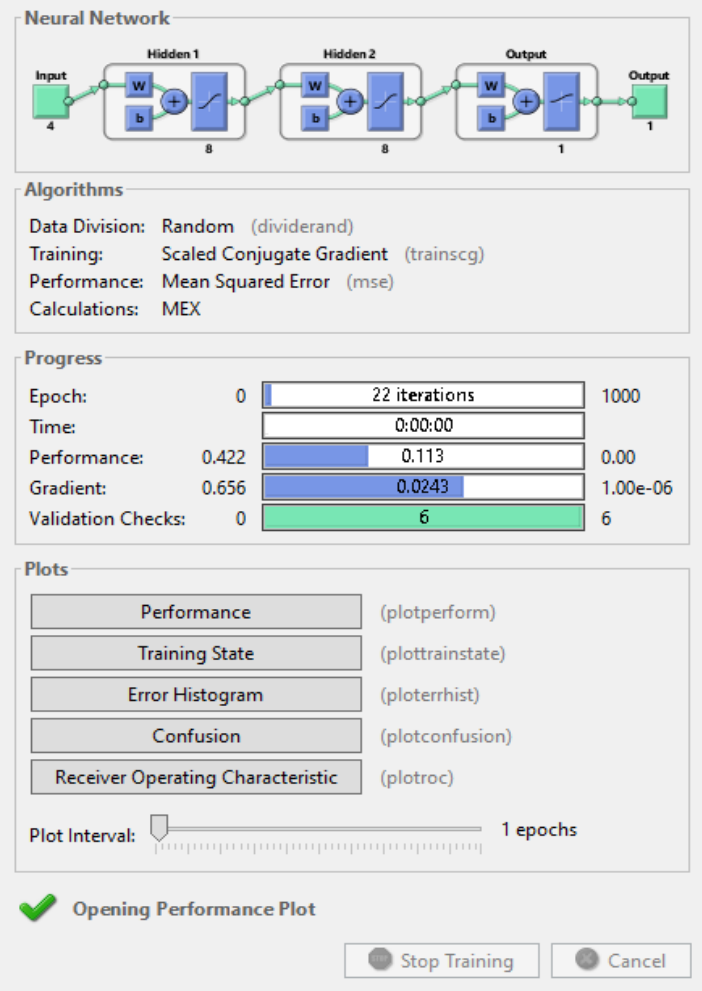

Figure 19: Thesis results. Architecture of Backpropagation Neural Network

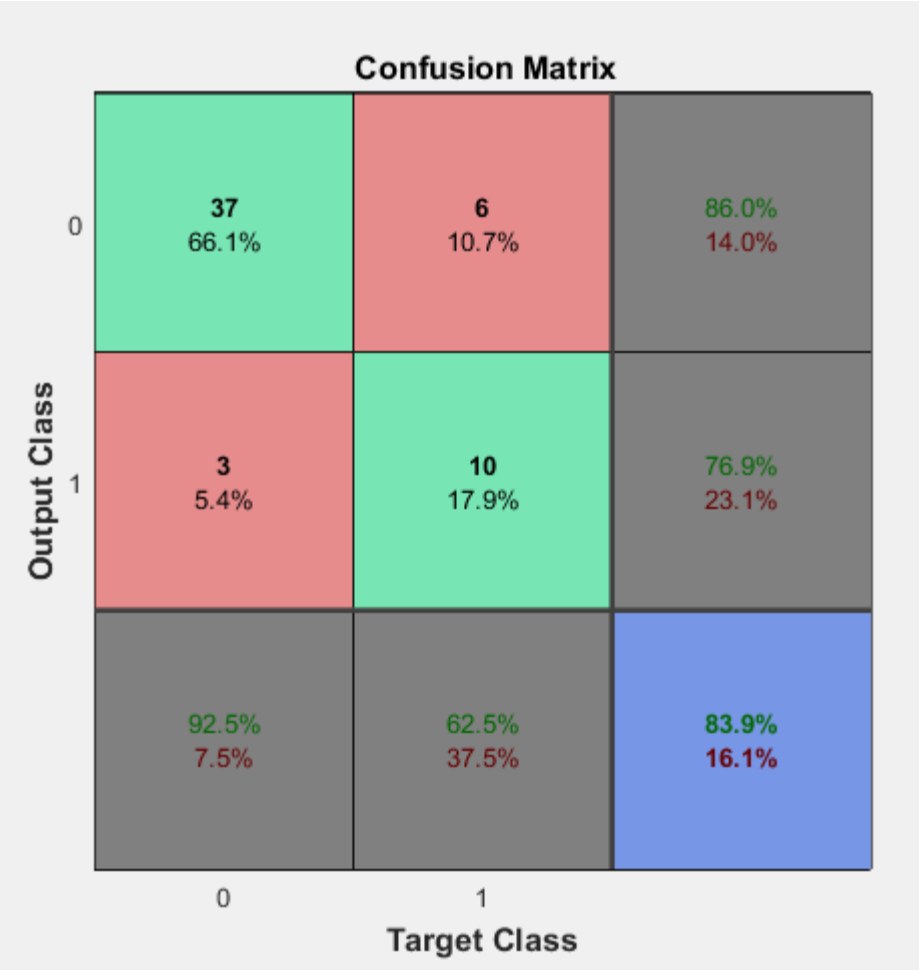

Figure 20: Thesis results of training ANN on Method 1 using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM)

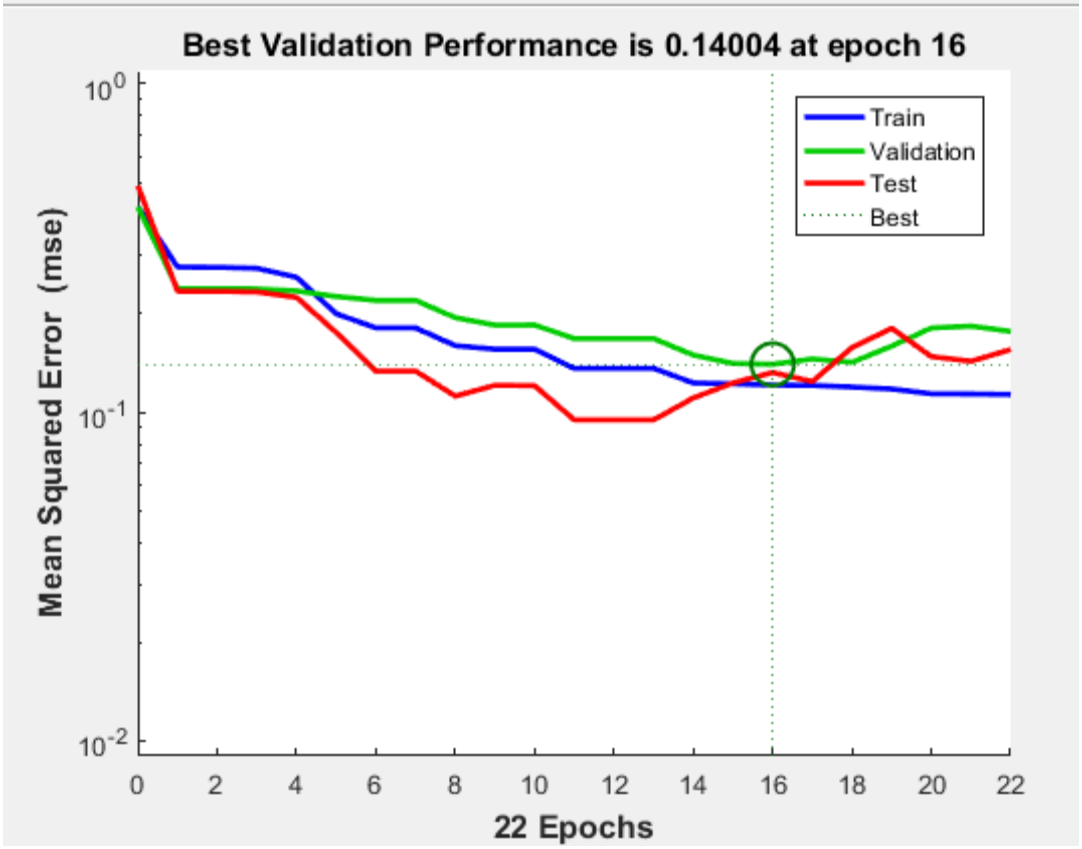

Figure 21: Thesis results of ANN on Method 1 using a subset of the database. Plot performance with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM)

From the performance plot we can see that the training of the neural network stops at the epoch 22 due to the *early stopping* method that *feedforwardnet* function uses. This method is used for improving generalization and prevent the problem of overfitting. When the validation error starts to increase it indicates that the network starts to overfit the data. In our network the default number of validation checks is used, so

```
net.trainParam.max_fail = 6
```

When the validation error increases for 6 iterations the training is stopped, and the weights and biases at the minimum of the validation error are returned.

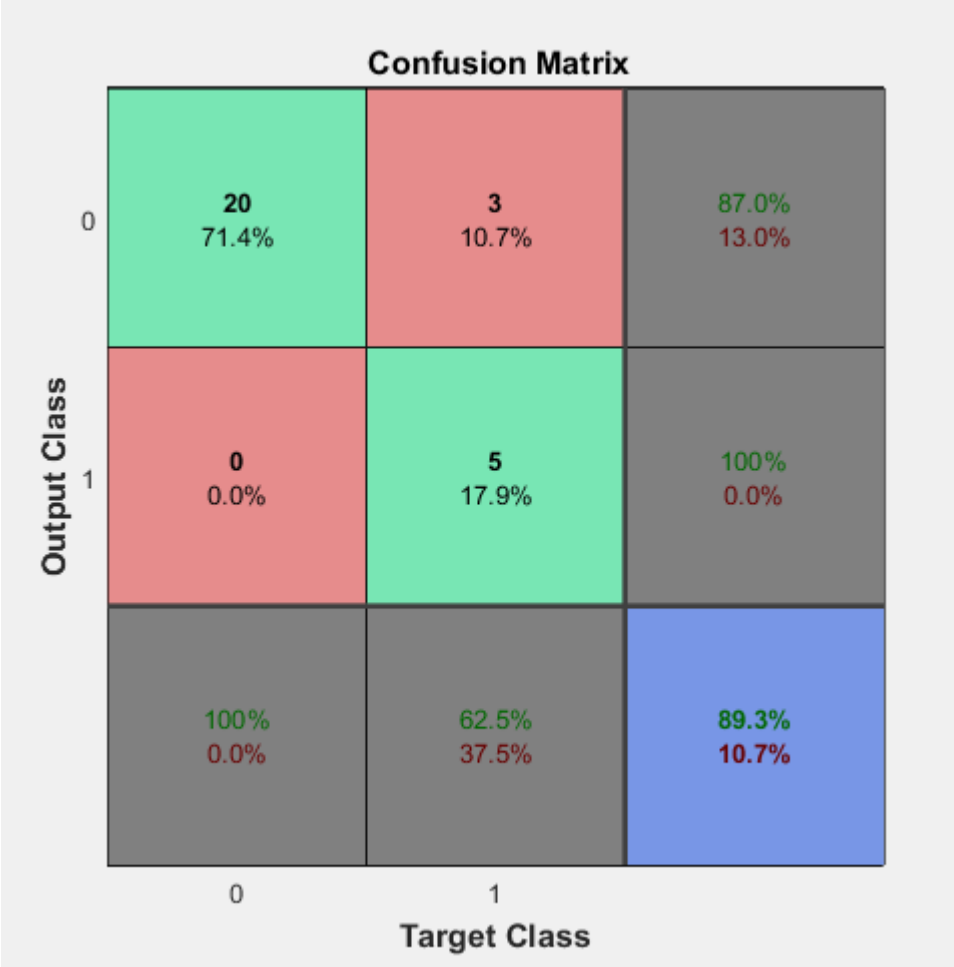

Figure 22: Thesis results of testing of ANN on Method 1 case of subset data. Confusion matrix of testing data with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM)

The implementation of the same network using the total number of patients gives the following results:

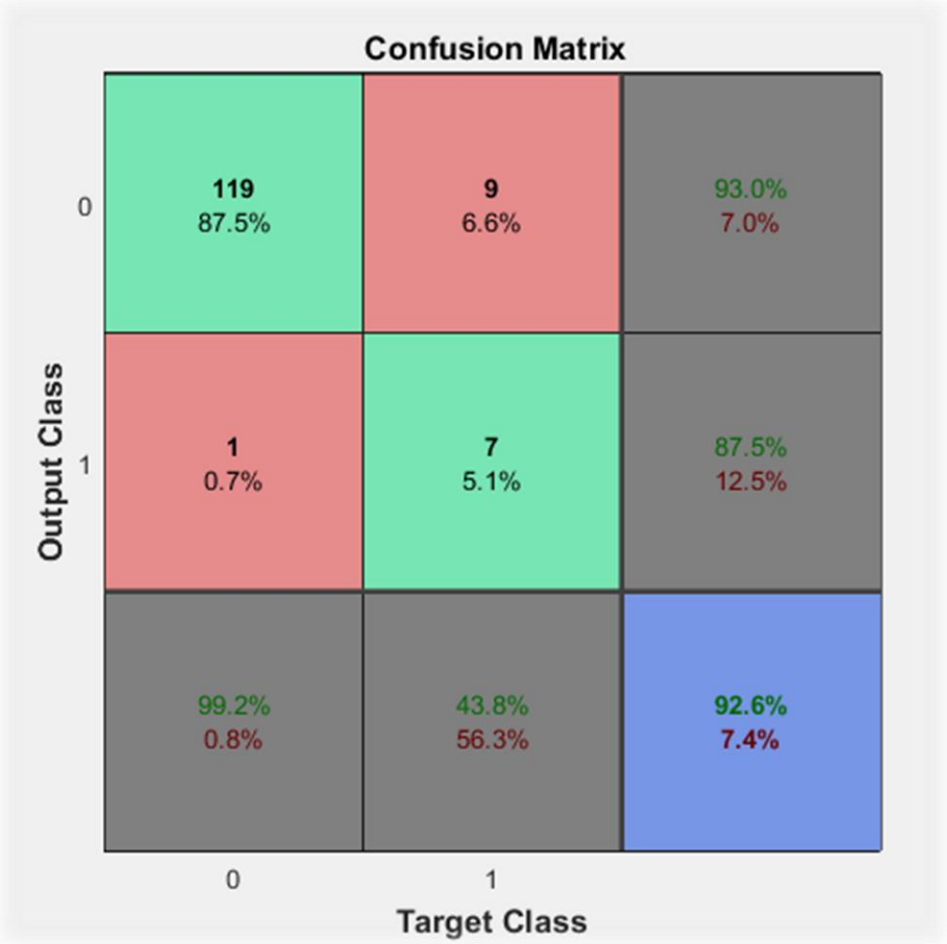

Figure 23: Thesis results of training ANN on Method 1 using the whole dataset. Confusion matrix of training data with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM)

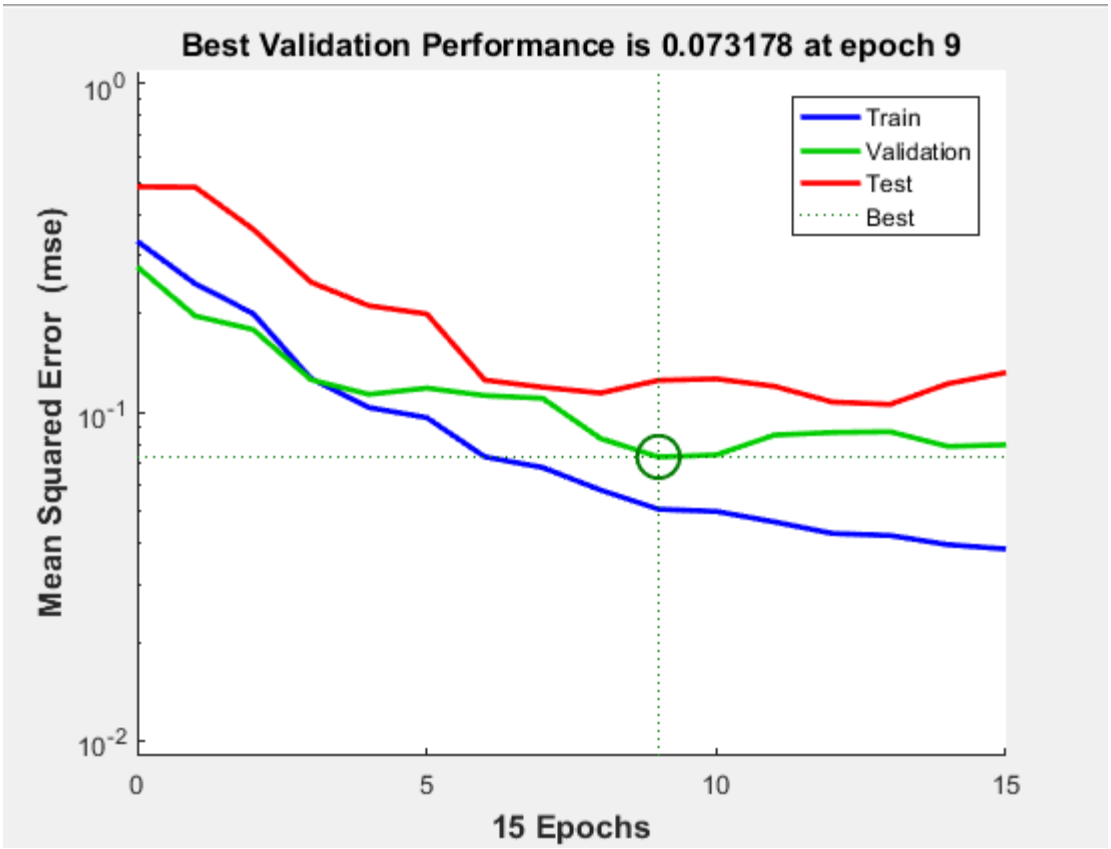

Figure 24: Thesis results of ANN on Method 1 using the whole dataset. Plot performance with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM)

From this performance plot we see that the validation error is better than the previous case using a subset of the database, although the training is stopped also before the desired validation goal.

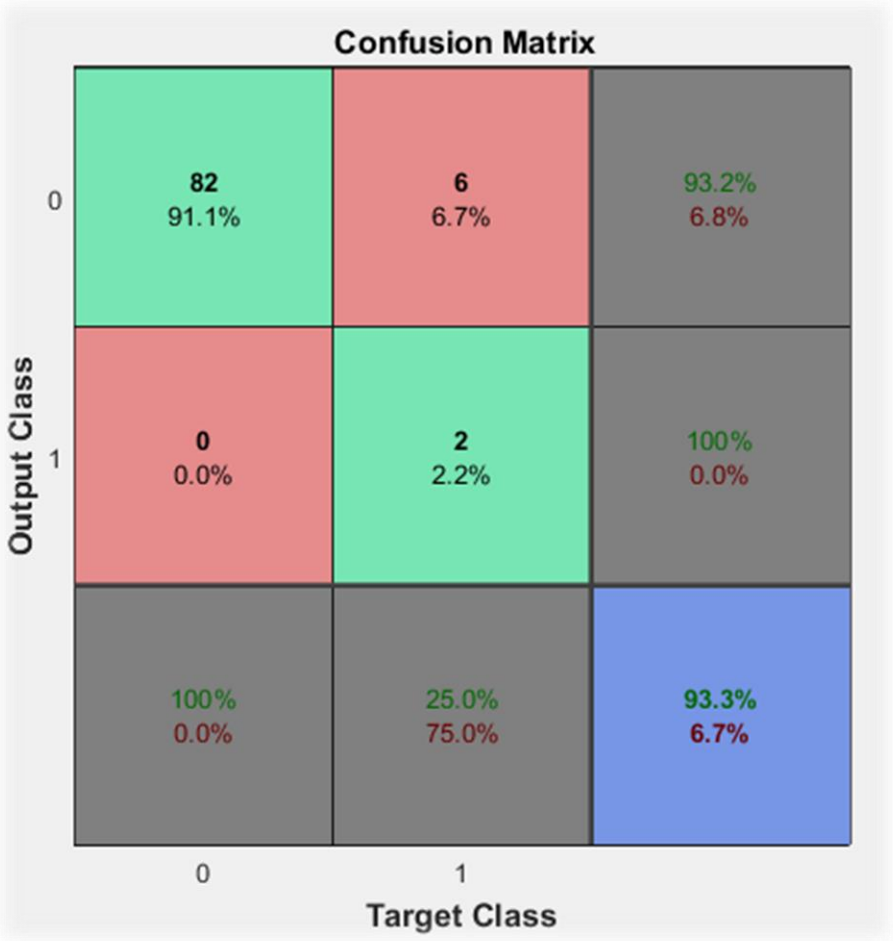

Figure 25: Thesis results of testing ANN on Method 1 using the whole dataset. Confusion matrix of testing data with inputs the features of energy, homogeneity, skewness and kurtosis based on Method 1(Otsu's Thresholding, Morphological Operators, GLCM)

We can observe that in Method 1 for the whole dataset the network is not generalized as we want it to be and struggles to classify properly the non-tumorous cases.

5.2.2 Results of ANFIS

After the implementation of the neural network on the features we extracted with Method 1 a different approach is examined. The use of ANFIS is considered for examining the improvement of our classification results.

Figure 26: Thesis results of training ANFIS on Method 1 using a subset of the database. Training plot of ANFIS for Method 1(Otsu's Thresholding, Morphological Operators, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness

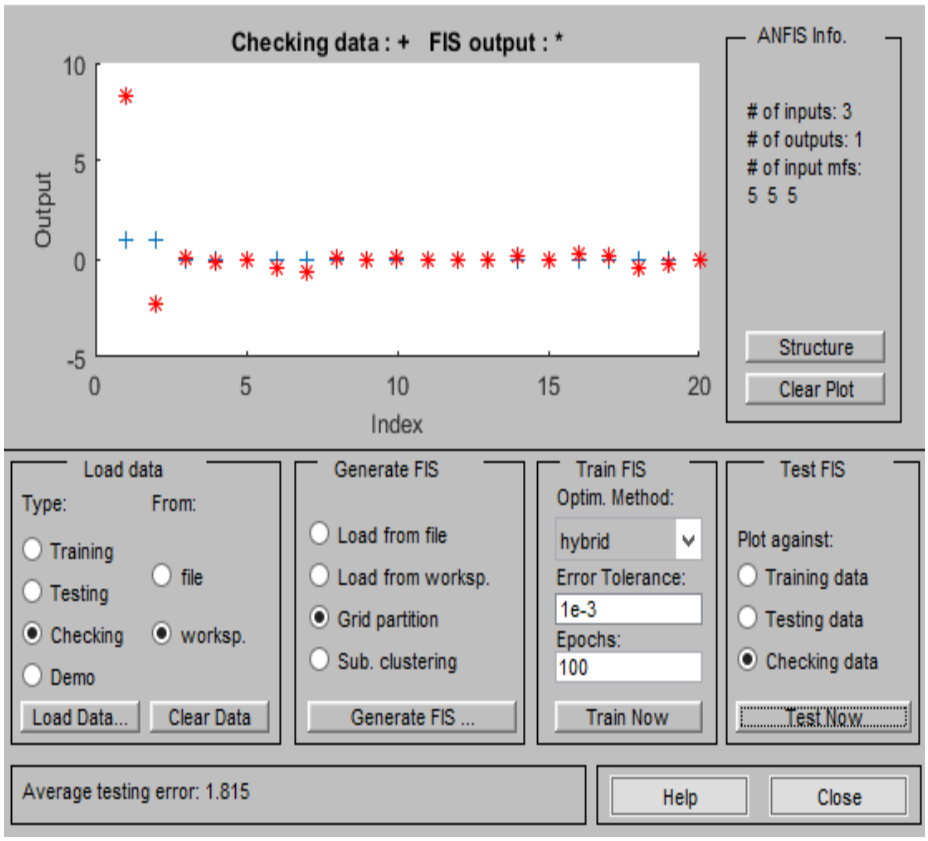

Figure 27: Thesis results of testing ANFIS on Method 1 using a subset of the database. Testing plot of ANFIS for Method 1(Otsu's Thresholding, Morphological Operators, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness

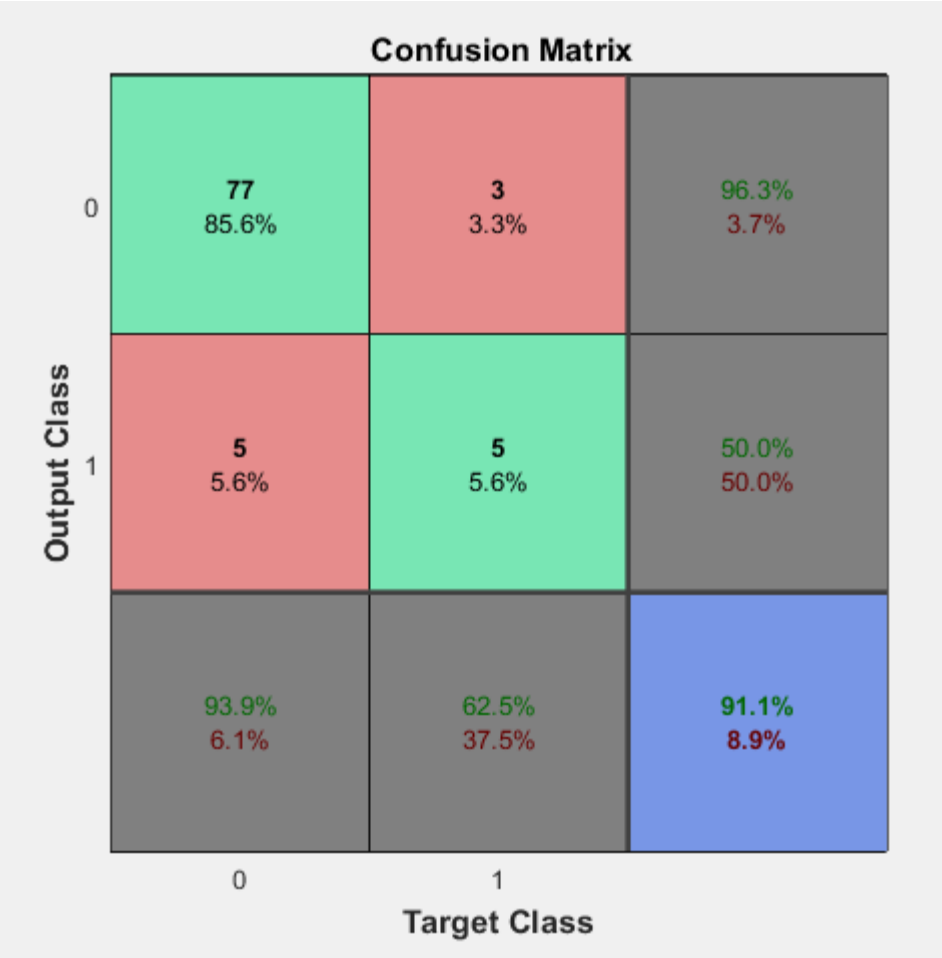

Figure 28: Thesis results of testing ANFIS on Method 1 using a subset of the database. Confusion matrix of testing data with ANFIS for Method 1(Otsu's Thresholding, Morphological Operators, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness

Evaluation parameters	Subset of database with ANN	Whole database with ANN	Whole database with ANFIS
True positive	20	82	77
False positive	3	6	3
True negative	5	2	5
False negative	0	0	5
Sensitivity (%)	100	100	93.9
Specificity (%)	62.5	25	62.5
Accuracy (%)	89.3	93.3	91.1

Table 5: Thesis results. Comparison of different classification results for Method 1 using different amount of database cases

We can clearly come to the conclusion that for Method 1 the best classification technique is the use of ANFIS which has worse *accuracy* value but far better *specificity* of the neural network that is used for the whole dataset of MRI.

5.3 Classification results of Method 2

5.3.1 Results of ANN

First we design a network for a dataset containing only some of the patients with ratio 40% healthy cases and 60% non-healthy cases. This network contains two hidden layers with 5 neurons each and has 179 inputs and one output.

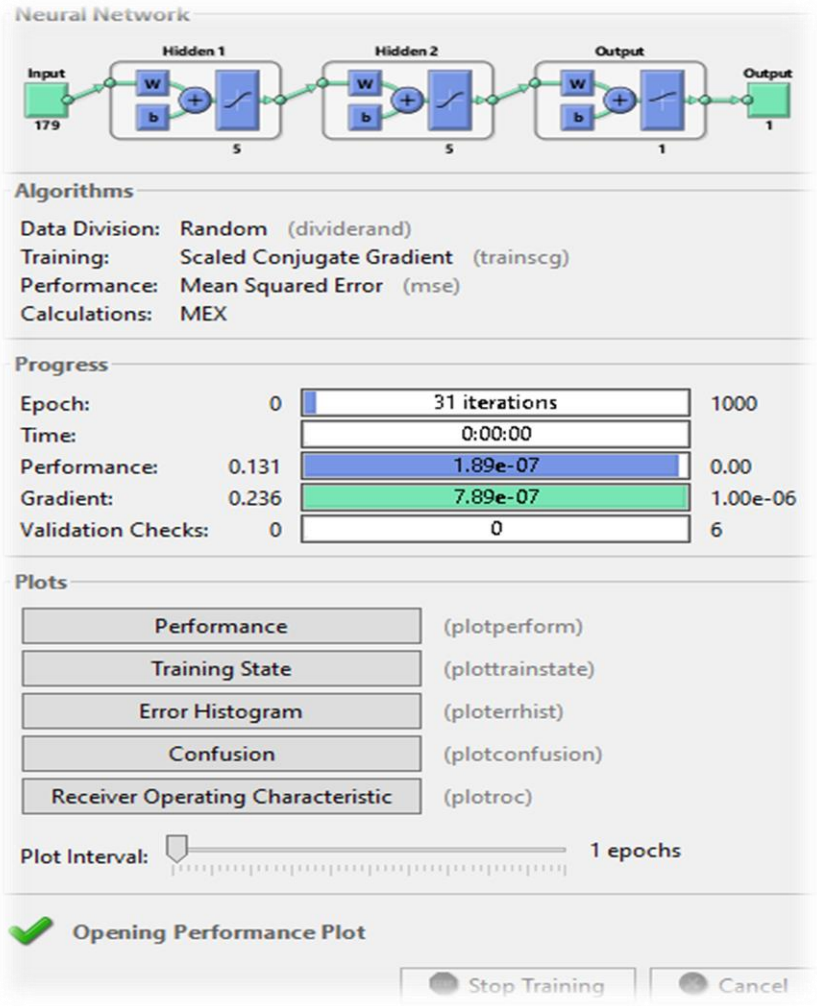

Figure 29: Thesis results of Method 2. Architecture of Backpropagation Neural Network

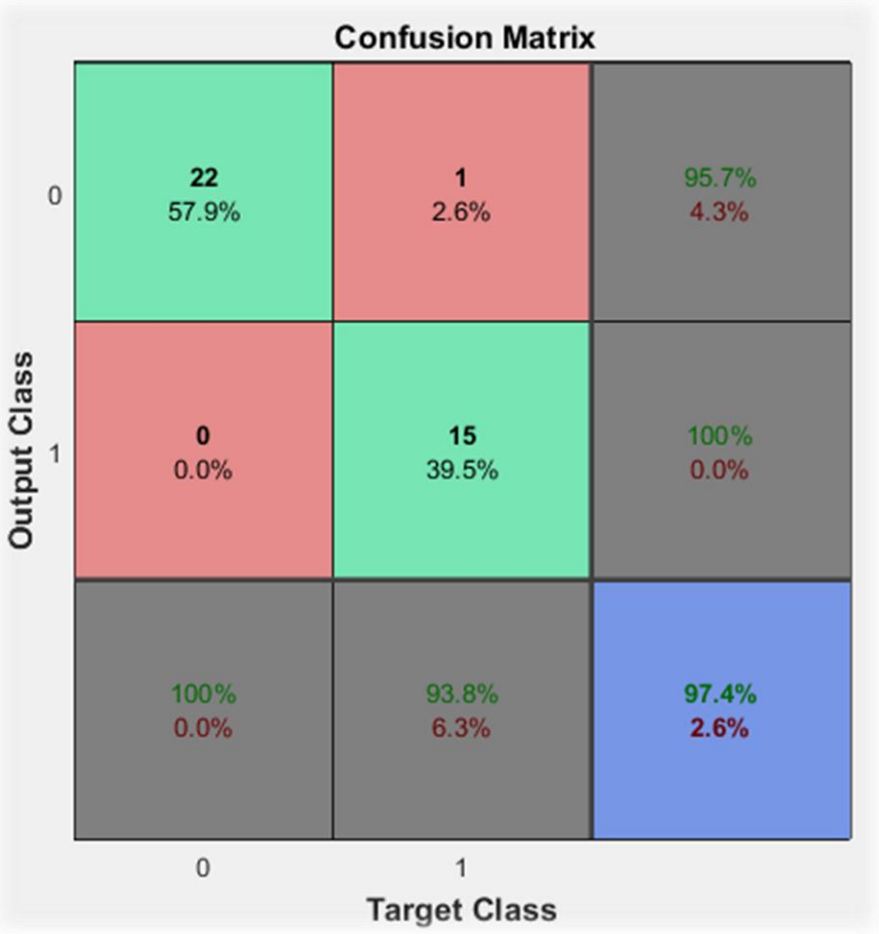

Figure 30: Thesis results of training ANN on Method 2 using a subset of the database. Confusion matrix of training data with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA)

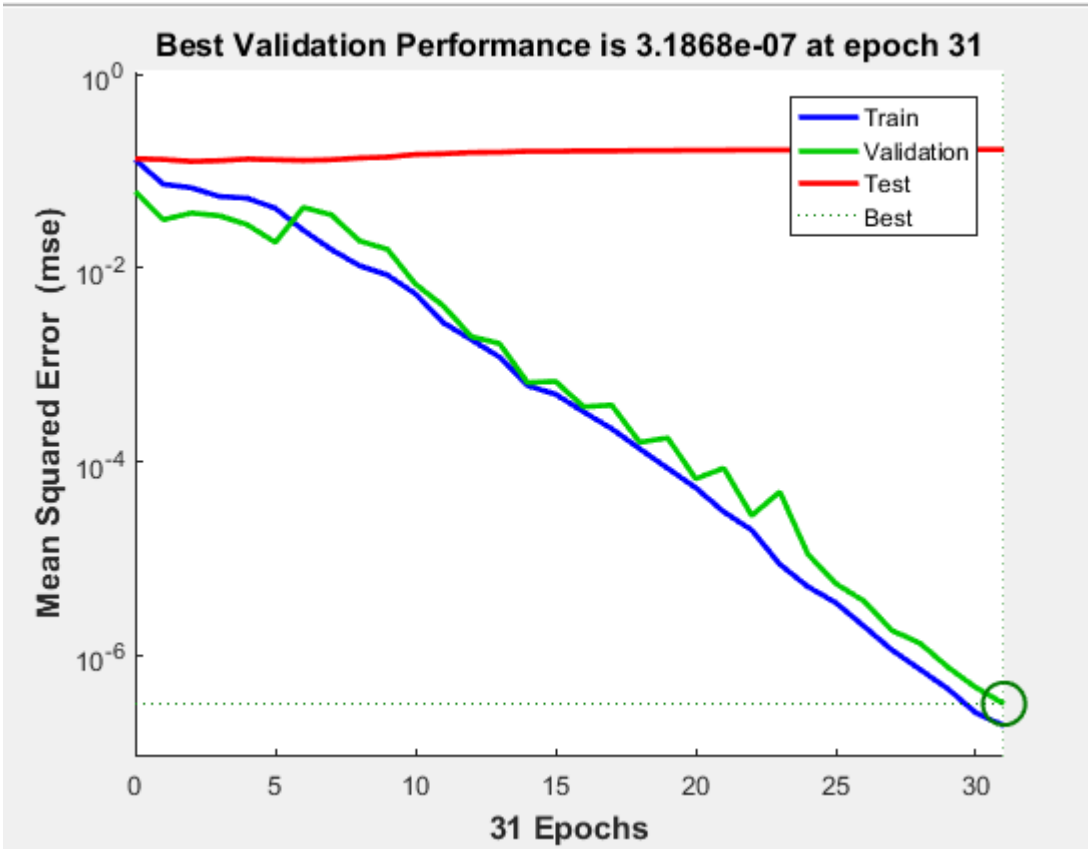

Figure 31: Thesis results of testing ANN on Method 2 using a subset of the database. Plot performance with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA)

The performance plot here indicates that the neural network is trained properly and the training process is stopped when the network reaches the specified acceptable validation error. The testing error does not have the same decreasing character which predisposes that the network is not perfect generalized.

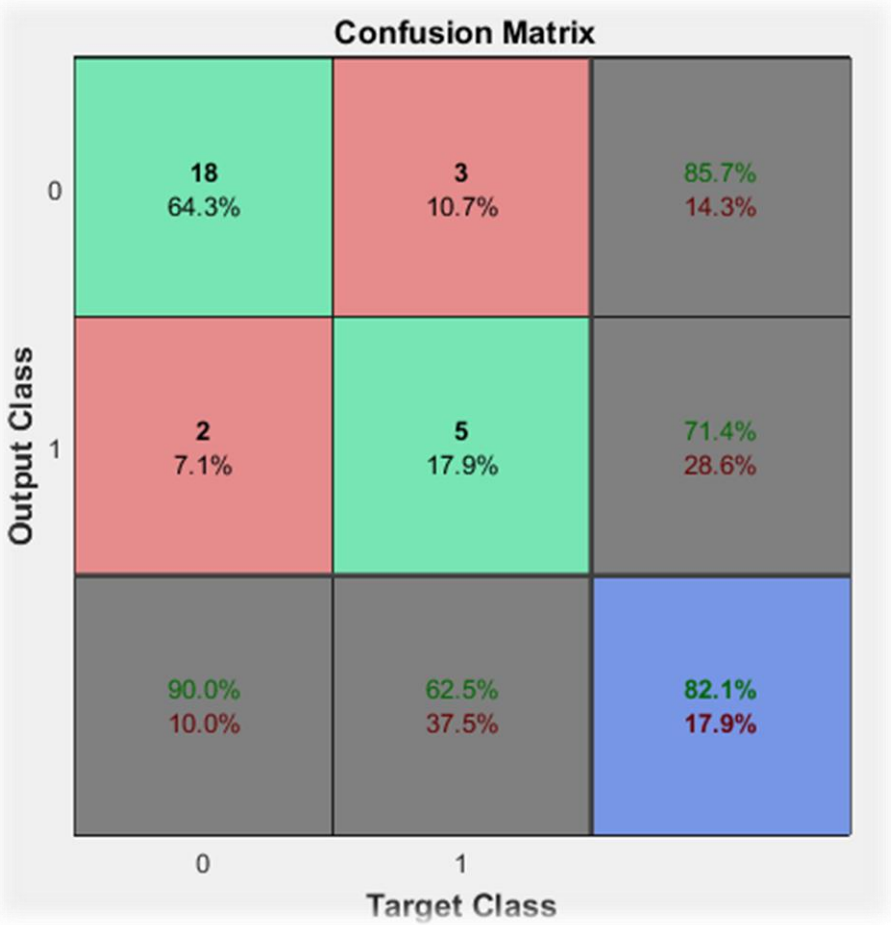

Figure 32: Thesis results of testing ANN on Method 2 using a subset of the database. Confusion matrix of testing data with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA)

Then, for classifying all cases from the whole dataset a new neural network is designed with *Levenberg-Marquardt* backpropagation function and two hidden layers with 25 and 30 neurons respectively. The network has 179 inputs and one output.

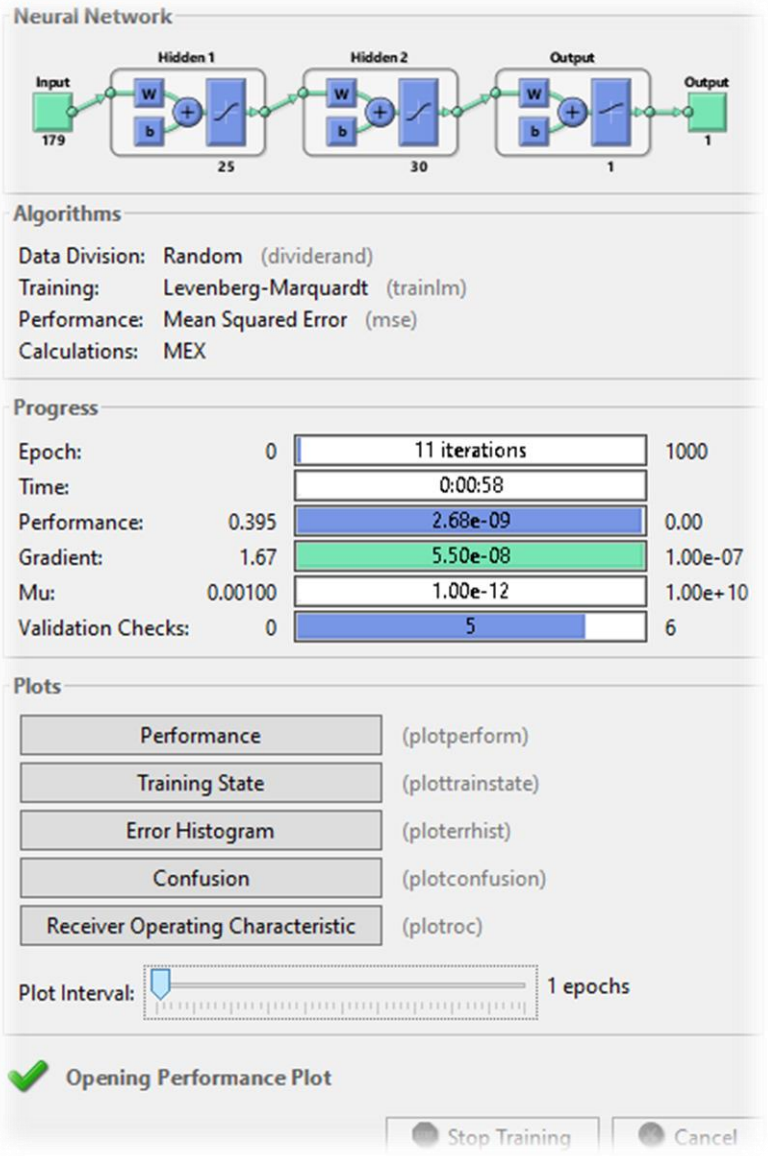

Figure 33: Thesis results of Method 2. Architecture of backpropagation neural network

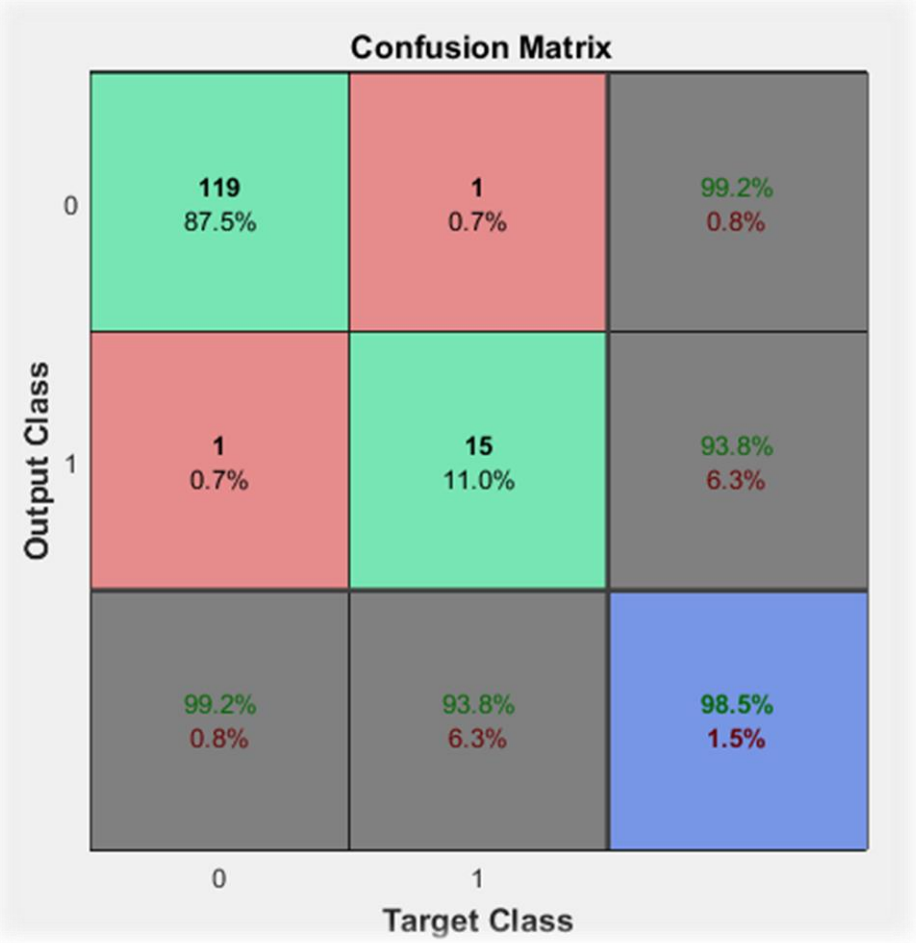

Figure 34: Thesis results of training ANN on Method 2 using the whole dataset. Confusion matrix of training data with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA)

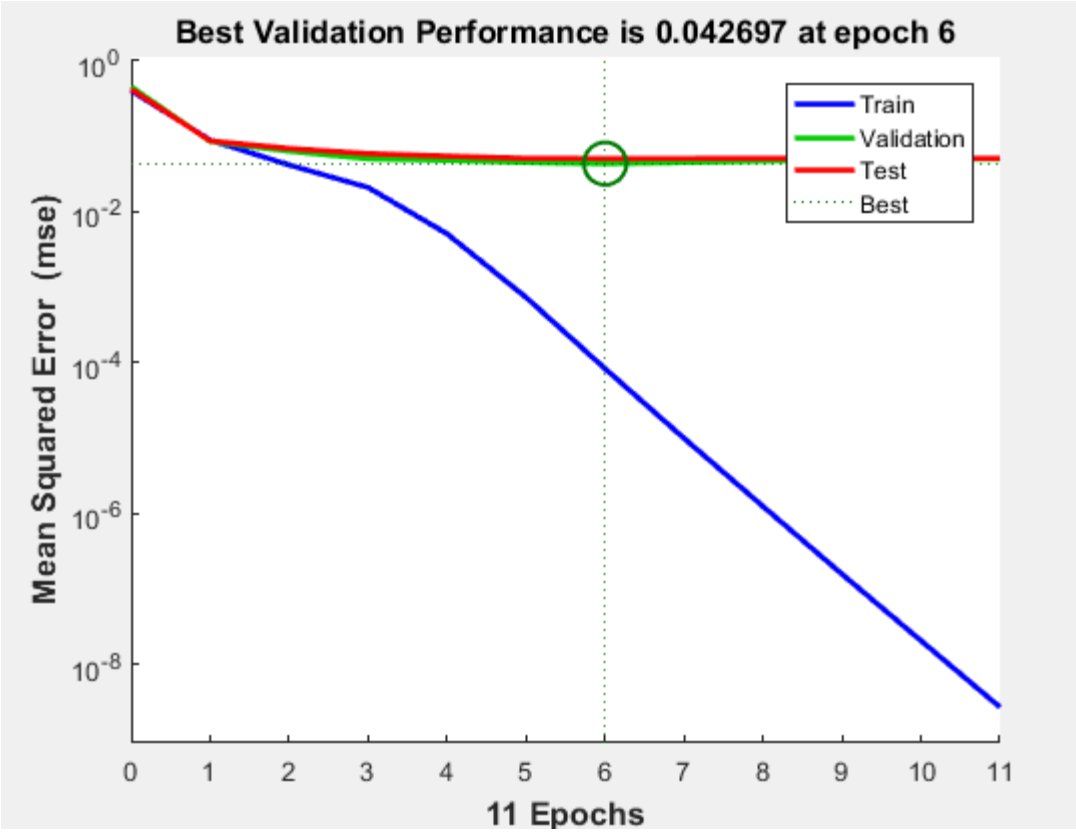

Figure 35: Thesis results of testing ANN on Method 2 using the whole dataset. Plot performance with inputs the features extracted from the most significant principal components based on Method 2 (Otsu's Thresholding, Morphological Operators, DWT, PCA)

The training error is decreasing significantly although the validation and testing error do not follow this decrease. This is expected as we increased the non-healthy cases. Here, the *early stopping* method is applied also for preventing overfitting and the training process stops at epoch 11.

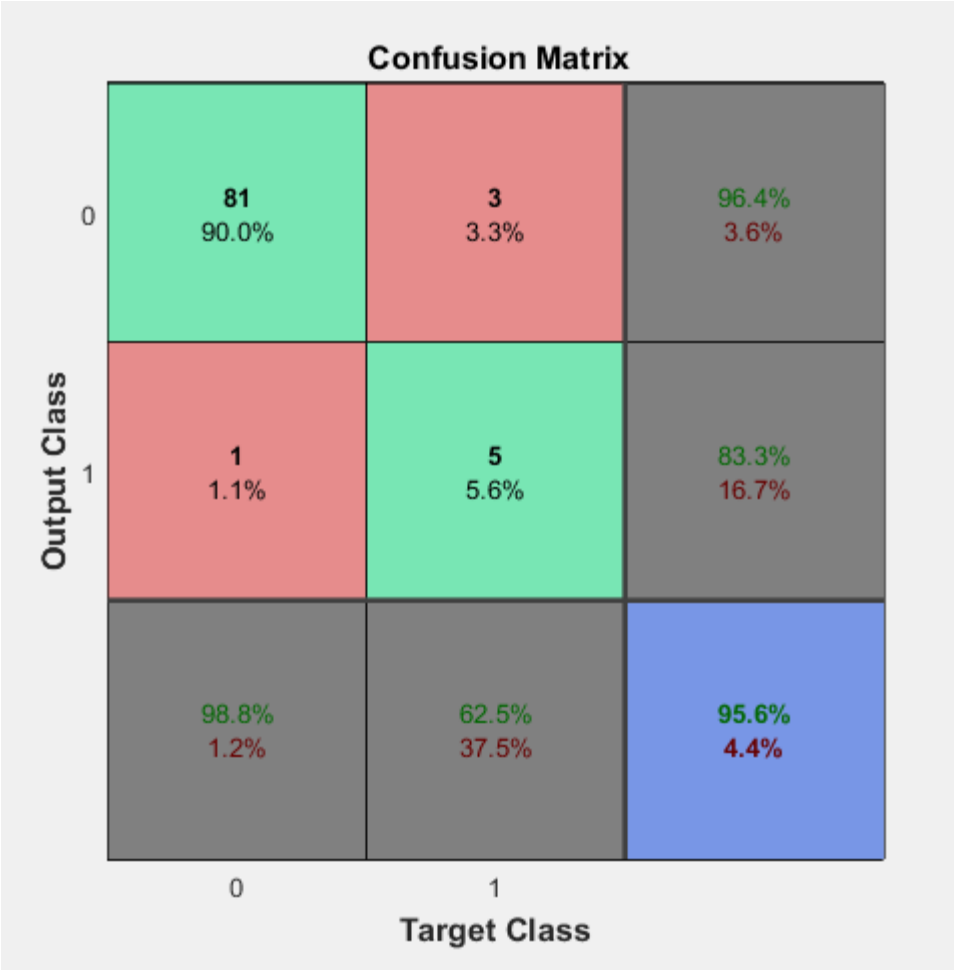

Figure 36: Thesis results of testing ANN on Method 2 using the whole dataset. Confusion matrix of testing data with inputs the features extracted from the most significant principal components based on Method 2 (Otsu' s Thresholding, Morphological Operators, DWT, PCA)

Evaluation parameters	Subset of database	Whole database
True positive	18	81
False positive	3	3
True negative	5	5
False negative	2	1
Sensitivity (%)	90	98.8
Specificity (%)	62.5	62.5
Accuracy (%)	82.1	95.6

Table 6: Thesis results. Comparison of different classification for Method 2 for different number of database cases

5.4 Classification results of Method 3

5.4.1 Results of ANN

First we designed a network for a dataset containing only some of the patients with ratio 30% healthy cases and 70% non-healthy cases. This network has 3 inputs and one output. It contains two hidden layers with 30 and 20 neurons respectively and uses the *Levenberg-Marquardt* backpropagation function.

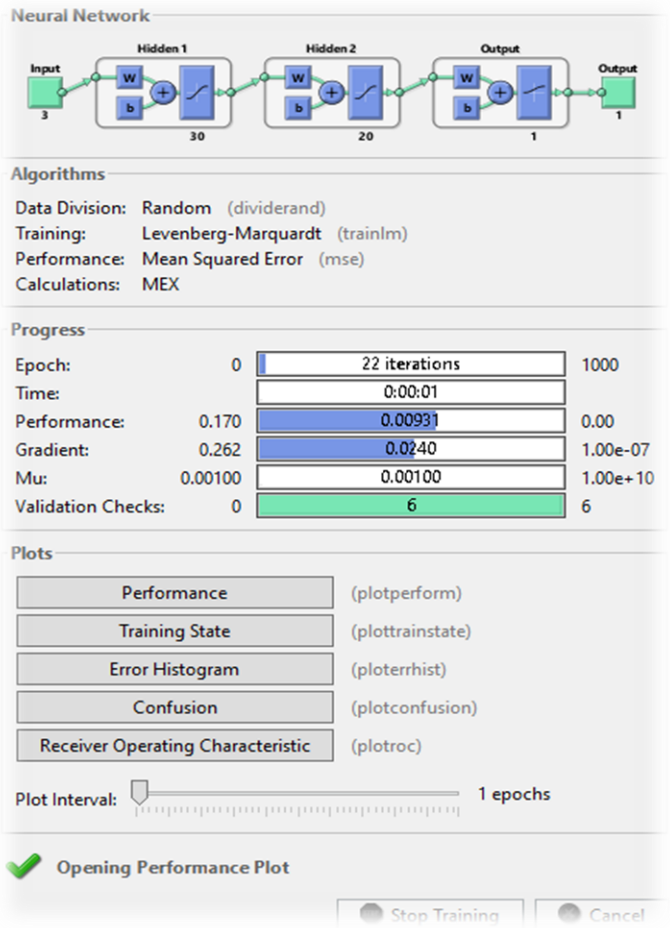

Figure 37: Thesis results for Method 3. Architecture of backpropagation neural network

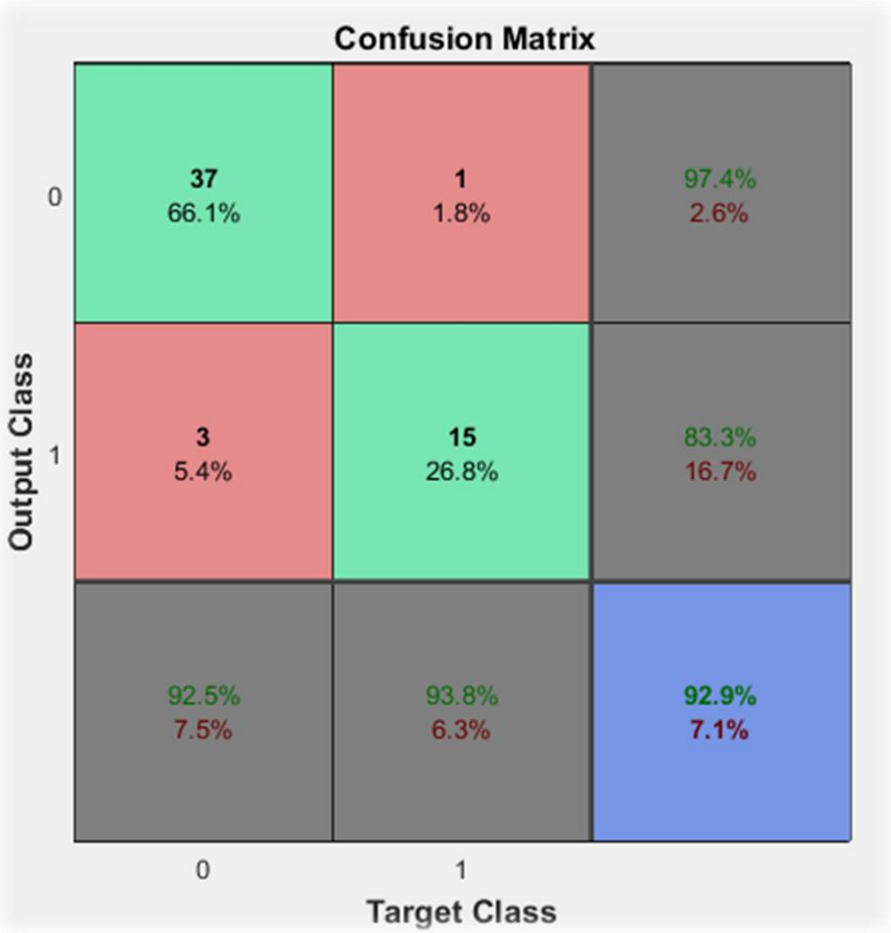

Figure 38: Thesis results of training ANN on Method 3 using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity and kurtosis based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

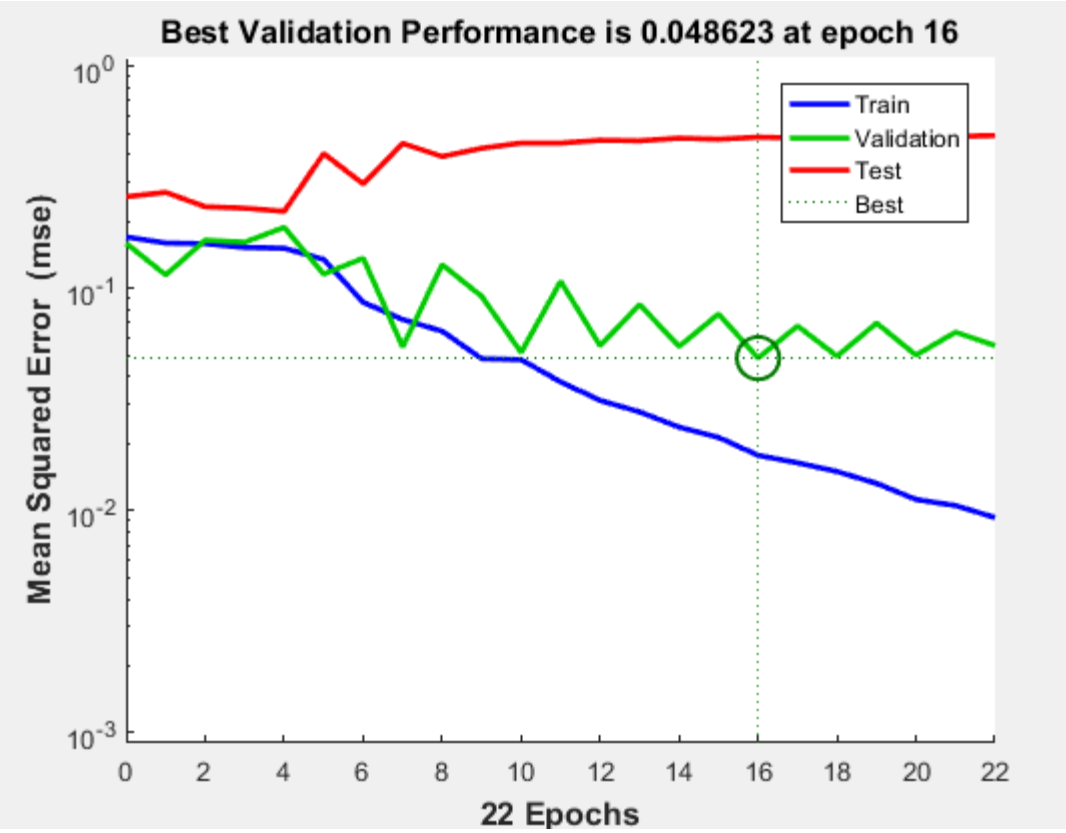

Figure 39: Thesis results of testing ANN on Method 3 using a subset of the database. Plot performance with inputs the features of energy, homogeneity and kurtosis based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

The training error is decreasing hesitantly while the validation error follows and the testing error remains steady. The training process stops at epoch 22 preventing the neural network to overfit the data and returns the best validation performance observed at epoch 16.

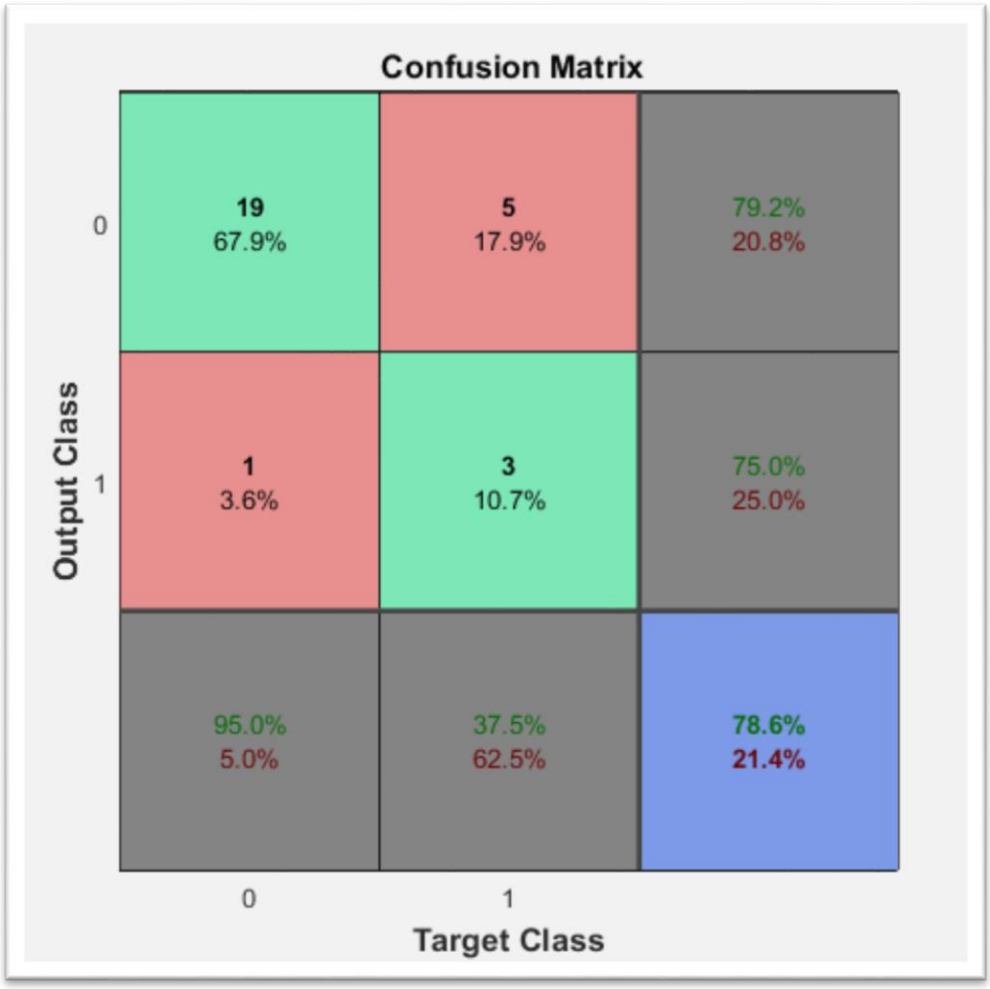

Figure 40: Thesis results of testing ANN on Method 3 using a subset of the database. Confusion matrix of testing data with inputs the features of energy, homogeneity and kurtosis based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

Then, for classifying all cases from the whole dataset a new neural network is designed with *Scaled Conjugate Gradient* backpropagation function and two hidden layers with 20 neurons each. The network has 4 inputs and one output.

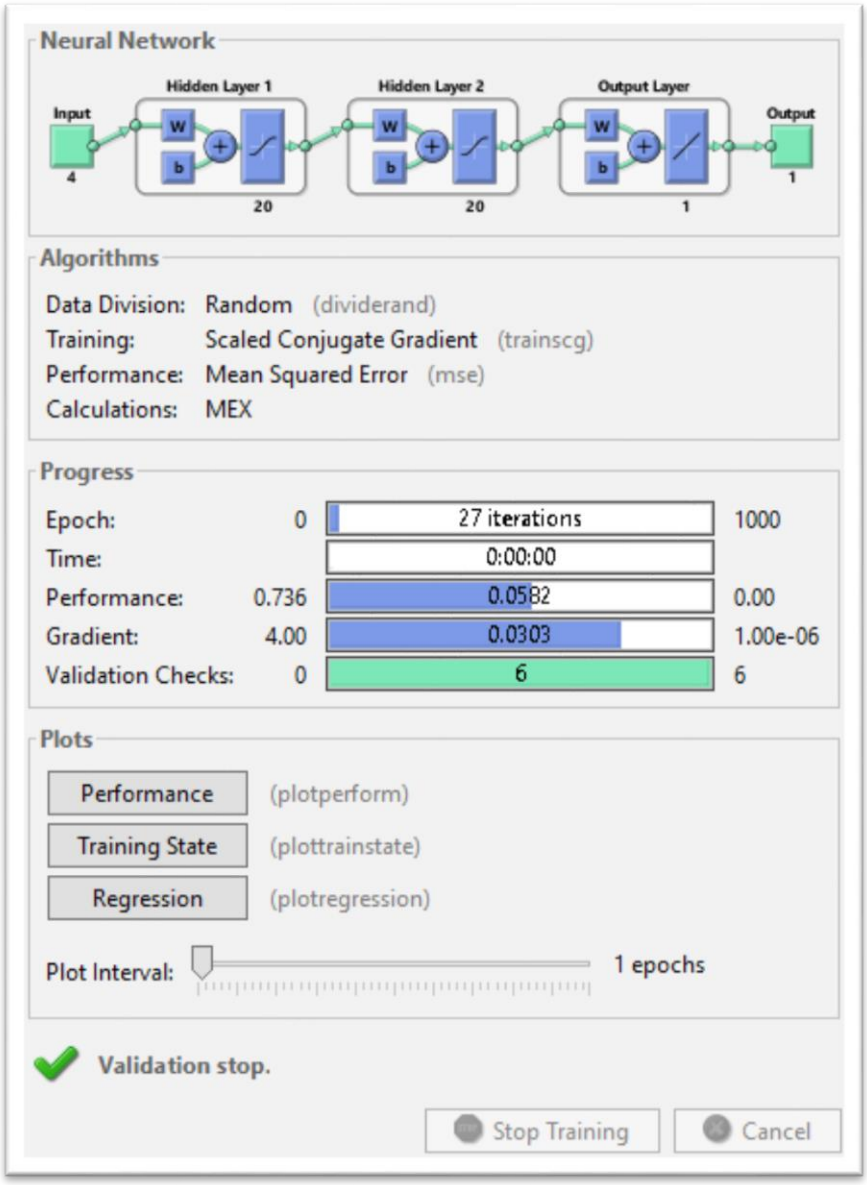

Figure 41: Thesis results for Method 3. Architecture of backpropagation neural network

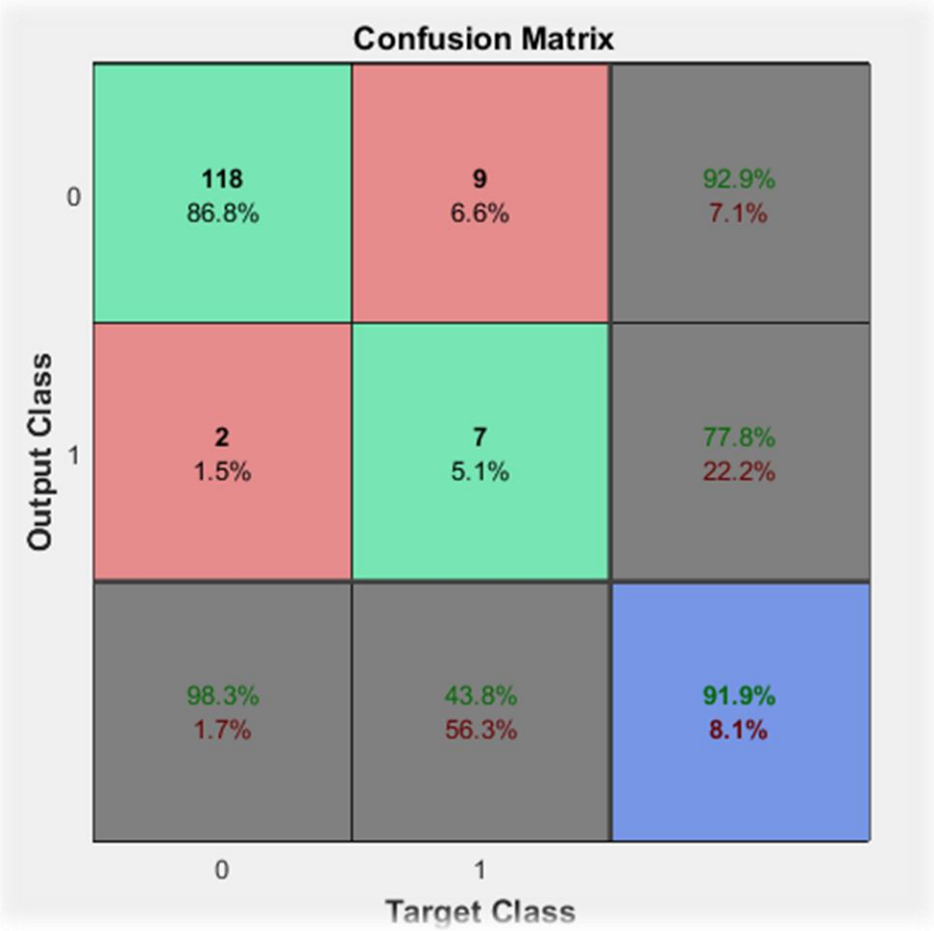

Figure 42: Thesis results of training ANN on Method 3 using the whole dataset. Confusion matrix of training data with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

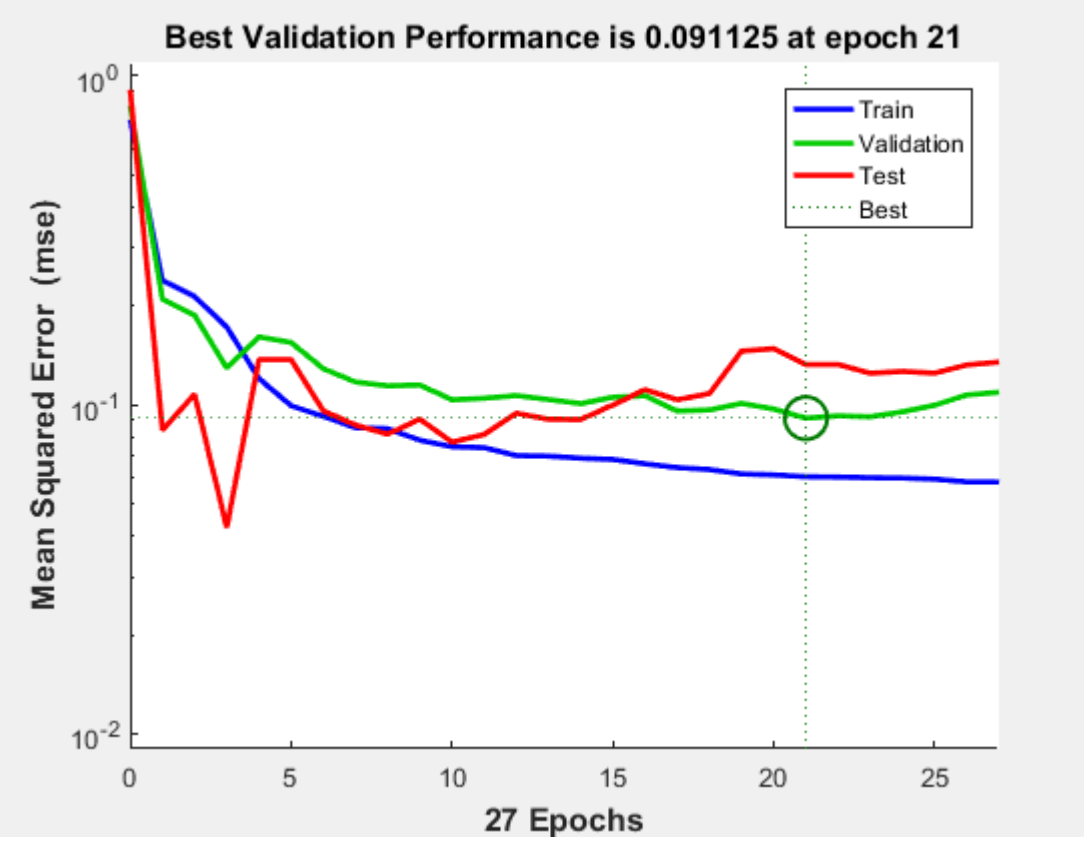

Figure 43: Thesis results of testing ANN on Method 3 using the whole dataset. Plot performance with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 3 (Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

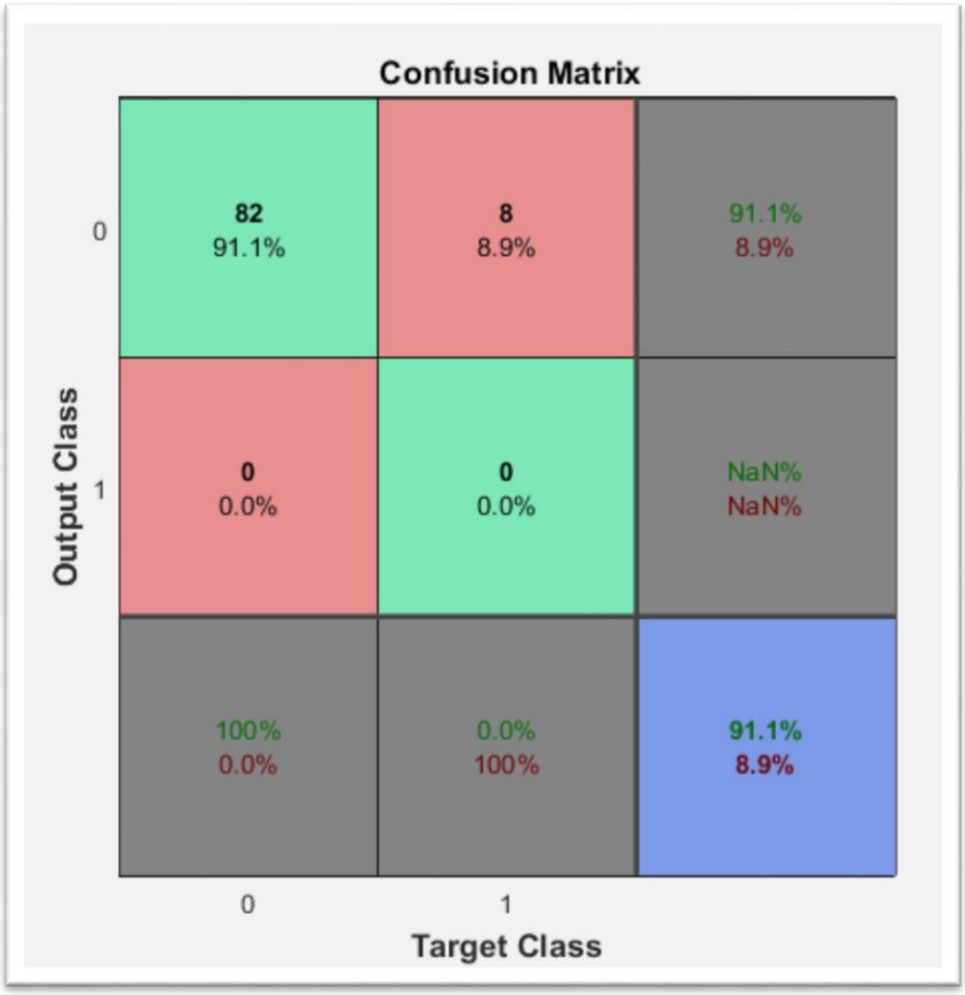

Figure 44: Thesis results of testing ANN on Method 3 using the whole dataset. Confusion matrix of testing data with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 3 (Otsus Thresholding, Morphological Operators, DWT, PCA, GLCM)

5.4.2 Results of ANFIS

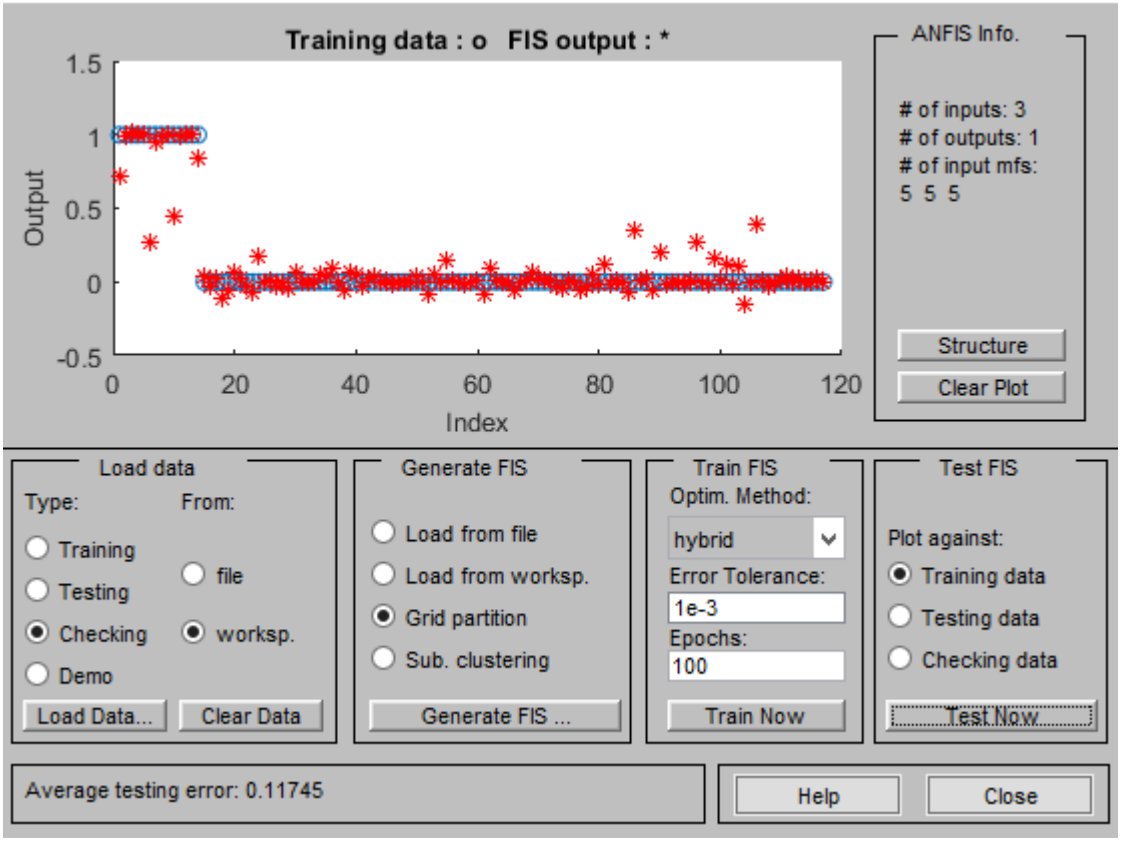

Figure 45: Thesis training results of ANFIS on Method 3 using the whole dataset. Training plot of ANFIS for Method 3(Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness

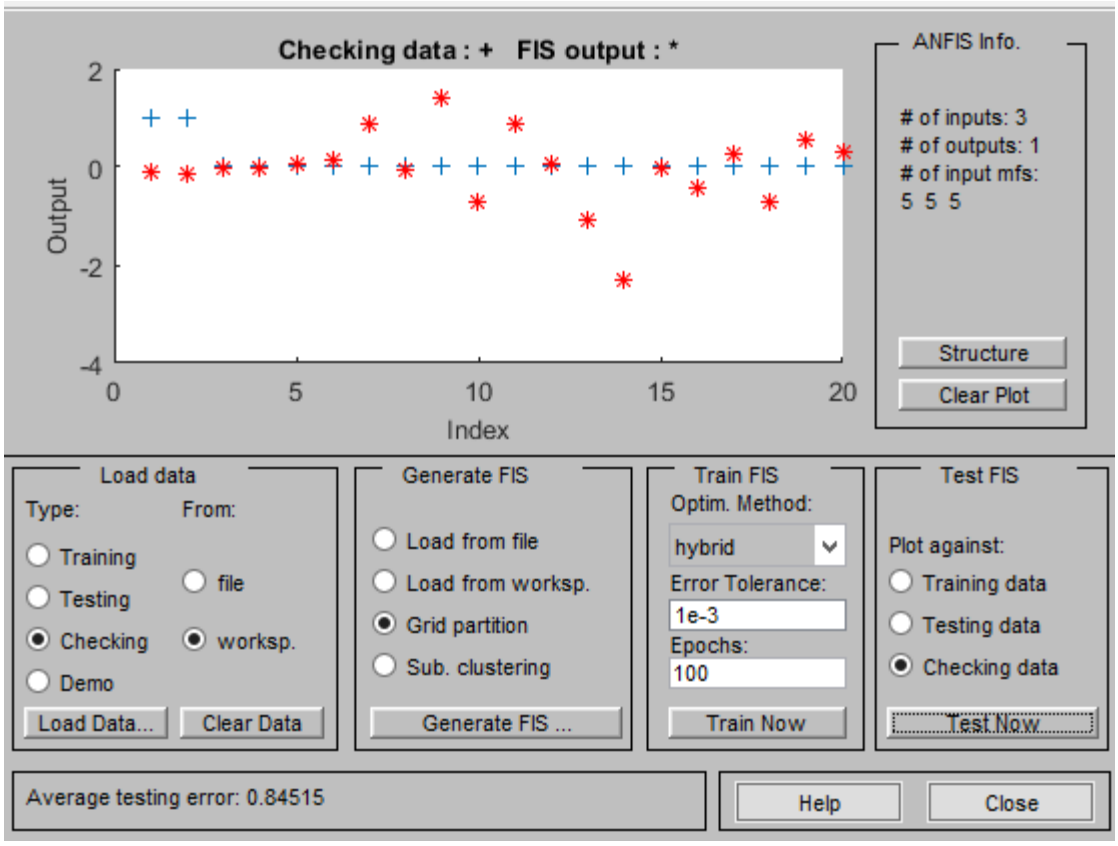

Figure 46: Thesis results of testing ANFIS on Method 3 using the whole dataset. Testing plot of ANFIS for Method 3(Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness

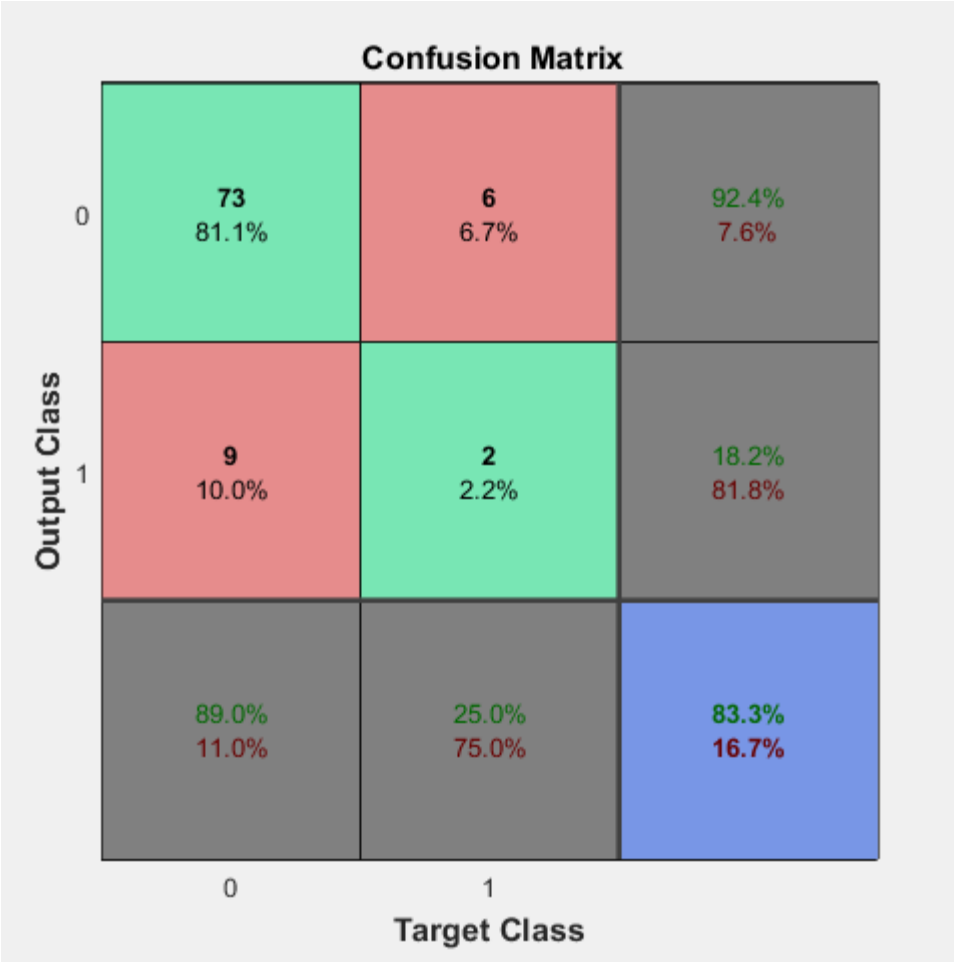

Figure 47: Thesis results of testing ANFIS on Method 3 using the whole dataset. Confusion matrix of testing data with ANFIS for Method 3(Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity, kurtosis and skewness

Evaluation parameters	Subset of database with ANN	Whole database with ANN	Whole database with ANFIS
True positive	19	82	73
False positive	5	8	6
True negative	3	0	2
False negative	1	0	9
Sensitivity (%)	95	100	89
Specificity (%)	37.5	0	25
Accuracy (%)	78.6	91.1	83.3

Table 7: Thesis results. Comparison of different classification results for Method 3 for different number of database cases

5.5 Classification results of Method 4

5.5.1 Results of ANN

First we designed a network for a dataset containing only some of the patients with ratio 30% healthy cases and 70% non-healthy cases. This network has 3 inputs and one output. It contains two hidden layers with 15 and 10 neurons respectively and uses the *Scaled Conjugate Gradient* backpropagation function.

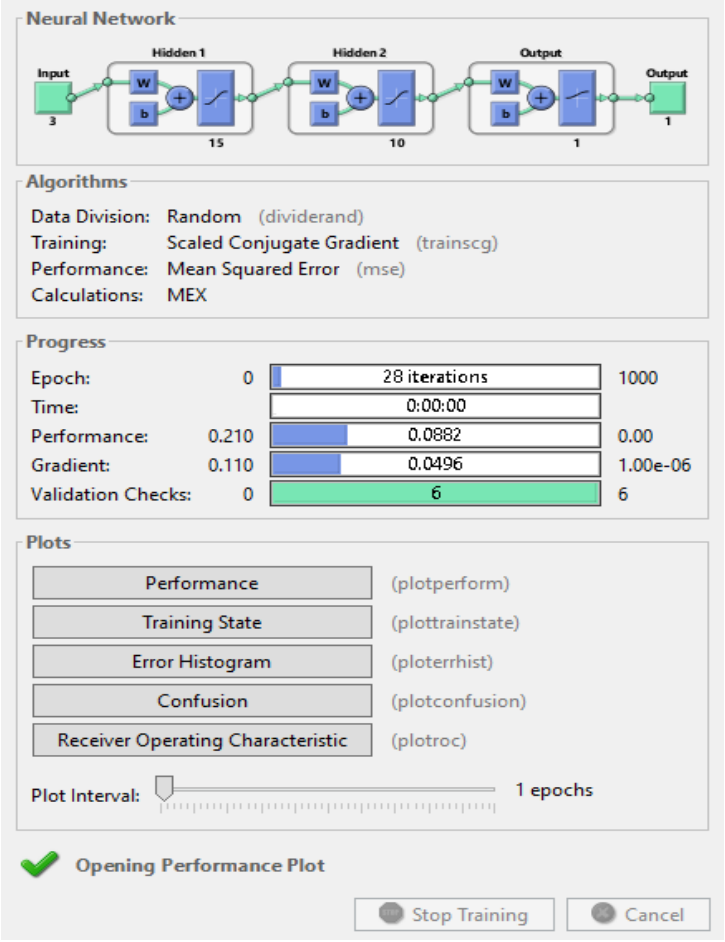

Figure 48: Thesis results of Method 4. Architecture of backpropagation neural network

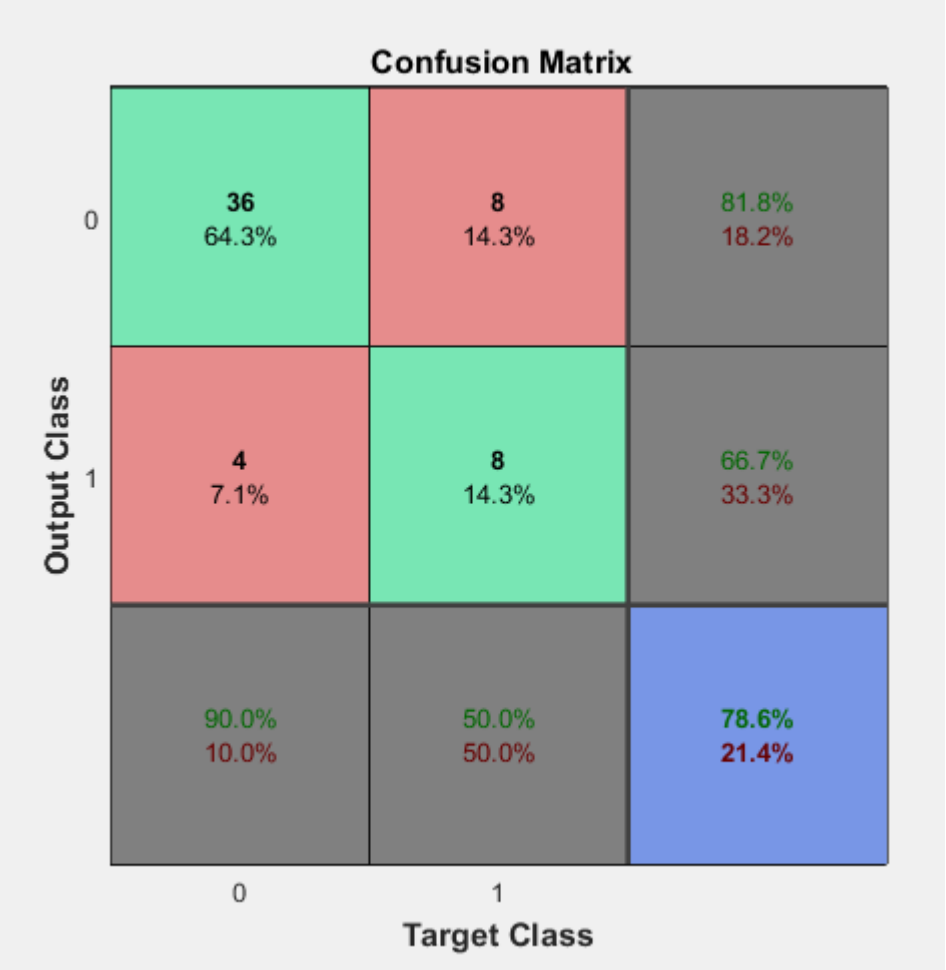

Figure 49: Thesis results of training ANN on Method 4 using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity and kurtosis based on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

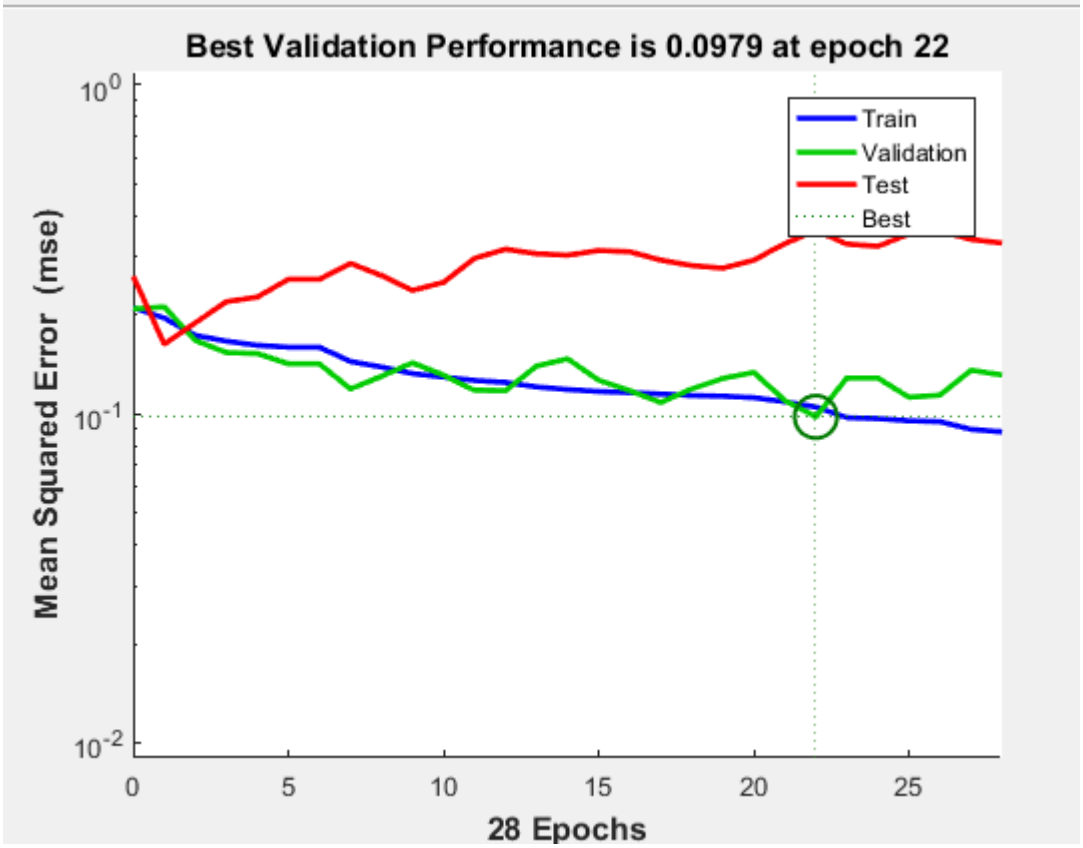

Figure 50: Thesis results of testing ANN on Method 4 using a subset of the database. Plot performance with inputs the features of energy, homogeneity and kurtosis based on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

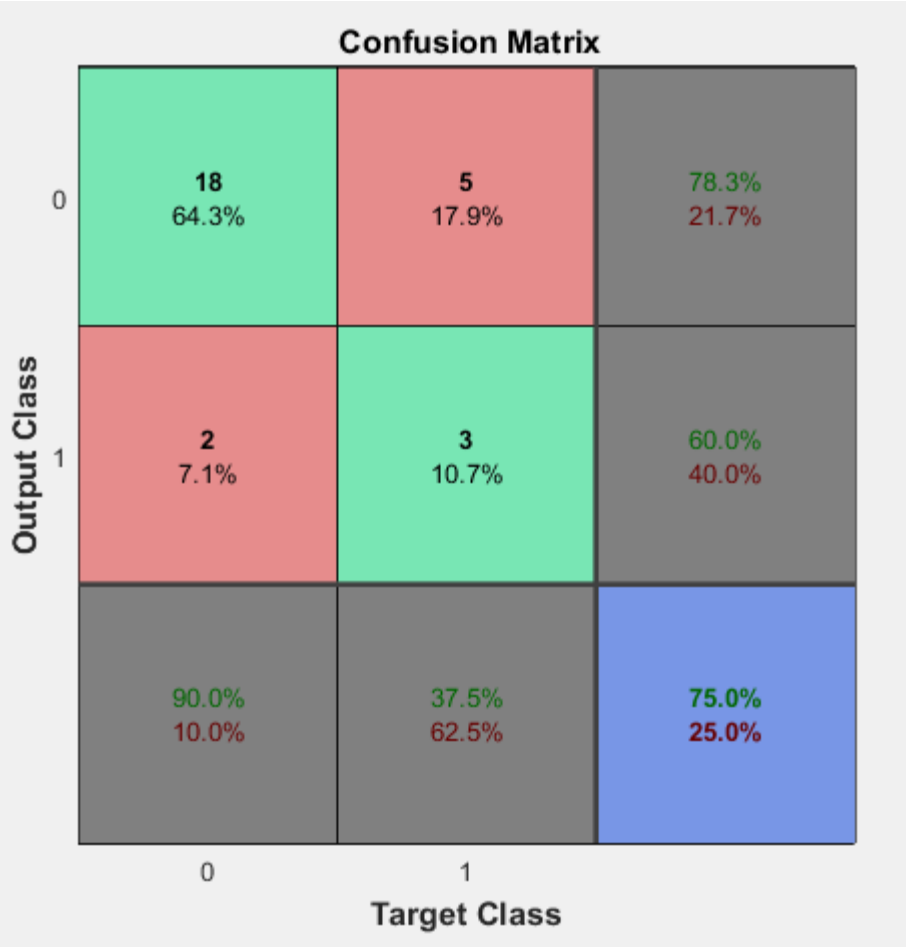

Figure 51: Thesis results of testing ANN on Method 4 using a subset of the database. Confusion matrix of testing data with inputs the features of energy, homogeneity and kurtosis based on Method 4 (Mean-Shift segmentation, Otsu' s Thresholding, Morphological Operators, DWT, PCA, GLCM)

Then, for classifying all cases from the whole dataset a new neural network is designed with *Scaled Conjugate Gradient* backpropagation function and two hidden layers with 30 and 20 neurons respectively. The network has 4 inputs and one output.

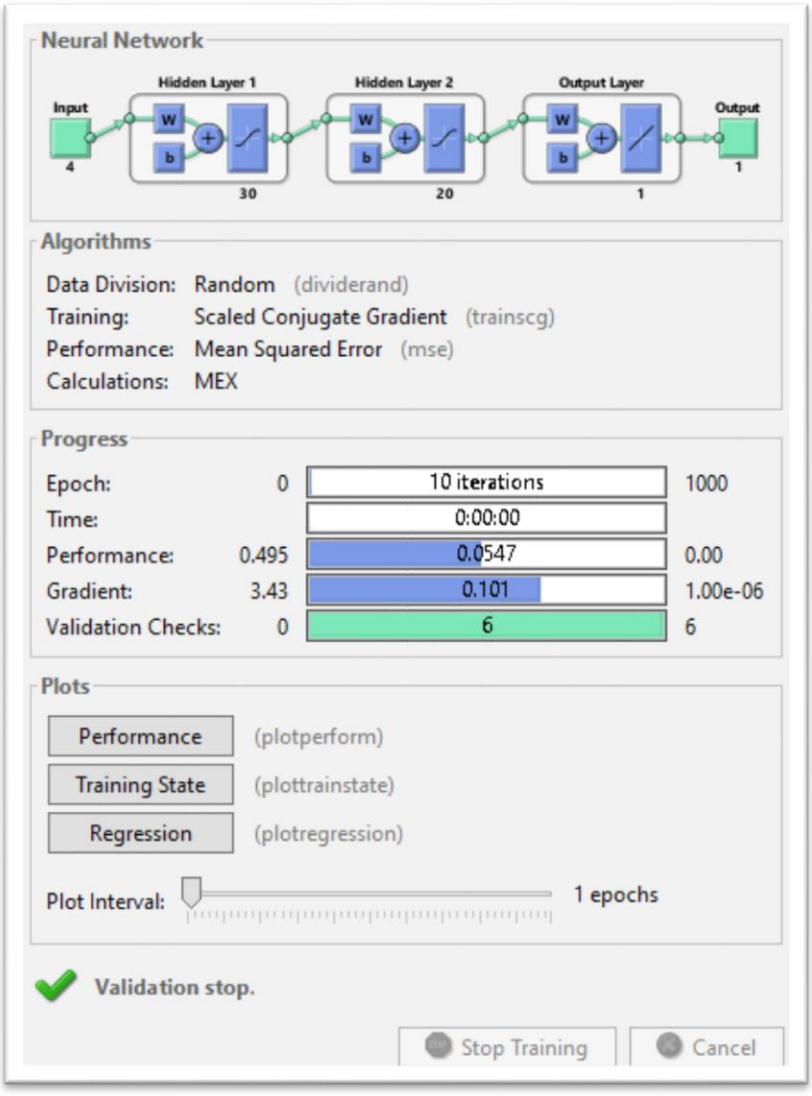

Figure 52: Thesis rsults for Method 4. Architecture of backpropagation neural network

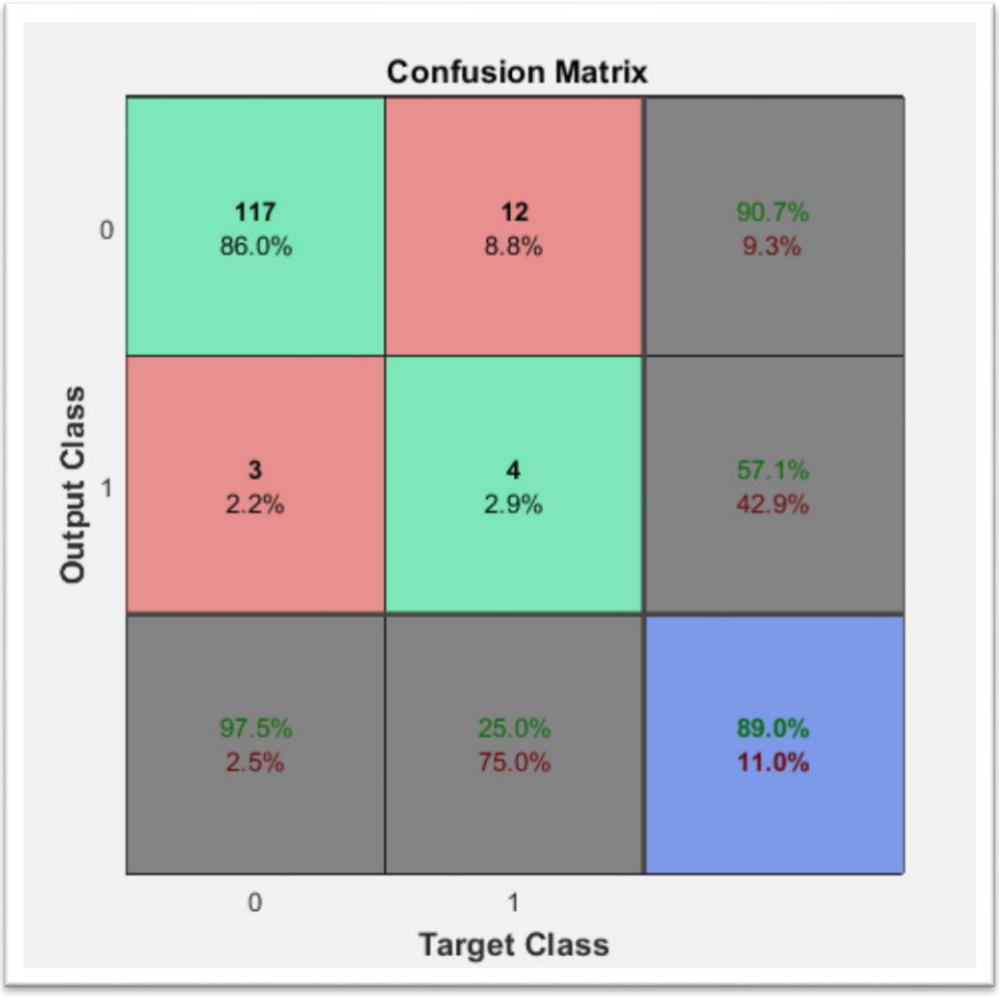

Figure 53: Thesis results of training ANN on Method 4 using the whole dataset. Confusion matrix of training data with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

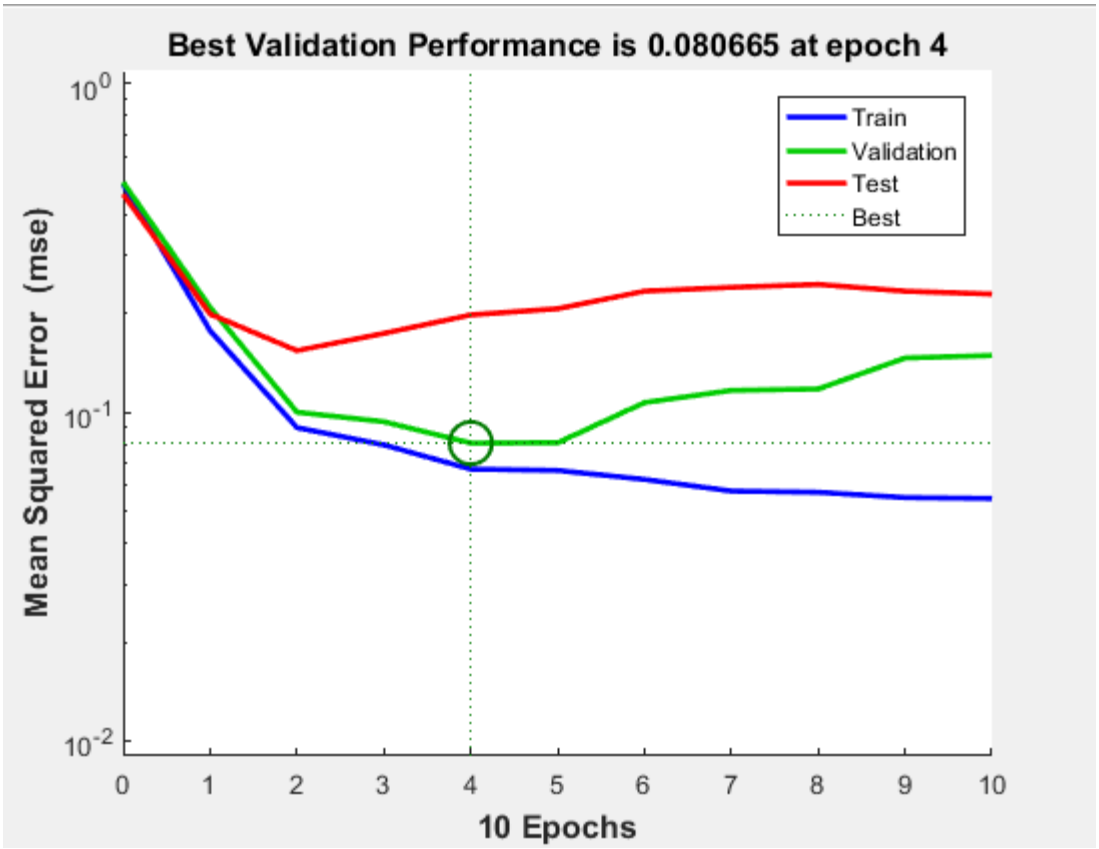

Figure 54: Thesis results of testing ANN on Method 4 using the whole dataset. Plot performance with inputs the features of energy, homogeneity, kurtosis and skewness based on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

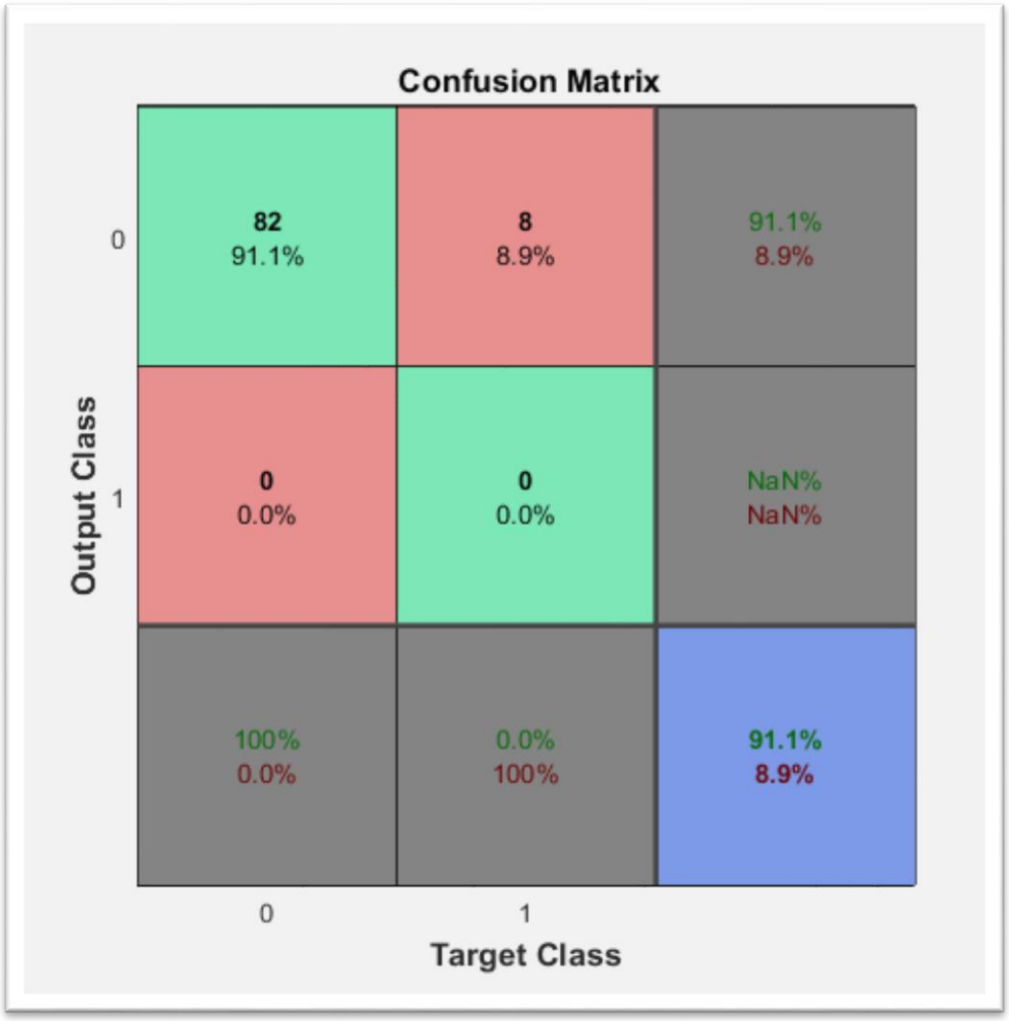

Figure 55: Thesis results of testing ANN on Method 4 using the whole dataset. Confusion matrix of testing data with inputs the features of energy, homogeneity, kurtosis and skewnessbased on Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)

5.5.2 Results of ANFIS

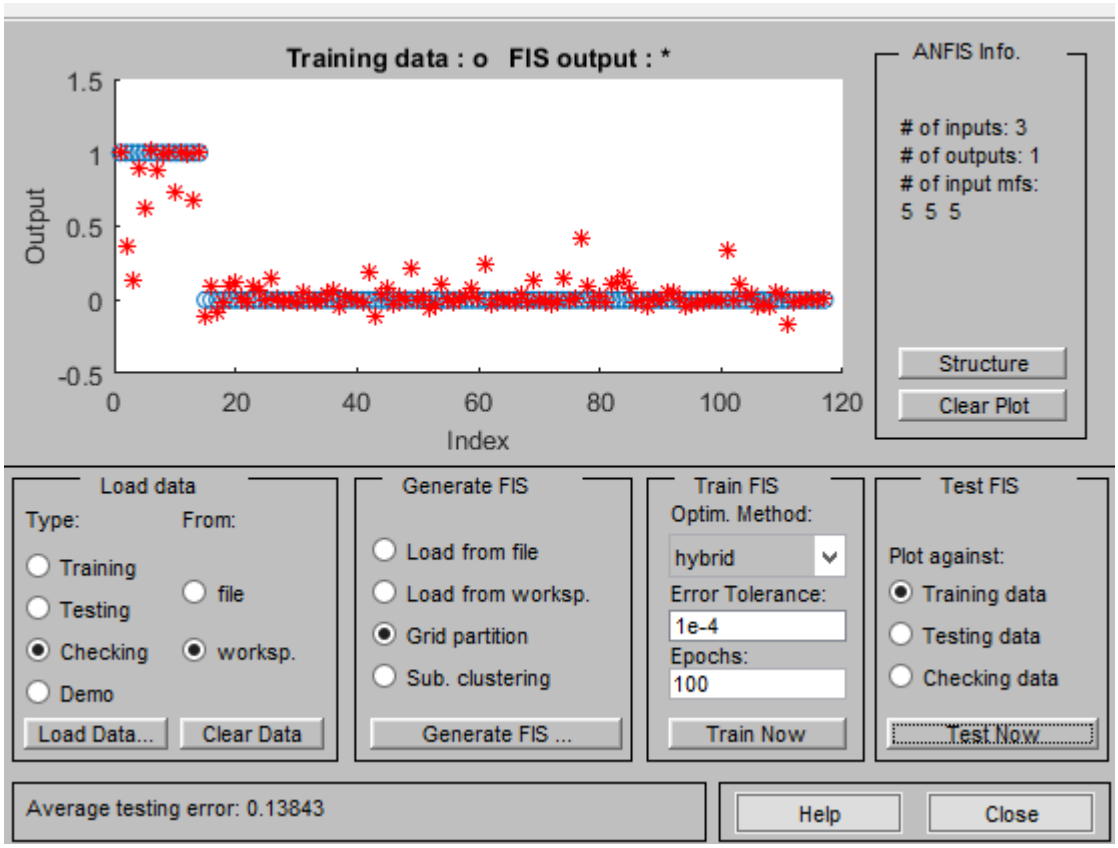

Figure 56: Thesis results of training ANFIS on Method 4 using the whole dataset. Training plot of ANFIS for Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM)with inputs the features of energy, homogeneity and kurtosis

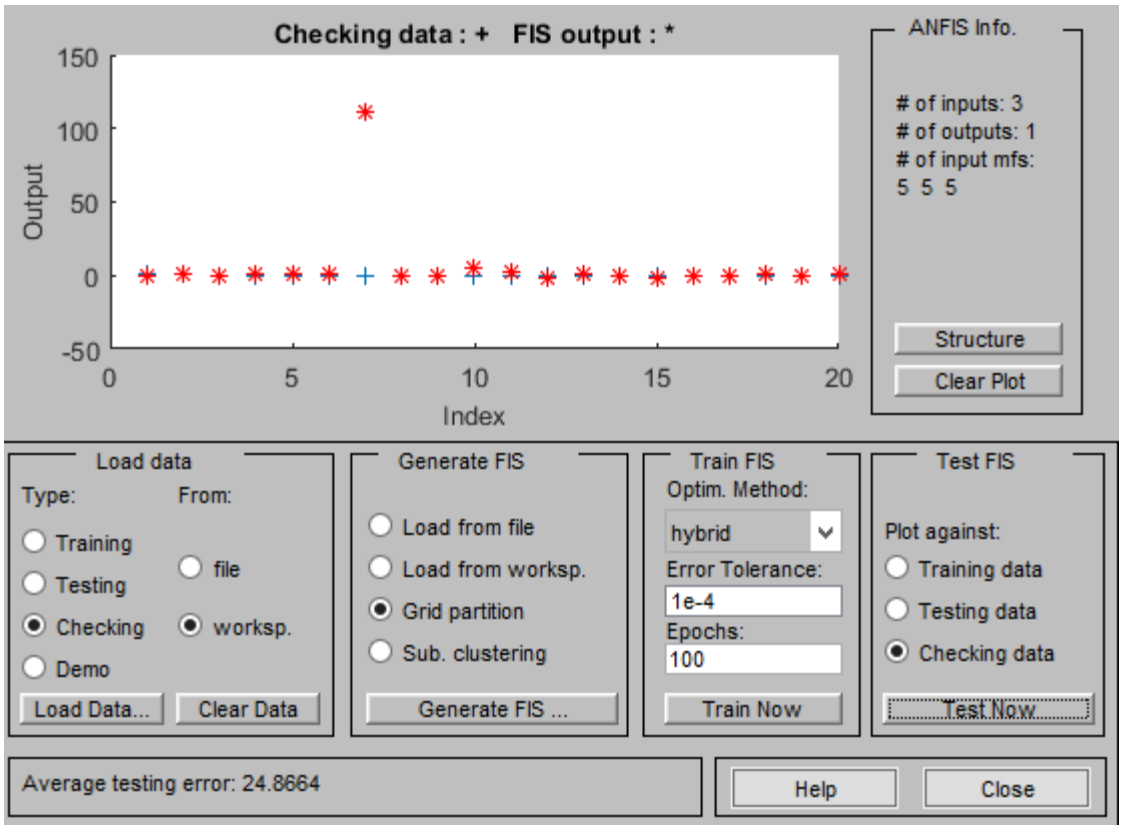

Figure 57: Thesis results of testing ANFIS on Method 4 using the whole dataset. Testing plot of ANFIS for Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity and kurtosis

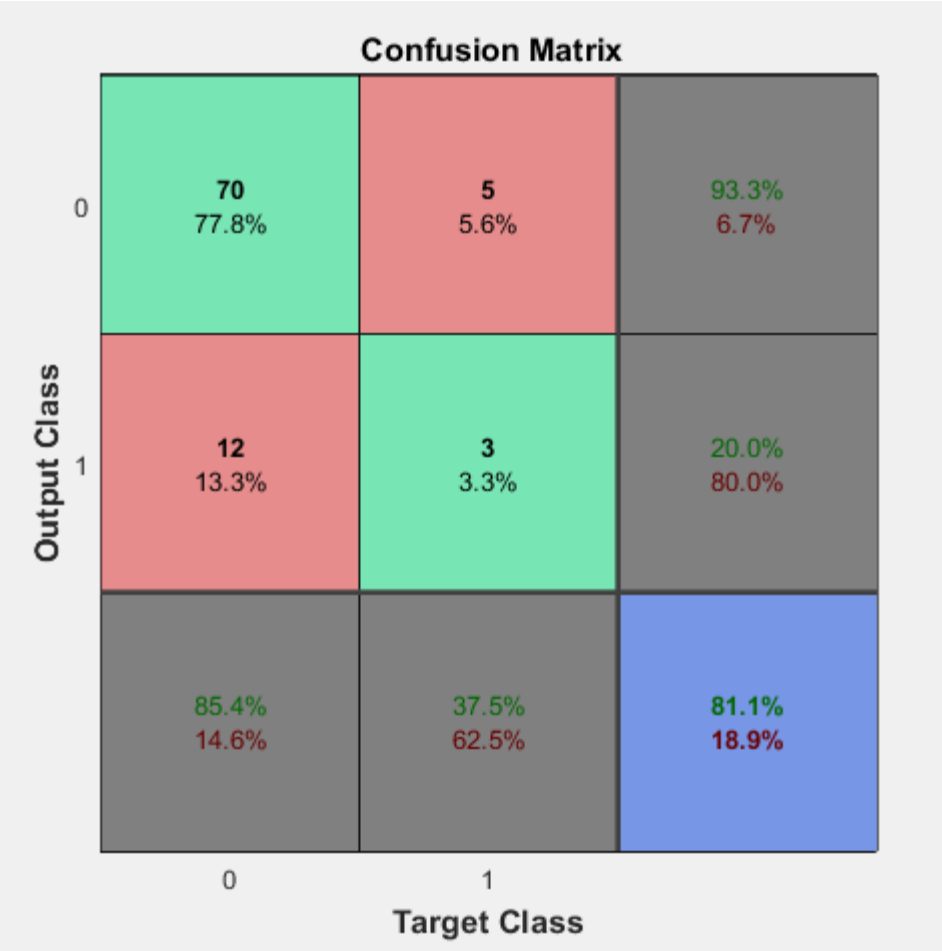

Figure 58: Thesis results of testing ANFIS on Method 4 using the whole dataset. Confusion matrix of testing data with ANFIS for Method 4 (Mean-Shift segmentation, Otsu's Thresholding, Morphological Operators, DWT, PCA, GLCM) with inputs the features of energy, homogeneity and kurtosis

Evaluation parameters	Subset of database with ANN	Whole database with ANN	Whole database with ANFIS
True positive	18	82	70
False positive	5	8	5
True negative	3	0	3
False negative	2	0	12
Sensitivity (%)	90	100	85.4
Specificity (%)	37.5	0	37.5
Accuracy (%)	75	91.1	81.1

Table 8: Thesis results. Comparison of classification results for Method 4 using different number of database cases

5.6 Classification results without pre-processing and ROI segmentation

For better evaluation of the pre-processing algorithms that are used we implemented neural network and ANFIS classification on the MRI without pre-processing and image enhancement. Of course the preparation stage of the original MRI is used, as well as the skull stripping algorithm, and after this the feature extraction is made by GLCM in the first case and DWT, PCA in the other. The results of these classification methods are presented below.

5.6.1 Results based on GLCM

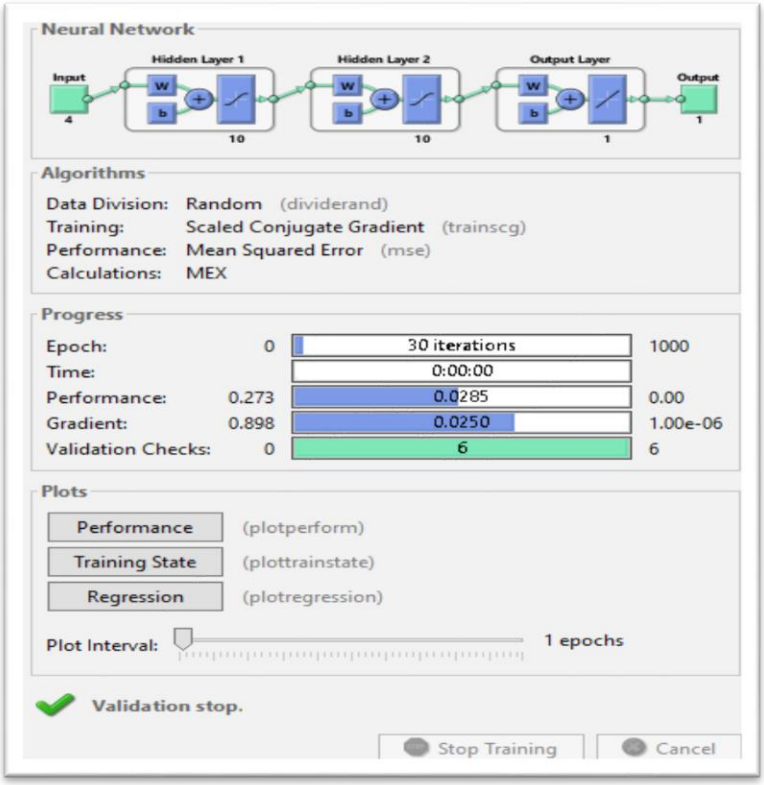

Figure 59: Thesis results based on GLCM. Architecture of backpropagation neural network

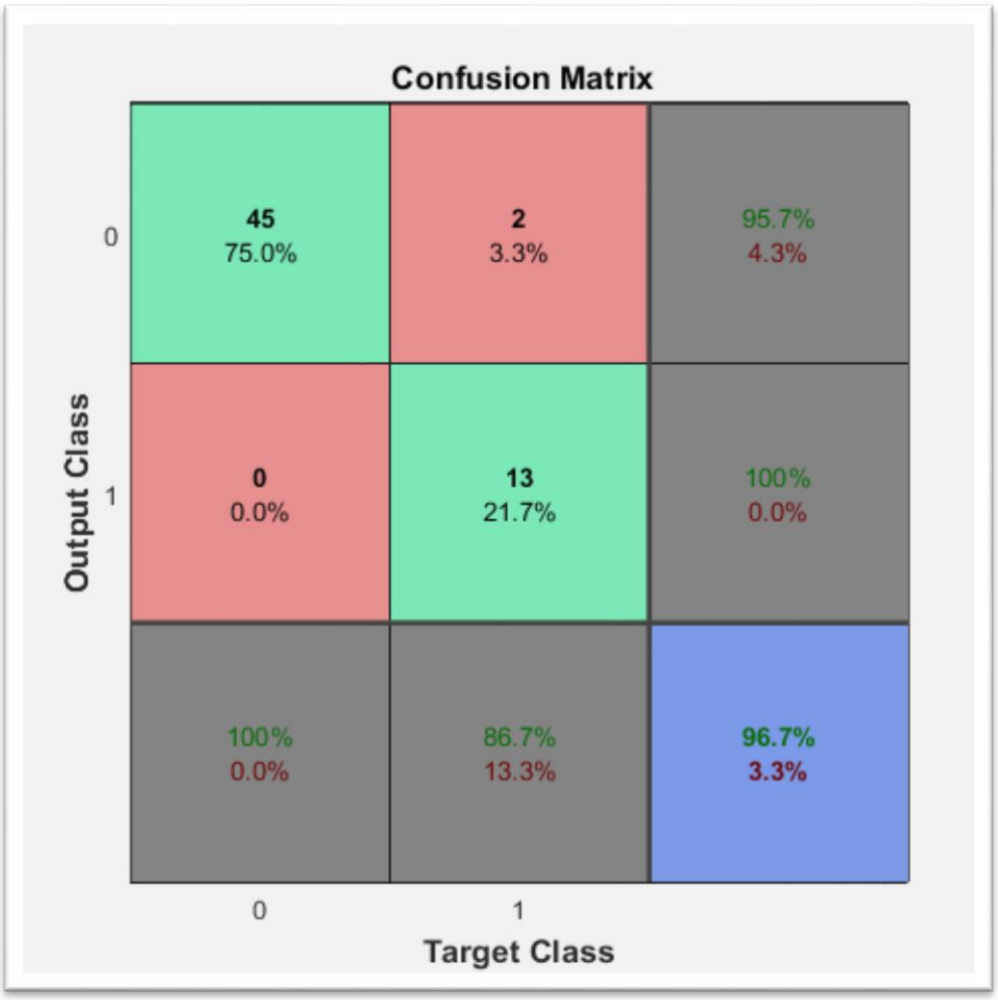

Figure 60: Thesis results of ANN based on GLCM using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity, skewness and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM

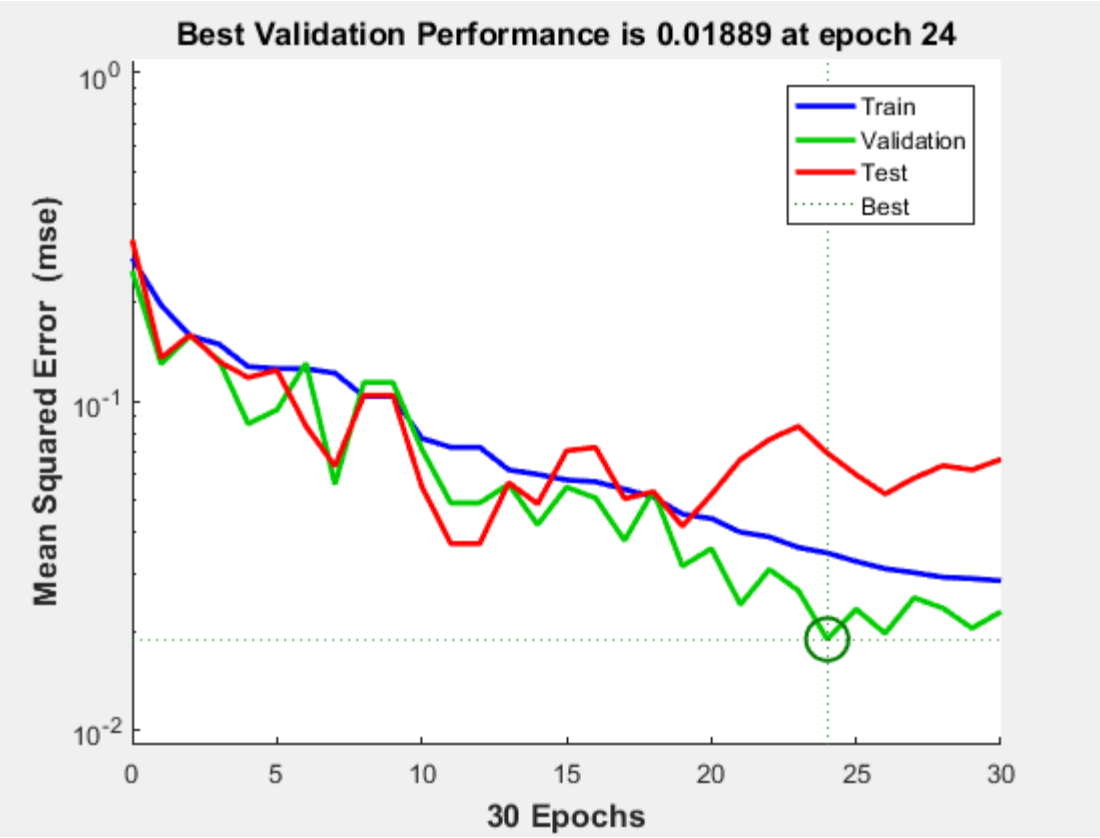

Figure 61: Thesis results of ANN based on GLCM using a subset of the database. Plot performance of neural network with inputs the features of energy, homogeneity, skewness and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM

The training error is decreasing while the validation and the testing error follow. After epoch 19 the testing error is starting to increase temporary while the training process stops at epoch 30, using the *early stopping* method and preventing the neural network to overfit the data.

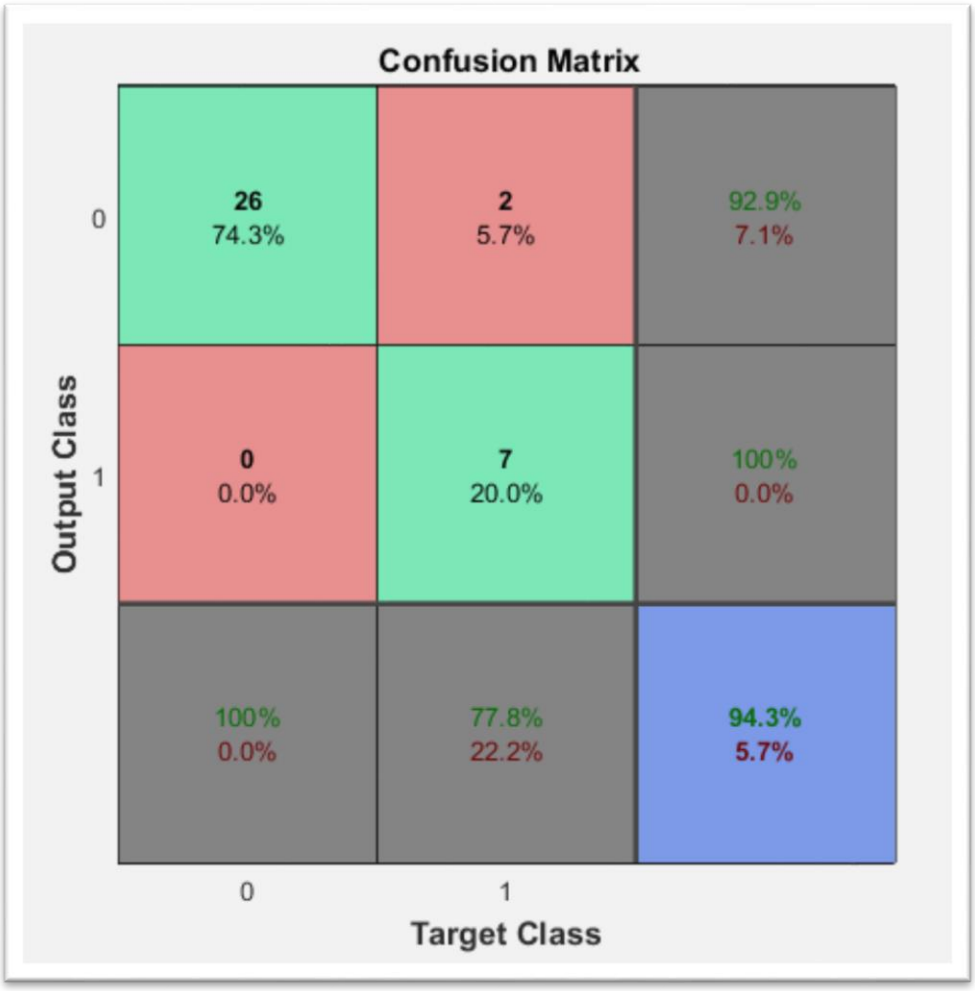

Figure 62: Thesis results of ANN based on GLCM using a subset of the database. Confusion matrix of testing data with inputs the features of energy, homogeneity, skewness and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM

Then, for classifying using the whole dataset a new neural network is designed with *Levenberg-Marquardt* backpropagation function and two hidden layers with 20 and 15 neurons respectively. The network has 3 inputs and one output

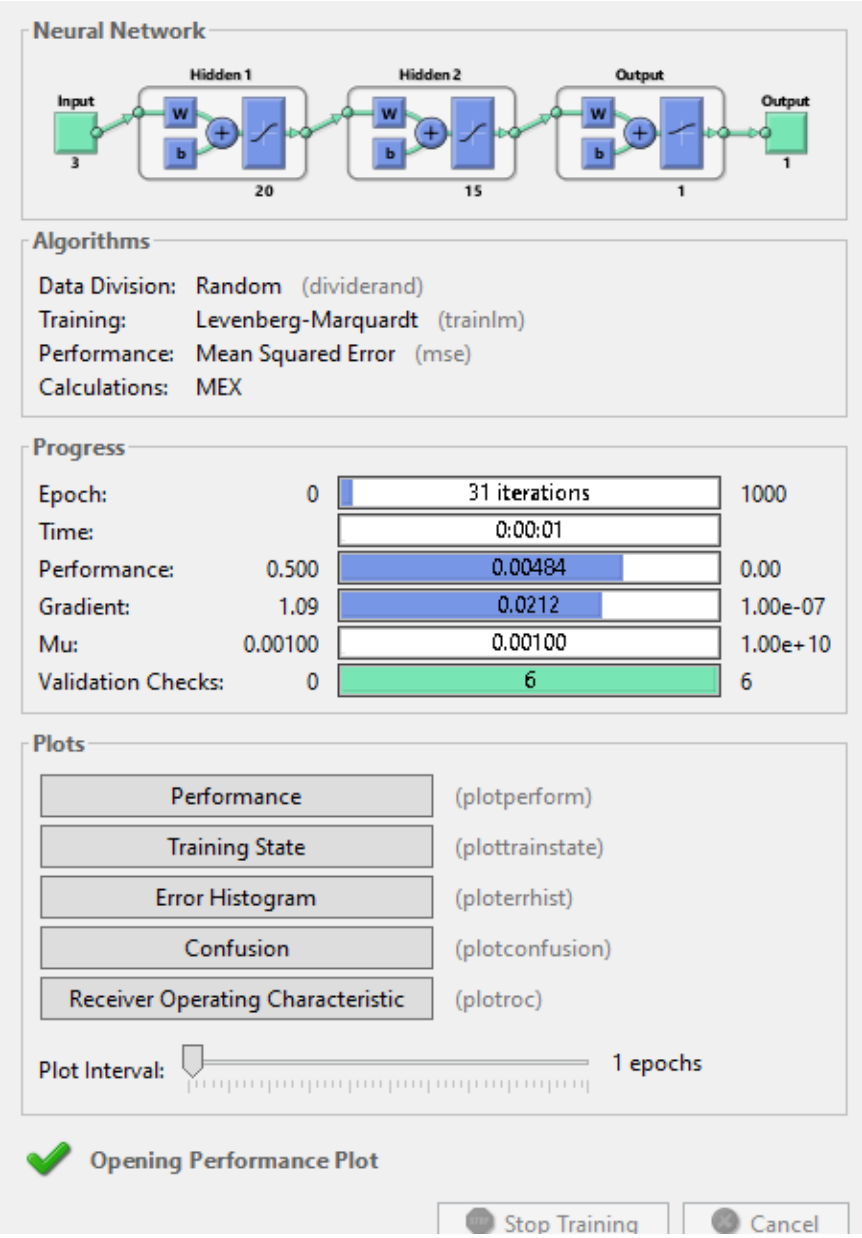

Figure 63: Thesis results of ANN based on GLCM using the whole dataset. Architecture of backpropagation neural network

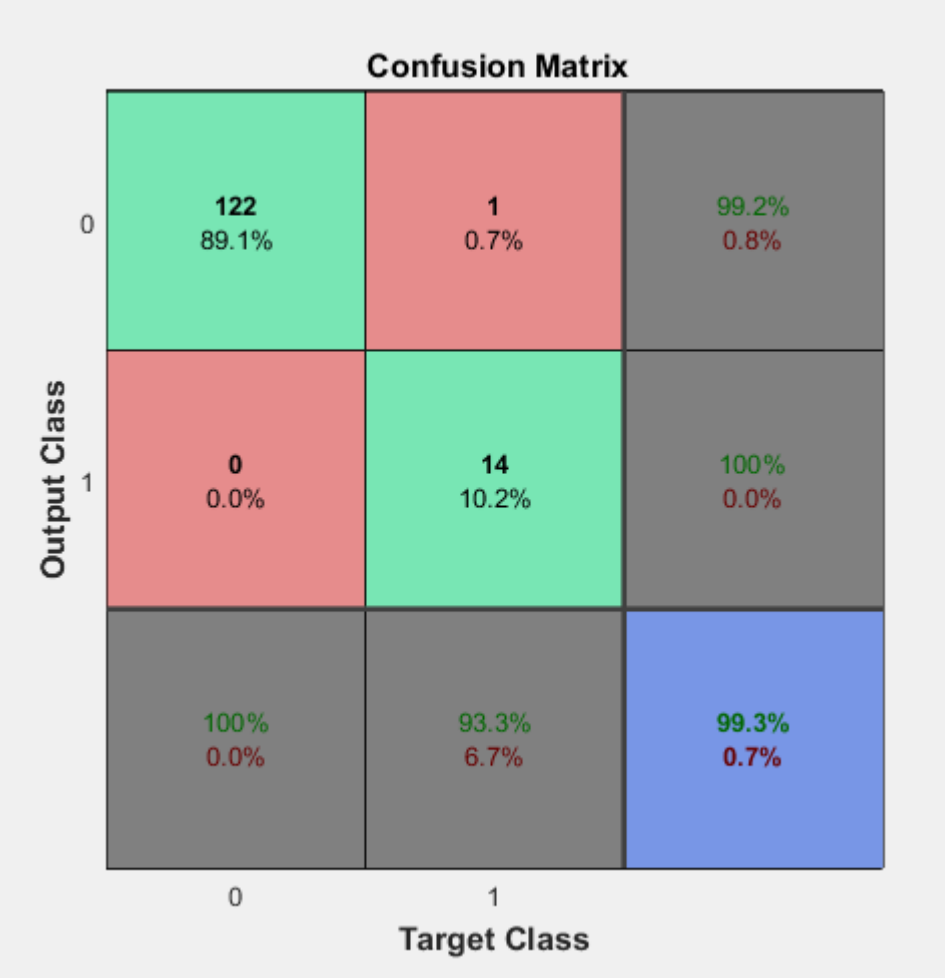

Figure 64: Thesis results of ANN based on GLCM using a subset of the database. Confusion matrix of training data with inputs the features of energy, homogeneity and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM

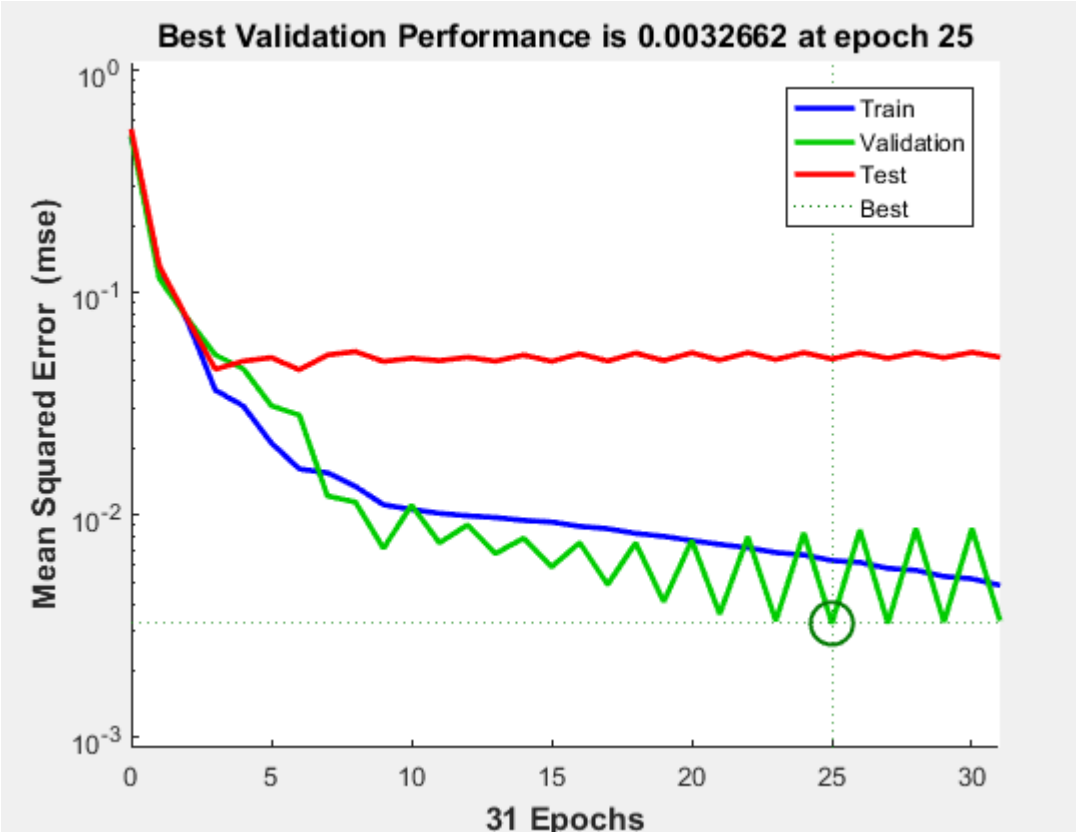

Figure 65: Thesis results of ANN based on GLCM using a subset of the database. Plot performance of neural network with inputs the features of energy, homogeneity and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM

Based on this performance plot we observe that the training error is decreasing while the validation follows. The testing error is stabilized while the training process continues. The training process stops at epoch 31, using the *early stopping* method and preventing the neural network to overfit the data. The best validation results are observed at epoch 25.

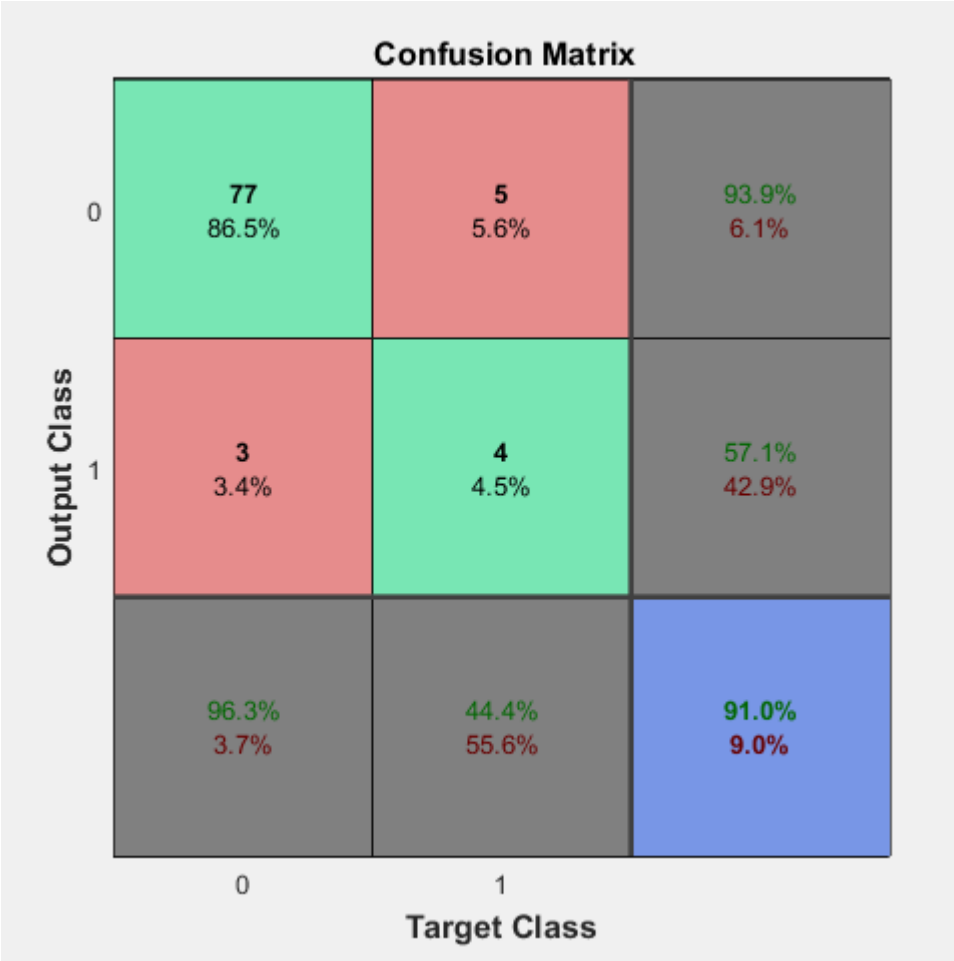

Figure 66: Thesis results of ANN based on GLCM using the whole dataset. Confusion matrix of testing data with inputs the features of energy, homogeneity and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM

Classification with ANFIS classifier is examined also for the case of GLCM extracted features of energy, homogeneity and kurtosis from MRI without pre-processing stage.

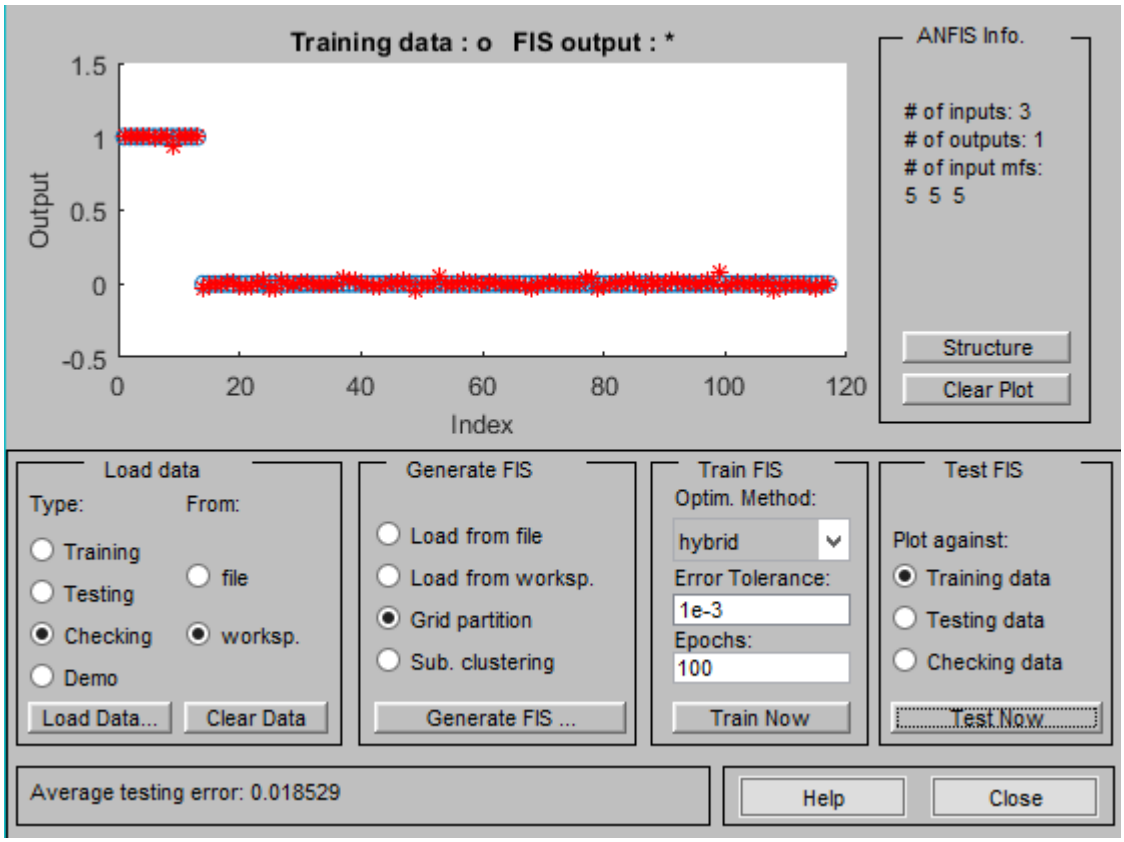

Figure 67: Thesis results of ANFIS based on GLCM using the whole dataset. Training plot of ANFIS with inputs the features of energy, homogeneity and kurtosis that were extracted with GLCM matrix from MRI without pre-processing and ROI segmentation

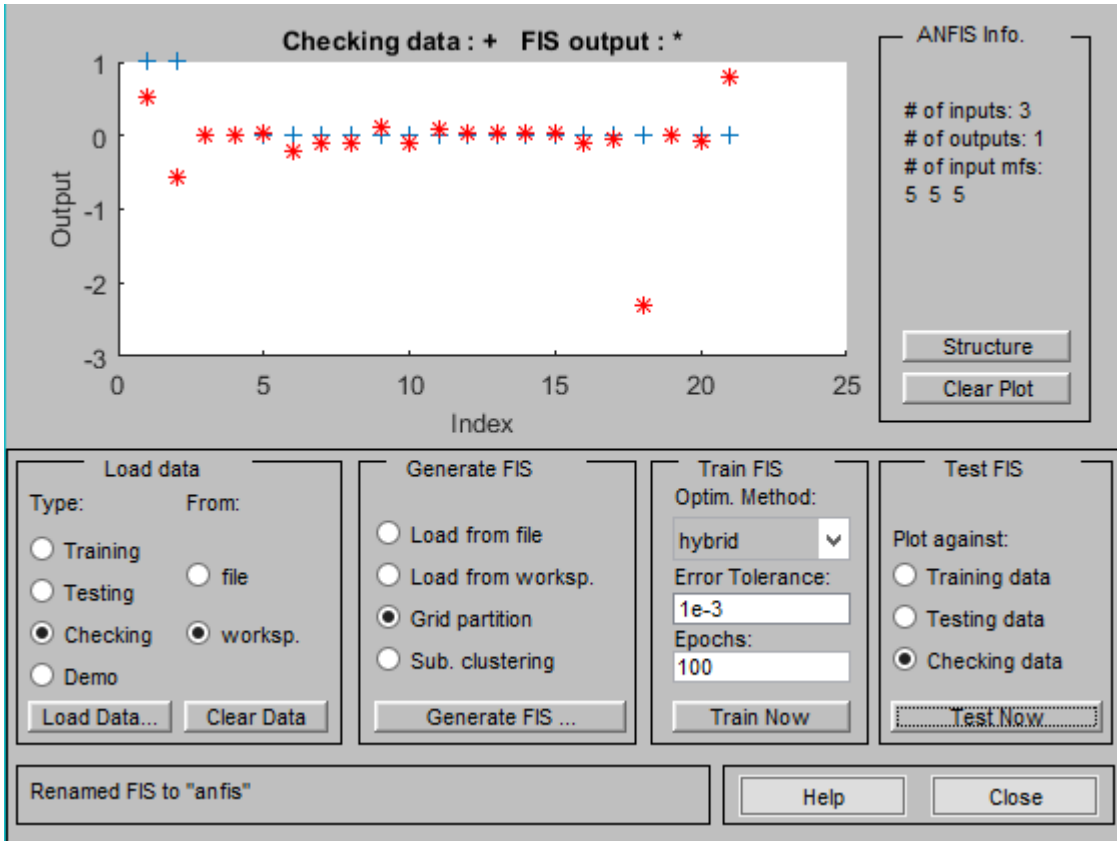

Figure 68: Thesis results of ANFIS based on GLCM using the whole dataset. Testing plot of ANFIS with inputs the features of energy, homogeneity and kurtosis that were extracted with GLCM matrix from MRI without pre-processing and ROI segmentation

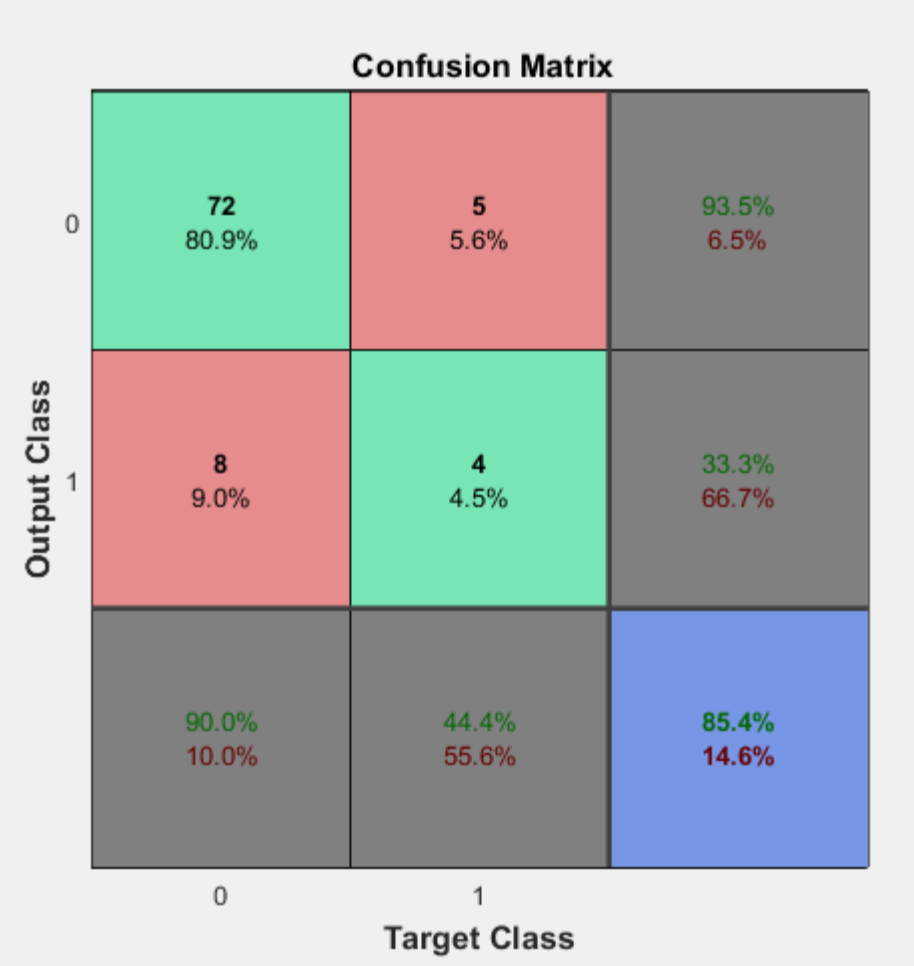

Figure 69: Thesis results of ANFIS based on GLCM using the whole dataset. Confusion matrix of testing data with ANFIS, with inputs the features of energy, homogeneity and kurtosis that have been extracted from the images without pre-processing and ROI segmentation, with GLCM

Evaluation parameters	Subset of database with ANN	Whole database with ANN	Whole database with ANFIS
True positive	25	77	72
False positive	2	5	5
True negative	7	4	4
False negative	0	3	8
Sensitivity (%)	100	96.3	90
Specificity (%)	77.8	44.4	44.4
Accuracy (%)	94.3	91	85.4

Table 9: Thesis results. Comparison of classification results from GLCM feature extraction without pre-processing and ROI segmentation

5.6.2 Results based on DWT and PCA

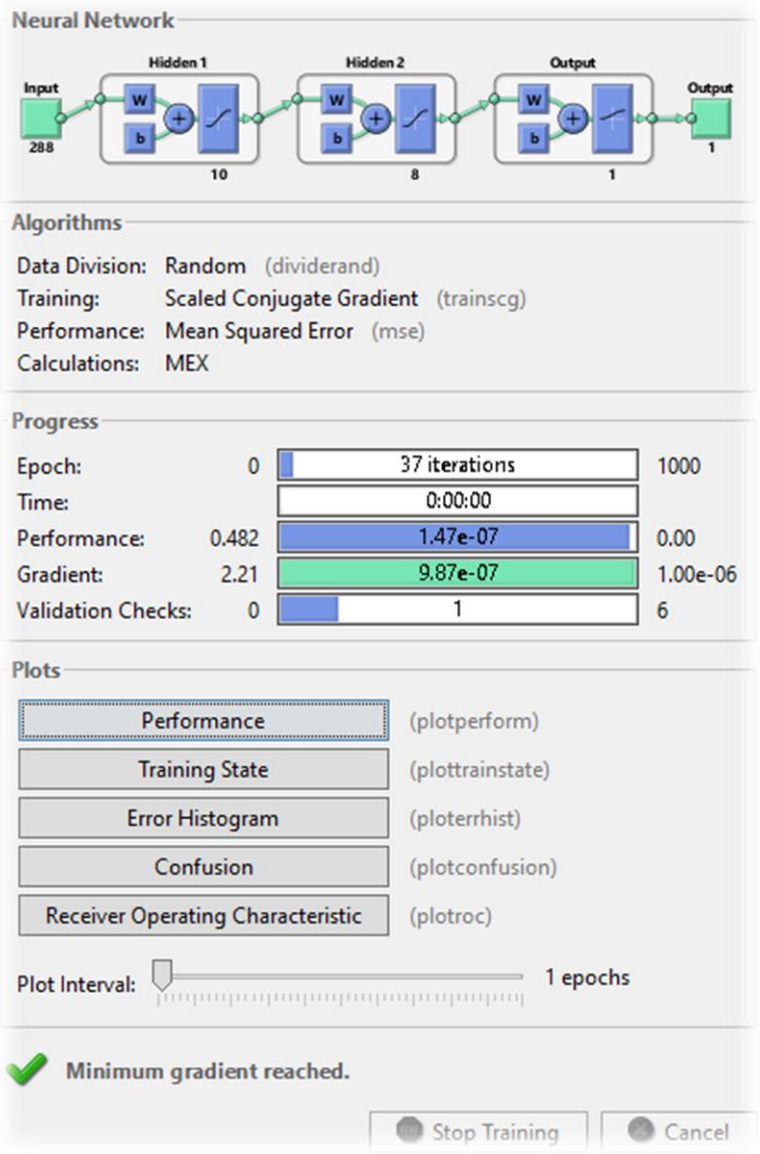

Figure 70: Theis rsults. Architecture of backpropagation neural network

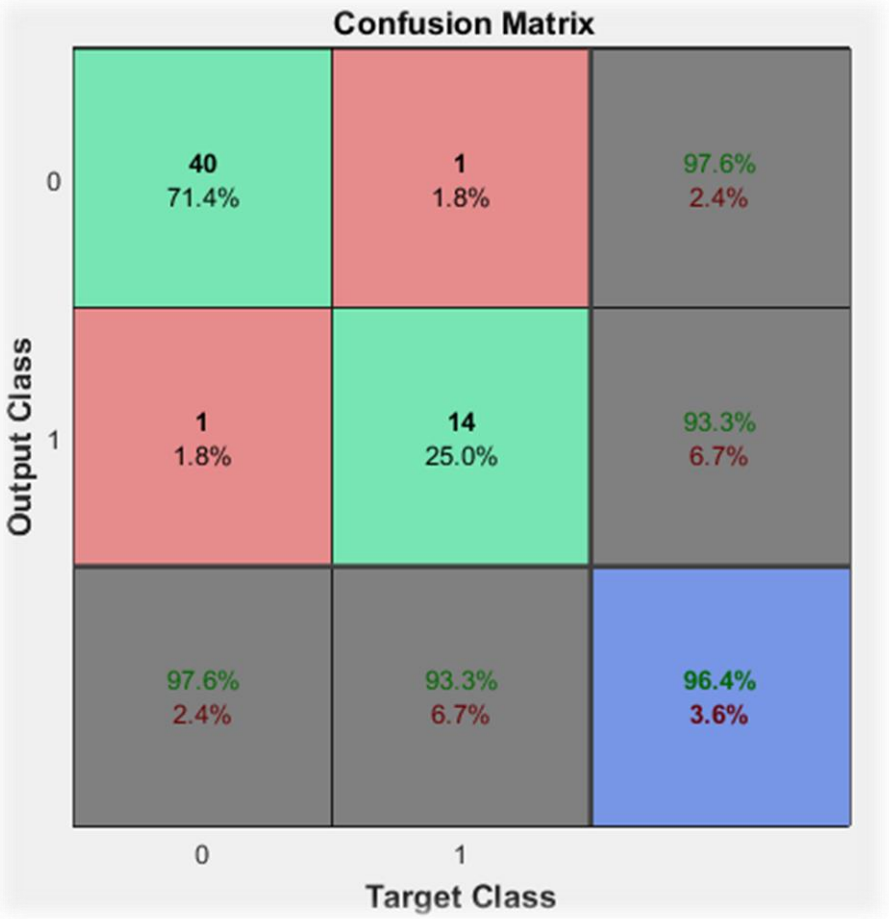

Figure 71: Thesis results of ANN based on PCA using a subset of the database. Confusion matrix of training data with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation

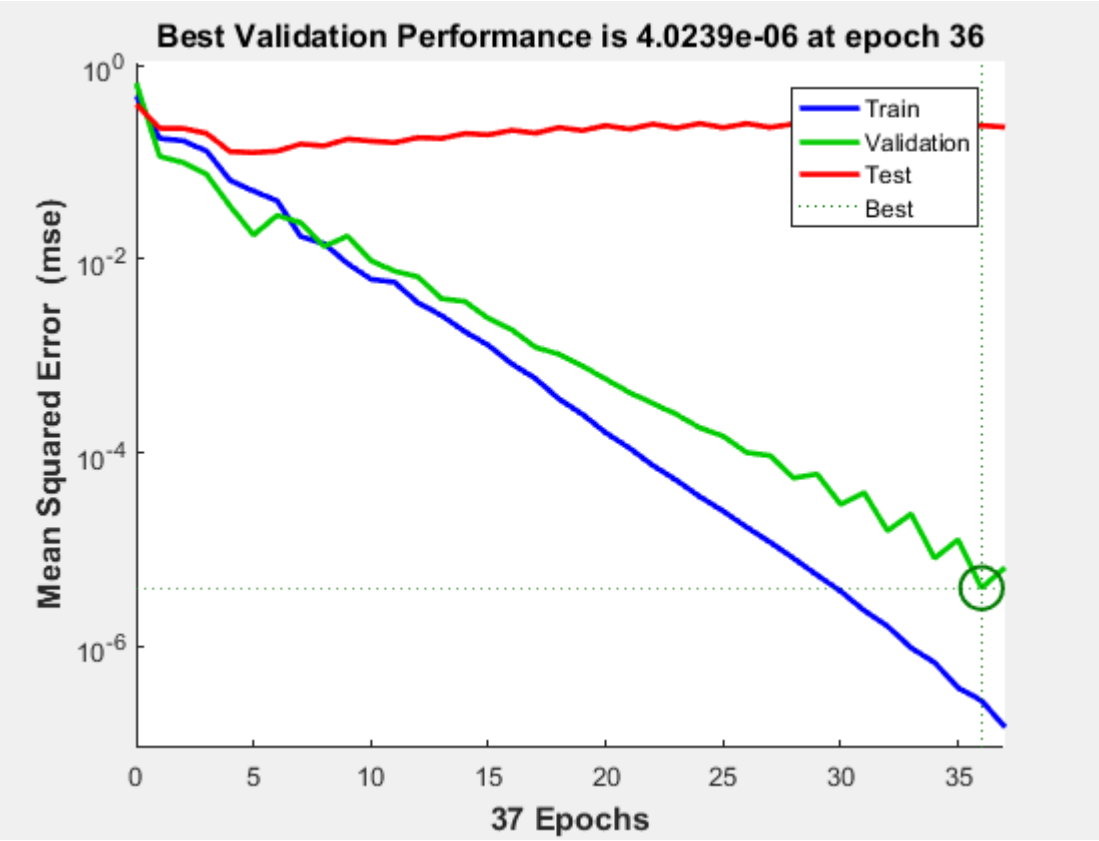

Figure 72: Thesis results of ANN based on PCA using a subset of the dataset. Plot performance of neural network with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation

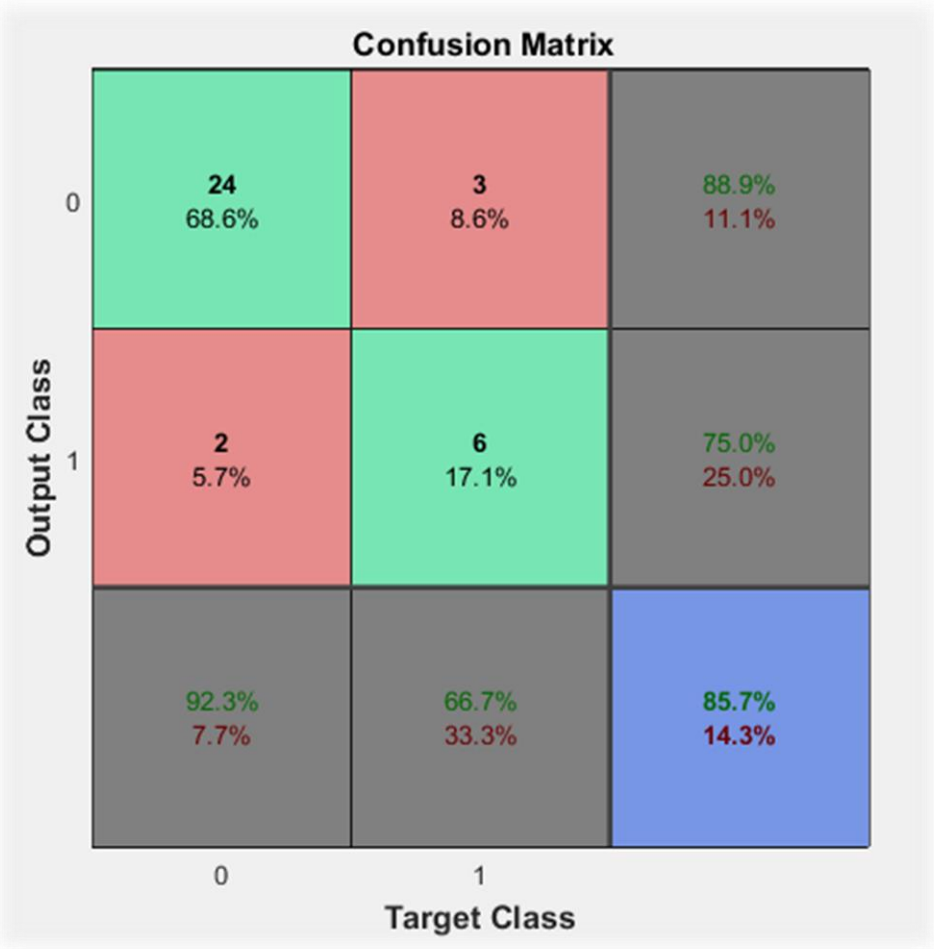

Figure 73: Thesis results of ANN based on PCA using a subset of the database. Confusion matrix of testing data with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation

Then, for classifying all cases from the whole dataset we designed a new neural network which is presented below.

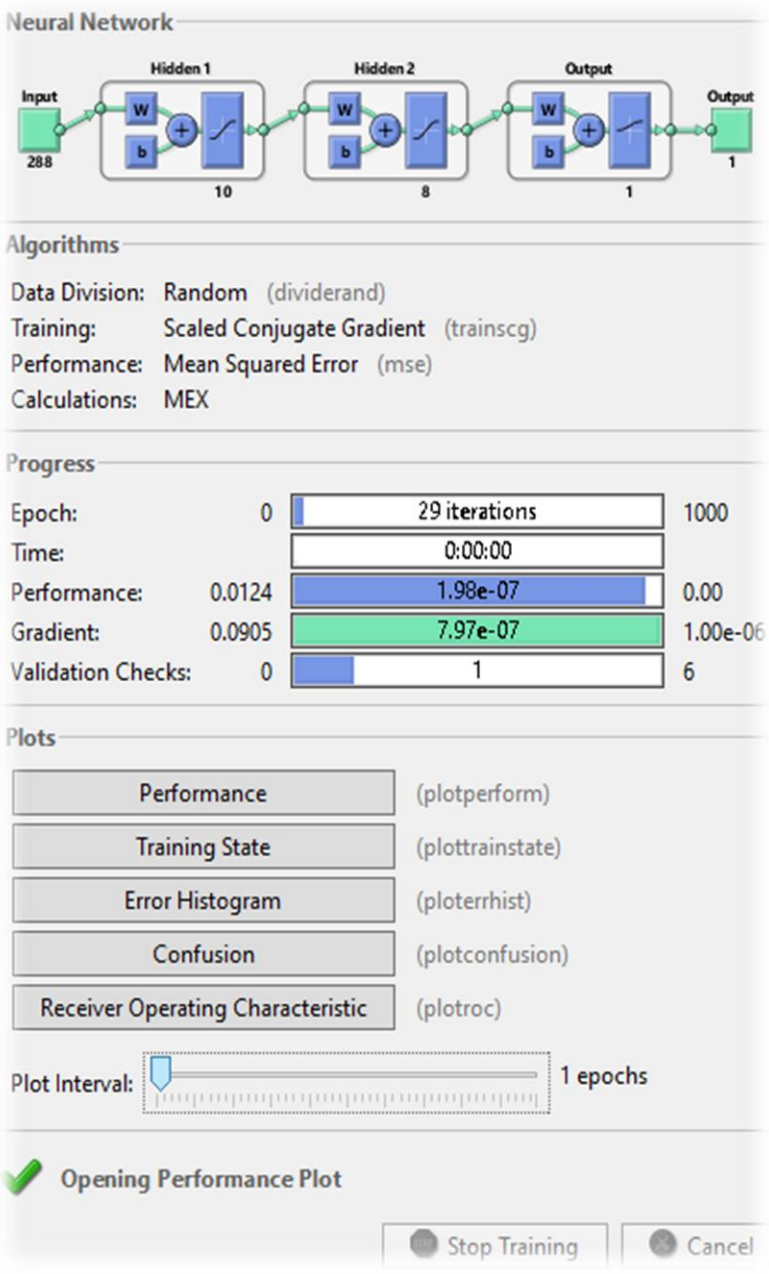

Figure 74: Thesis results. Architecture of backpropagation neural network

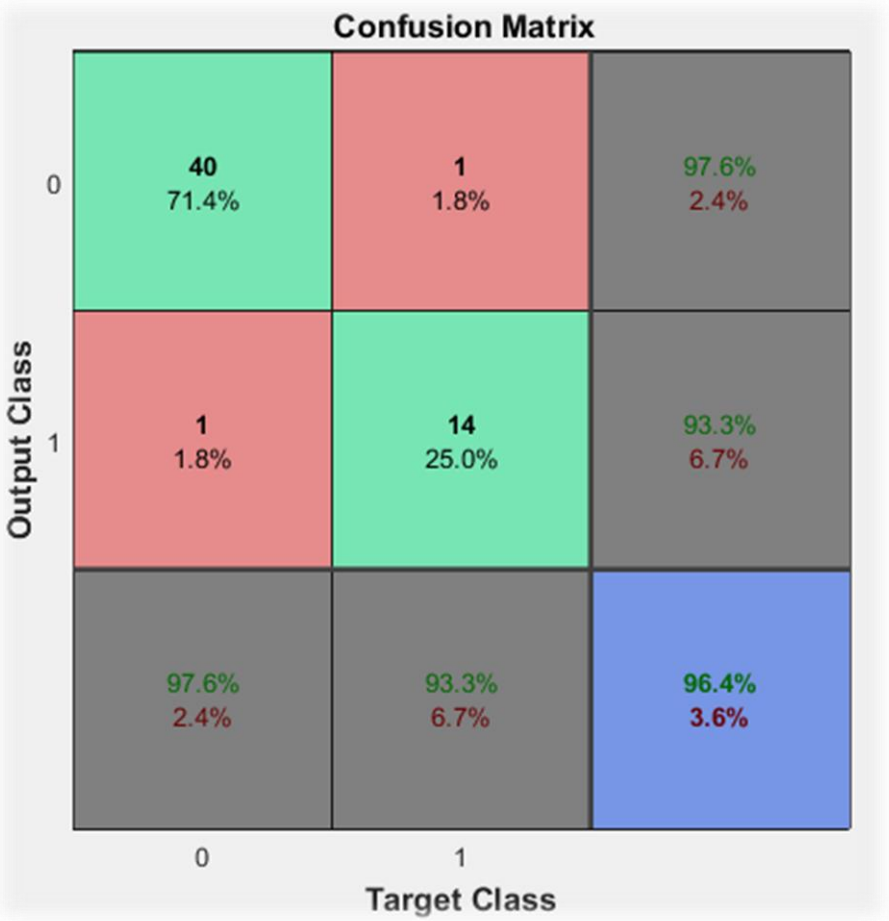

Figure 75: Thesis results of ANN based on PCA using the whole dataset. Confusion matrix of training data with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation

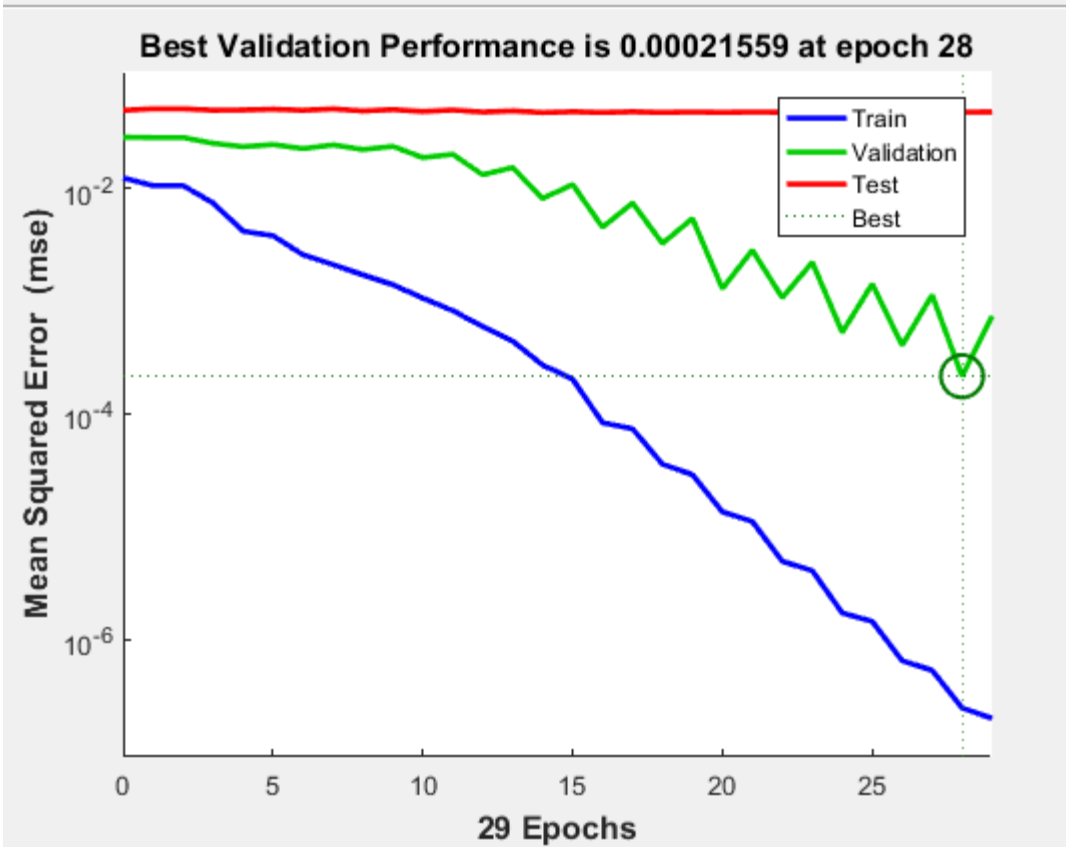

Figure 76: Thesis results of ANN based on PCA using the whole dataset. Plot performance of neural network with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation

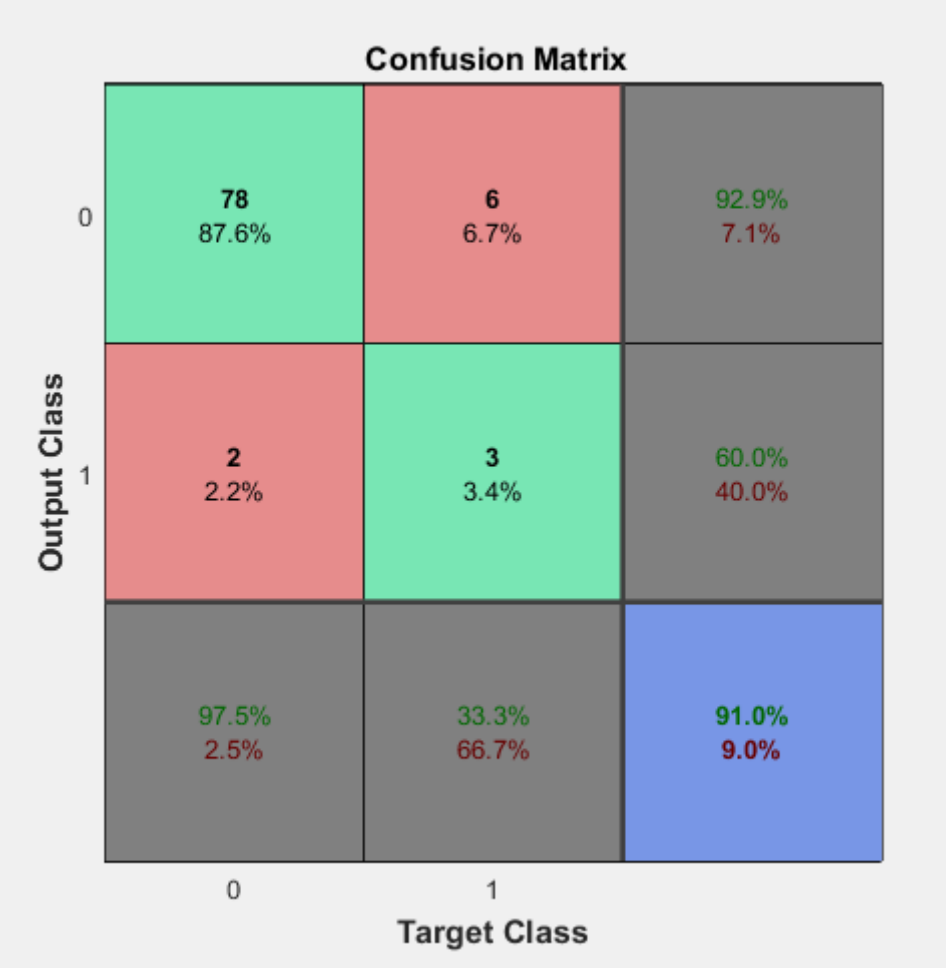

Figure 77: Thesis results of ANN based on PCA using the whole dataset. Confusion matrix of testing data with inputs the features extracted from the most significant principal components that have been extracted from the images without pre-processing and ROI segmentation

Evaluation parameters	Subset of database	Whole database
True positive	24	78
False positive	3	6
True negative	6	3
False negative	2	2
Sensitivity (%)	92.3	97.5
Specificity (%)	66.7	33.3
Accuracy (%)	85.7	91

Table 10: Thesis results. Comparison of ANN classification results from DWT and PCA feature extraction without pre-processing and ROI segmentation

CHAPTER 6. CONCLUSION, DISCUSSION AND RECOMMENDATIONS

6.1 Conclusions

This thesis aims to classify healthy from non-healthy patients based on MRI weighted T2 modality taken from axial plane. For this purpose several algorithms and methods are examined and implemented for pre-processing and post-processing stages. Furthermore, for the classification stage we examined two basic classifiers and presented our results in detail.

In the **Table 11** below we conclude the best classification results for each proposed CAD process Method we presented, for the dataset containing the whole amount of MRI from our database. The proposed algorithm based on our results is the one which combined pre-processing, the Method's 2 post-processing algorithms and the feed-forward backpropagation artificial neural network classification. This proposed algorithm uses high-pass and median filter for image enhancement and noise removal, Otsu's thresholding and morphological operators for ROI segmentation, DWT and PCA algorithms for feature extraction and dimensionality reduction and for classification, a feed-forward backpropagation artificial neural network. Finally, our proposed algorithm has been successfully tested and achieved the best results with accuracy 95.6%, specificity 62.5% and accuracy 95.6%.

Thesis proposed Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)
Method 1 (GLCM+ANFIS)	93.9	62.5	91.1
Method 2 (DWT+PCA+ANN)	**98.8**	**62.5**	**95.6**
Method 3 (DWT+PCA+GLCM+ANFIS)	89	25	83.3
Method 4(MeanShift+Method3+ANFIS)	85.4	37.5	81.1
GLCM+ANN (without pre-processing and ROI segmentation)	96.3	44.4	91
DWT+PCA+ANN (without pre-processing and ROI segmentation)	97.5	33.3	91

Table 11: Comparison of different classification results from the presented proposed Methods using the whole dataset

For the purpose of completeness, for our proposed algorithm we present below in Figure 78 the ROC curve of testing data for this algorithm.

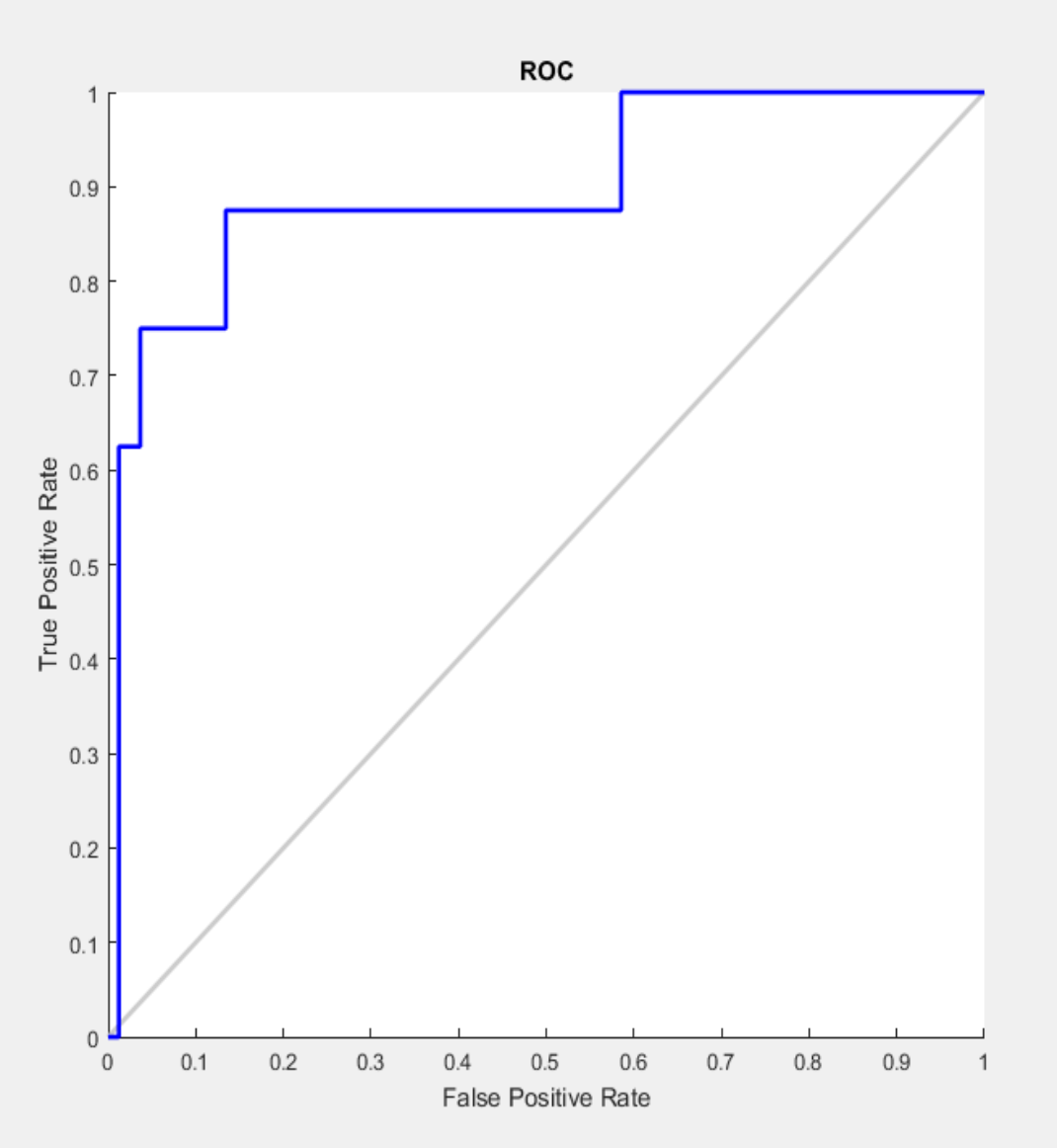

Figure 78: Thesis results. ROC curve of testing data of the proposed algorithm for classification of human brain MRI images using ANN clasifier

In **Table 12** below we compare our proposed algorithm with the algorithms designed in [13].

Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)
Method 2	98.8	62.5	95.6
BWT+GLCM+ANFIS	96.72	79.74	90.04%
BWT+GLCM+ANN	97.5	76.54	85.57%
BWT+GLCM+SVM	97.72	94.2	96.51%
BWT+GLCM+k-NN	93.33	77.77	87.06%

Table 12: Comparison of our proposed algorithm and the algorithms presented in [13].

6.2 Recommendations

There is no end to research and the same is applicable to this study also. The proposed algorithm of this study is designed to classify human brain MRI images and separate them to healthy and tumorous class.

The following recommendations are suggested:

1. First of all the proposed algorithm should be trained in a larger dataset containing more healthy MRI images and uses further methods such as *cross-validation* to deal with overfitting problems
2. Furthermore, the proposed methods could be developed to classify benign from malignant tumors, or recognizing and classifying the grade of malignant tumors.
3. The proposed algorithm could be implemented in a dataset containing other modalities also, like T1 weighted or FLAIR.
4. The proposed algorithm could be extended to different parts of body like lung cancer, stomach etc.

REFERENCES

[1] K. Bigos, A. Hariri and D. Weinberger , Neuroimaging Genetics: Principles and Practices, Oxford University Press, 2015.

[2] A. Lobera, "Imaging in Glioblastoma Multiforme," *Medscape,* 2017.

[3] S. P. M. Kumar and S. Chatteijee, "Computer aided diagnostic for cancer detection using MRI images of brain (Brain tumor detection and classification system)," *IEEE Annual India Conference (INDICON),* pp. 1-6, 2016.

[4] D. H. Miller, R. I. Grossman, S. C. Reingold and H. F. McFarland, "The role of magnetic resonance techniques in understanding and managing multiple sclerosis," *Oxford University Press,* vol. 121, pp. 3-24, 1998.

[5] [Online]. Available: http:/casemed.case.edu/clerkship/neurology/Web%20Neurorad/MRI%20Basics.

[6] H. Fujita, Y. Uchiyama, T. Nakagawa, D. Fukuoka, Y. Hatanaka, T. Hara, G. N. Lee, Y. Hayashi, Y. Ikedo, X. Gao and X. Zhou, "Computer-aided diagnosis: The

emerging of three CAD systems induced by Japanese health care needs," *Computer Methods and Programs in Biomedicine,* vol. 92, no. 3, pp. 238-248, 2008.

[7] N. Gordillo-Castillo, E. Montseny and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation," *ELSEVIER,* vol. 31, no. 8, p. 1426=1438, 2013.

[8] K. Fukunaga and L. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications," *IEEE Transactions on Information Theory,* vol. 21, no. 1, pp. 32-40, 1975.

[9] J. Shing and R. Jang, "ANFIS: Adaptive Network Based Fuzzy Inference System," *IEEE Transactions on Systems, Man and Cybernetics,* vol. 23, no. 3, 1993.

[10] R. C. Patil and A. S. Bhalchandra, "Brain Tumor Extraction from MRI Images Using MATLAB," *International Journal of Electronics, Communication & Soft Computing Science and Engineering,* vol. 2, no. 1, 2012.

[11] E. El-Dahshan, H. Mohsen, K. Revett and A. Salem, "Computer-aided diagnosis of human brain through MRI: A survey and a new algorithm," *Expert Systems with Application,* vol. 41, no. 11, pp. 5526-5545, 2014.

[12] Y. Zhang and L. Wu, "An MR brain images classifier via Principal Components Analysis and Kernel Support Vector Machine," *Progress in Electromagnetics Research*, vol. 130, pp. 369-388, 2012.

[13] N. Bahadure, A. Kumar and H. P. Thethi, "Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biological Inspired BWT and SVM," *International Journal of Biomedical Imaging ,* 2017.

[14] X. Zhou, Y. Zhang, G. Ji, J. Yang, Z. Dong, S. Wang, G. Zhang and P. Phillips, "Detection of abnormal MR brains based on wavelet entropy and feature selection," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 11, no. 3, pp. 364-373, 2016.

[15] A. Mayer and H. Greenspan, "An Adaptive Mean-Shift Framework for MRI Brain Segmentation," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1238-1250, 2009.

[16] S. Haykin, ”Neural Networks and Learning Machines”, Pearson Prentice Hall, 3rd Edition, 2009.

[17] Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Grey_matter.

[18] Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Cerebrospinal_fluid.

[19] Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Brain_tumor.

[20] Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/White_matter.

[21] D. Comaniciu and P. Meer, "Mean-Shift Analysis and Applications," *Department of Electrical and Computer Engineering Rutgers University*, 1999.

[22] S. Bauer, R. Wiest, L. Nolte and M. Reyes, "A Survey of MRI-based Medical Image Analysis for Brain Tumor Studies," *Physics in Medicine and Biology*, vol. 58, no. 13, 2013.

[23] Σ. Τζαφέστας, Υπολογιστική Νοημοσύνη, Τόμος Α: Μεθοδολογίες, Αθήνα, Ελλάς, 2012.

[24] I. Diaz, P. Boulanger, R. Greiner and A. Murtha, "A critical review of the effects of de-noising algorithms on MRI brain tumor segmentation," *IEEE Engineering in Medicine and Biology Society*, 2011.

[25] R. C. Gonzalez and R. E. Woods, Ψηφιακή Επεξεργασία Εικόνας, Pearson Education, 2008.

[26] M. Vidya and S. Dilip Kumar, "Study and Analysis of Various Pre Processing Techniques used in Breast Cancer Detection using Image Processing," *International*

Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 4, no. 5, pp. 495-499, 2016.

[27] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, pp. 62-66, 1979.

[28] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610-621, 1973.