ALZHEIMER DISEASE CLASSIFICATION USING NEURAL NETWORK

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

By JALALEDDIN MOHAMED

In Partial Fulfilment of the Requirements for the Degree of Master of Science in Electrical and Electronic Engineering

NICOSIA, 2017

ALZHEIMER DISEASE CLASSIFICATION USING NEURAL NETWORK

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

By JALALEDDIN MOHAMED

In Partial Fulfilment of the Requirements for the Degree of Master of Science in Electrical and Electronic Engineering

NICOSIA, 2017

JALALEDDIN MOHAMED: ALZHEIMER DISEASE CLASSIFICATION USING NEURAL NETWORK

Approval of Director of Graduate School of Applied Sciences

Prof. Dr. Nadire ÇAVUŞ

We certify this thesis is satisfactory for the award of the degree of Masters of Science in Electrical and Electronic Engineering

Examining Committee in Charge:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Jalaleddin Mohamed

Signature:

Date:

ACKNOWLEDGMENTS

I would like to express my sincere appreciation and thanks to my supervisors, Assistant Prof. Dr. Sertan Kaymak and Assistant Prof. Dr. Kamil Dimililer; for their guidance and mentorship during my graduate studies. Their impressive knowledge and creative thinking have been source of inspiration throughout this work. My deepest gratitude goes to my mother, my brothers, sisters, and the soul of my father who has passed away, to whom I am most indebted. I thank them for constant love, prayers, patience and support while I was studying abroad. I know I can never come close to returning their favour upon me. I will always be thankful to my friends and colleagues for their unlimited support. I extend my thanks to all the Libyan community that gave me a second family away from home. I also would like to thank the Libyan government for its support during my studies.

To my parents and family....

ABSTRACT

Alzheimer's disease is known as one of the most particular types of mental disorder for elder people. The first signs of the disease can appear in form of memory losses, cognitive and functional difficulties. The Alzheimer patient loses his ability to perform the simplest human tasks in the right way. The number of Alzheimer patients is increasing all over the world. The early detection of Alzheimer disease and its treatment can reduce its effects and symptoms. In this context, many researchers have focused on the implementation of different technologies for the detection and prediction of Alzheimer disease. Artificial neural network have proven their ability to perform well in extracting common patterns in medical images. They can be used for the detection of early Alzheimer symptoms in the brain MRI images. This work proposes the implementation of back propagation algorithm in the detection of Alzheimer disease in MRI brain images.

Keywords: Alzheimer; Artificial neural networks; back propagation; medical images; database

ÖZET

Alzheimer hastalığı, yaşlılarda görülen en yaygın hastalıklardan biri olan zeka bozukluğu hstalığıdır. Hastalığın ik belirtileri hafiza kaybı, anlama ve fonksiyonel zorluklar olarak kendiini gösterir. Alzheimer hastası en basit insan davranışını doğru olarak yapma kabiliyetini kaybeder. Dünyada alzheimer hastalarının sayısı gittikçe artmaktadır. Hastalığın erken teşhisi ve tedavisi etkilerini ve semptomlarını azaltabilir. Bu konuda, bir çok araştırmacı, alzheimer hastalığının belirlenmesi ve geleceği için farklı taknolojilerin uygulanmasına odaklanmışlardır. Yapay sinir ağlarının, tıbbi görüntülemede yaygın şekilleri çıkarmada oldukça başarılı oldukları görülmüştür. Bu ağlar, MRI görüntüleriyle beyindeki alzheimer semptomlarını erken belirlemede kullanılabilirler. Bu çalışmada, alzheimer hatalığının MR! Beyin görüntülenmesinde back propagation algoritmasının uygulanması önerilmektedir.

Anahtar Kelimeler: Alzheimer. Yapay sinir ağları, back propagation, tıbbi görüntüleme, veri tabanı

TABLE OF CONTENTS

ACKNOWLEDGMENTS	i
ABSTRACT	iii
ÖZET	iv
TABLE OF CONTENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	ix
CHAPTER 1: INTRODUCTION1	
1.1 Overview	1
1.2 Literature Review	2
1.3 Purpose of the Research	3
1.4 Methods and Plan	3
CHAPTER 2: ARTIFICIAL NEURAL NETWORK	
2.1 Introduction	6
2.2 Related Works	7
2.2.1 OASIS database	7
2.2.2 Other database	8
2.3 Historical Revision of Neural Networks	8
2.4 Structure of the Neuron	9
2.4.1 Biological neuron	9
2.4.2 Artificial neurons	10
2.4.3 Activation function in artificial neural network	13
2.4.4 Step activation function	14
2.4.4.1 Linear ramp transfer function	15
2.4.4.2 Sigmoid transfer function	16
2.4.5 Feed forward networks	
2.5 Training of Neural Networks	
2.5.1 Unsupervised learning of a network	19

2.5.2 Supervised learning neural network	19
2.6 Back Propagation Learning Algorithm	20
2.6.1 Formulation of the back propagation	21
CHAPTER 3: ALZHEIMER DISEASE AND IMAGE PROCESSING	
3.1 Alzheimer Disease	24
3.2 Magnetic Resonance Imaging MRI	24
3.3 Image Processing	24
3.3.1 Digital Image Representation	25
3.3.2 RGB image format or coloured image	
3.3.3 RGB to grey scale image	
3.3.4 Median filtering of the images	27
3.3.5 Image adjustment	
3.4 Work Structure	
CHAPTER 4: RESULTS AND DISCUSSIONS	
4.1 Introduction	
4.2 Processing Phase of MRI Images	
4.3 Part 1: Without Image Processing	
4.4 Part 2: With Image Processing	
4.4.1 ANN system using tangent sigmoid functions and	
4.4.2 ANN using logarithmic transfer function	
4.4.3 ANN using linear ramp transfer function	
4.5 Part 3: Two Equal Parts of Database with Image Processing	41
4.5.1 Training using tangent sigmoid functions	41
4.5.2 Training using ramp functions	42
4.5.3 Training of different parameters	43
4.6 Comparison with Previous Works	44
CHAPTER 5: CONCLUSIONS AND FUTURE WORKS	
REFERENCES	

LIST OF TABLES

Table 4.1: Parameters of the used neural network	32
Table 4.2: Results obtained from ANN (without DIP)	33
Table 4.3: Parameters of the ANN (with DIP)	34
Table 4.4: Output of the neural network (with DIP)	35
Table 4.5: Parameters of the system (with DIP)	36
Table 4.6: Results of the training (with DIP)	
Table 4.7: Output of the neural network (with DIP)	
Table 4.8 : Parameters of the system, linear ramp (with DIP)	39
Table 4.9 : Output results of the network (ramp function with DIP)	40
Table 4.10: Obtained results of test (tangent sigmoid with DIP)	41
Table 4.11: Ramp function training results with DIP	42
Table 4.12: Network results (ramp with DIP)	43
Table 4.13: Comparison of different training results	44
Table 4.14: Comparison of results of OASIS database	45
Table 4.15: Comparison of results of other database	45

LIST OF FIGURES

Figure 1.1: Samples of the used database for AD MRI and normal MRI	4
Figure 1.2: Flowchart of the classification work	5
Figure 2.1: Structure of the biological neurons	10
Figure 2.2: General model of the artificial neuron	11
Figure 2.3: Structure of the SLP neuron	12
Figure 2.4: Structure of multi layer perceptron	12
Figure 2.5: step function in ANN	15
Figure 2.6: The form of ramp function	16
Figure 2.7: General form of the sigmoid function	17
Figure 2.8: Feed forward neural network structure	18
Figure 2.9: Error back propagation algorithm	21
Figure 3.1: High resolution vs. low resolution image	26
Figure 3.2: Conversion of RGB image to gray image	27
Figure 3.3: Effect of applying median filter on noisy image	
Figure 3.4: Flowchart of the processing of images	
Figure 4.1: Sample data base of the training images	32
Figure 4.2: MSE curve of the training process	33
Figure 4.3: MATLAB interface of ANN training tool	34
Figure 4.4: Curve of the mean squared error during training	35
Figure 4.5: Curve of test data mean squared error	36
Figure 4.6: Mean Squared Error curve (Logarithmic sigmoid)	37
Figure 4.7: Curve of the training MSE	
Figure 4.8: Mean Squared Error curve (ramp function)	40
Figure 4.9: MSE curve during the training (tangent sigmoid)	42
Figure 4.10: MSE curve of the training network	43

LIST OF ABBREVIATIONS

AD: Alzheimer disease
ANN: Artificial Neural Networks
AI: Artificial Intelligence
BP: Back Propagation
DIP: Digital Image Processing
MRI: Magnetic Resonance Image
MLP: Multi-Layer Perceptron
NAD: Not Alzheimer Disease
OASIS: Open Access Series of Imaging Studies
RGB: Red, Green, and Blue
SLP: Single Layer Perceptron

CHAPTER 1

INTRODUCTION

1.1 Overview

Alzheimer's disease is known as one of the most particular types of mental disorder for elder people. This disease is defined by scarce presence in the brain of neuro-pathological structures. The first indications of the disease appear in the form of memory losses, cognitive and functional difficulties. The patient who is affected by the symptoms of the Alzheimer's disease suffers generally from difficulties in performing simple human tasks the right way. Alzheimer is classified as the first cause of brain malfunctions for elder people. The global number of disease affected people is in continuous increase and expected to go more than 100 million by the year 2050 (Katherine, 2012).

The disease abbreviated AD was actually named after the name of Alois Alzheimer who is a German scientist and physician. It has no treatment therapy in the medical point of view of modifying the disease structure; however, there are some treatments that can assist the patient to perform daily tasks in better way. Laboratory and clinical researches are pointed toward finding treatments that can reduce the risks of disease infection or to treat the actually infected people. Many other studies are pointed toward the study of different early signs of disease infection possibility.

As any other disease, early detection of the AD or expectation of its possibilities is a very important key in its treatment (Patil & Kale, 2016). It is very important to detect the early stages of the disease of Alzheimer for persons who are expected to suffer from it symptoms in order to be able to generate powerful treatments. Medically, different criterions are used for the early detection of the AD symptoms. One of the criterions is the NINCDS-ADRDA criteria (McKhann et al., 2011). This criterion is considered very useful in classifying patients according to their ability to have the AD. In order to give an exact judgment of the AD symptoms, brain tissues must be directly analyzed by specialist person. Medical researchers site that a successful therapy to modify the disease of AD can reduce the danger of having the disease. It sites that a delay of 12 months of the progression of the AD can reduce the number

of infected patients by about 10%. This shows that an early detection process of the disease acceptance is very important in reducing the global number of patients.

The artificial intelligence sciences are developing in a very fast rhythm in the last few decades. This development is due mainly to the huge development encountered in the processing devices in addition to the success of such artificial intelligence in solving complex problems faster than ever. Such ease of use, high performance, and low cost technologies are being more and more implemented in the development of auto systems that can perform different medical and industrial tasks.

Intelligent algorithms are being investigated widely in the medical sciences recently. They have shown a very important success in different fields and attracted researchers from different fields to more and more interest. Artificial neural networks have been in the core of these artificial intelligence systems. Artificial neural networks are parallel computing processors that use the structure of human brain to perform non linear functions and imitate biological thinking process. They can learn by examples and update their knowledge based on special training algorithms.

This work proposes the implementation of artificial neural networks for the classification of the brain MRI images and the detection of AD symptoms. The back propagation learning algorithm will be used in the training of the artificial neural network to perform the required classification tasks.

1.2 Literature Review

The early detection of Alzheimer's disease is one of the important research topics in the field of medical science. Different techniques have been implemented for the detection of the symptoms of the AD. Different tools were also used in the evaluation of the AD stage of the disease. These tools include blood, neurophysiologic signs, clinical tests, and magnetic resonance imaging techniques. The researchers in (Patil & Kale, 2016) were motivated to present their work on the early detection of AD. They presented a historical review of the disease and its features. Different image processing, enhancement, and segmentation techniques were studied and presented in this work. Finally, the use of artificial neural networks in the cognitive impairment and control subjects was proposed and studied. Gharaibeh & Kheshman, (2013) have presented their work on the subject of AD detection.

They proposed a fusion method to classify the MRI images of 27 Jordanian patients according to their AD medical evaluation. They claimed that they collected their database from the Jordanian hospitals during the preparation of their research. Artificial neural networks were used in the classification task of their research based on statistical analysis of the MRI images. Multi-scale fractional analysis of brain magnetic resonance images was proposed in (Lahmiri & Boukadoum, 2013). They use different image processing techniques to accomplish the classification task of AD detection. An automated method for the detection of Alzheimer disease based on structural MRI of brain images was proposed in (Eskildsen, Coupé, Fonov, & Collins, 2014). (Anirban & Kumar, 2015) have presented MATLAB software prepared specially for the presentation and study of MRI images and detection of AD possibilities. (You, 2007) has presented a collection of data base containing MRI images, voice, and other data for the use of AD detection in patients medical assessment. An artificial neural networks based classification system of tomography brain images was presented in (deFigueiredo et al., 1995). Different image processing techniques and ANN structures were studied and evaluated in this work. A special segmentation and image processing techniques was presented for the early detection of AD using MRI images was presented in (Wagner, 2000).

1.3 Purpose of the Research

The main purpose of this research is to develop an automated process that is able to support in the classification of MRI images and detection of Alzheimer's symptoms prior to its advanced stages. It is aimed to apply image processing techniques and artificial neural network on the brain MRI images to detect the special features in the brain images. These features are going to be used in the detection of AD disease in order to take special care of the patient to reduce the possible malfunction of the brain.

1.4 Methods and Plan

In order to be able to accomplish the purpose of this work, different steps must be carried out and assessed to evaluate the proposed methods. These steps include the collection of special MRI images for different subjects that are going to be used in the work. In order to accomplish this task, an extended search process in different universities and medical resources for specialized database was executed. This search has resulted in the collection of an acceptable amount of MRI images for two types of people. 243 images of AD infected people and 243 images on non-infected people were prepared and collected from the OASIS database (Buckner, 2012). Sample of the images is shown in the next Figure 0.1.



(a) Normal MRI

(b) AD MRI

Figure 0.1: Samples of the used database for AD MRI and normal MRI

These images are going to be carefully treated using different image processing techniques to extract special features and concentrate them before being used with the ANN. The images will be converted to small size images and filtered using simple median filter to remove noise. The result images will then be segmented using canny edge segmentation. They will then be divided into many images whose mean and standard deviation values are going to be calculated and added to the original images. Normalization of the image pixels is then going to be applied to improve the efficiency of the ANN during the training and test of database. The flowchart below of Figure 1.2 presents a brief description of the proposed method.



Figure 1.2: Flowchart of the classification work

CHAPTER 2

ARTIFICIAL NEURAL NETWORKS

2.1 Introduction

Artificial neural networks are nowadays very famous for being one of the most powerful and efficient artificial intelligence technology. They are used in many areas to accomplish a vast variety of tasks that include tracking, control, classification, recognition, identification and many other applications. Neural networks are of the important applications of the artificial intelligence that are very powerful. They offer simple solutions for complicated human tasks that were regarded as to be impossible to be automatically carried out.

The origins of neural networks can be found in the neuro-biological researches of the last century. They started to be spotted in the beginning of the 20th century. Biological specialists started to be attracted to the structure of the human nervous system. William James (1980) has a great insightful declaration that says: "The activity at any spot of the brain cortex is equal to the total tendencies sum of the other regions to discharge in that spot". These tendencies are related proportionally to the next factors (Mehrotra, Mohan, & Ranka, 2001):

- The number of times these regions were excited in accordance with the point.
- The density of the excitation
- The inexistence of any leaking points to which the discharges may happen.

Questions about the method in which nerves react for events, transformation of data through nerves, threshold of signals to be able to transmit and make effect were all question to be answered by biologists. Psychologists also were working to understand the processes of learning, thinking, person identification, and emotions are expressed and happen in humans and animals. Psychological and physical experiments in this field were very important and greatly enhanced our understanding of the way in which neurons accomplish their tasks (Hudson & Cohen, 1999; Mehrotra et al., 2001).

2.2 Related Works

Alzheimer disease is a neural muddle that causes dementia for elder people. This disease is characterized by scarce presence in the brain of neuro-pathological structures. The first indications of the disease appear in the form of memory losses, cognitive and functional difficulties. It causes difficulties in doing simple jobs like moving things or speaking normally. The numbers of infected people is increasing all over the world (Katherine R. Gray, 2012; Savio, Garcia-Sebastian, Hernandez, Grana, & Villanua, 2009).

2.2.1 OASIS database

Database images were collected from OASIS (Open Access Series of Imaging Studies) database. This database has cross sectional collection of 416 subjects or person. The age of these subjects is between 18 and 96 including those who have early Alzheimer disease (Savio et al., 2009). The used images contain multiple images for each study subject in different stages of life. OASIS database was used and implemented in different researches that are interested in Alzheimer disease. It was used by (Savio et al., 2009) in their paper to study the disease. They used 49 AD subjects and 49 normal subjects in their study. A segmentation and feature extraction process was implemented in this research with application of artificial neural network. Different structures of artificial neural networks were implemented in their study. The OASIS database was also used by (Saraswathi, Mahanand, Kloczkowski, Suresh, & Sundararajan, 2013). 198 persons' images were used in this paper and divided into three separate groups. 30 persons were classified as moderate, 98 persons were classified as normal, and 70 persons were classified as very mild persons. The paper proposed the use of feature extraction based on Voxel Morphometry method. This is a method used to identify changes in the gray matter between the brains of the normal and infected person. Extremely learn machine which is a single hidden layer network with back propagation algorithm was used. The work implemented genetic algorithms to extract 500 different features in the images that can be useful in the classification process of the network.

(Mahmood & Ghimire, 2013) in their paper used 457 OASIS database images for the training and test of their system. Normalized images were converted to vectors from where the average images were subtracted. The Eigen values of the result were computed and ordered in order to take the first 150 values that contain the important information according to the paper. Multilayer neural network was then applied on the database for training and test of results. (Long & Wyatt, 2010) has proposed an unsupervised method based on symmetric log-domain diffeomorphic algorithm for the classification of MRI images into normal and AD images. 416 OASIS images were used in this work.

2.2.2 Other database

Other researches also have studied the problem of AD detection using different database types. (Gharaibeh & Kheshman, 2013) have applied the artificial neural network on 37 different MRI images of brain. These images were collected from the Jordanian hospital. Morphological filters were applied on the images and they were converted to gray scale images. The standard deviation of images was also calculated to obtain statistic values of the images. Enhancement procedure was also applied on images to ameliorate the quality of detection. Obtained results were 100% performance in the training and 95% in the test of images. (Ramirez et al., 2009) in their research have used 97 different patients MRI images form "Virgen de las nieves" hospital in Spain. They used a random forest algorithm for feature extraction and selection from the MRI images. They reported good results out of their work. (Margarida Silveira, 2010) have presented their work using Alzheimer Disease Neuro-imaging Initiative database. They used 268 different subjects from the database in their work. A boosting classification algorithm was used in this research. The results obtained using this classifier were compared with those of support vector machine and gave better results.

2.3 Historical Revision of Neural Networks

The history of the neural network extends back to the early 1940s. Despite the fact it is not a very new idea, neural networks development wasn't that fast during the first two trimesters of the last century. It was not more than different works to develop a system that works similar to the real neural system. These works had a purpose to understand human's behaviours rather than the simulation of human nervous system. In 1943, McCulloch and Pitts presented their developed mathematical representation of the neural cell. Their model was insightful for many other works and has been used in various works. Different modifications were applied to this model to understand whether this model in its simple form is able to explain all mental processes of living beings. Before the end of the fifth decade of the 20th century, Hebb

introduced a learning rule that modifies the strength of connection between neurons based on defined criteria. Rosenblatt also presented another perceptron concept associated with a learning rule that uses the gradient descent theory to update. His rule used the terms "reward & punish" a weight based on how satisfactory the behaviour of the neuron is. The perceptron models were simple and easy to implement, however, they were not satisfactory for all types of processes. Between the years 1960 and 1962 Widrow and Hopf works showed that some classification tasks can't be easily accomplished based on the simple perceptron model. Minsky and Papert also reached to the same conclusion few years later (Hudson & Cohen, 1999).

The next two decades have witnessed a lot of research activities whose purpose is to overcome the weakness points of the simple perceptron model. Combinations of different neurons all together to accomplish a single task was thought to be powerful by different researchers. Special learning algorithms for multiple neuron networks were presented by Dreyfus, Bryson, and Wyrbos during the 60s and 70s. During the 80s of the 20th century, new developments in the neural networks aroused opening the doors for creating complex and efficient types of neural networks. Recently, a lot of researchers were encouraged to share in the development of the neural network technology. Nowadays, neural networks are widely used in different areas and applied in different sciences.

2.4 Structure of the Neuron

2.4.1 Biological neuron

The biological neuron is constructed of different parts that accomplish the different tasks required from the neuron. It consists of a cell body, an axon, and a multiple of dendrites. Dendrites are surrounding the cell body. The axon is a long tube split into different parts that are connected to the dendrites. They are responsible for the transmission of information between different cells. Each neuron of the biological neural network receives between 100 and 100000 synapses. Figure 2.1 presents the model of the biological neuron (Mehrotra et al., 2001).



Figure 2.1: Structure of the biological neurons

The thickness of the biological neuron is approximately 70 to 100 Angstroms. The dendrites receive signals from the neighbour neurons. The axon is a communication channel that sends these signals to the next neurons. The rest voltage of the neuron cell is -70mV. The cell keeps this voltage by forcing chemical elements out of it. The biological neuron is distinctive form other types of cells because it can be excited by an external source. When dendrites receive a signal, the cell becomes incapable of keeping the rest voltage at its set value. This results in a pulse that moves along the axon and affects the neighbour cells. The same will happen in these cells causing the transmission f the signal through the whole neural network (Hudson & Cohen, 1999).

2.4.2 Artificial neurons

The general model of the artificial neuron is shown in Figure 2.2. It this figure, a single neuron with multiple inputs is presented. The inputs are evaluated by being multiplied by weights of the neuron. The results are collected at the summing junction of the neuron where a transfer function is applied on these inputs to generate the corresponding output of the neuron. As seen from the figure, the transfer function is applied on the sum of the weighted inputs. There exist a variety of transfer functions in artificial neural networks such as linear and hard limit functions. These functions are key point in the efficiency of the neural network as it will be discussed later in this chapter and during the discussion of the results of this work.



Figure 2.2: General model of the artificial neuron

The simplest form of the neural network is known by the single layer perceptron. A single layer perceptron is created from one or more neurons connected in parallel. Each one of the single layer perceptron neurons provides an output of the network. Usually, the neuron in a SLP is connected to all input signals. The diagram of a single layer perceptron is presented in the Figure 2.3 below. It contains an output layer of single neuron in addition to m input neurons that are connected to the output neuron through the weights. In a single layer perceptron, one of the inputs is special and treated as fixed unit that is equal to one. This unit is called bias and it help too much in the stability of the SLP. The SLP is a form of supervised learning neural network structures (Kiran, 2009). These are trained or taught to generate targeted outputs for different types of inputs. Supervised type ANNs are especially designed to fit for the control and modeling of dynamically changing systems. They also can be used in the classification of different things, the prediction of future events, and many other simple and complicated tasks. The weights of a neural SLP are chosen arbitrarily in the first stage of the creation of the network. However, after the learning process is started, these weights are updated following a well established procedure to ensure the convergence of the required process.



Figure 2.3: Structure of the SLP neuron

The other type of neural network is known by the name multilayer perceptron. A MLP is a neural network that is created from different parallel connected neurons that construct layer. Different layers are also connected together in series to form the MLP.



Figure 2.4: Structure of multi-layer perceptron

Figure 2.4 presents the structure of a simple multilayer perceptron. Multilayer perceptrons are believed to be capable of modelling any input output relation correctly for real number combinations. In MLP networks, different learning techniques can be used to train the network. The back propagation algorithm has been one of the most spread and famous algorithms until recently. It is very efficient and capable of generating the suitable weights for the network based on the error between the outputs and their desired controversial. The error is being used as feed-back signal (the same way as in control systems). This feed-back signal is used to create a small variation signal of the weights. Continuous variation of the weights leads to the convergence of the neural network (Kiran, 2009). After the reduction of the error to a suitable value, the network is called learned network. This network will be able to generate outputs based on their experience built during the training.

A recently method of training neural networks appeared that is called deep learning. This method is said to be very efficient compared to the Back Propagation Artificial Neural networks (BPANN) in the fact that it guarantees faster convergence and better results. However, this learning algorithm can't be separated from the back propagation algorithm because it is based on it. The deep learning can be considered as an extension of the back propagation application for the training of ANN.

2.4.3 Activation function in artificial neural network

The neuron in an artificial neural network behaves like its counterpart in a biological system. It is considered as a processing system whose output is generated from the weighted some of all of its inputs. The output is fed to an activation function to decide the strength of this output. The use of an activation function transforms the signal level to a suitable representation from input to output. Activation function squashes the inputs and put it in a suitable range for the output. The range of the activation function is defined generally by the activation function itself; however, the selection of the activation function is sensible to the type of application and the form of the output signals. In some applications, it is preferred to fit the input output forms to suit a selected transfer function. This makes easy the application of the neural network and increases the speed and efficiency of the training process. In literature, there are various types of activation functions that can be used in the training of neural networks.

Sigmoid functions, linear functions, hard limits, and other types of functions are all examples of these types of activation functions.

2.4.4 Step activation function

As it is clear from its name, a step transfer function is a type of transfer function that looks like a step. It's one of the oldest and common types of transfer functions used in ANN. It is simply defined by the function:

$$f(net) = \begin{cases} a, net < c \\ b, net \ge c \end{cases}$$
(2.1)

The principle of this function is very simple; it passes either of two possible outputs based on the input level (Mehrotra et al., 2001). It is similar to a logic on-off switch. It is an easy to implement type of transfer function. It can be translated by the idea of having a minimum threshold in order to be able to trigger the system to generate a pulse or an output. If the sum of input signals is strong enough to trigger the output a, it is then generated; otherwise, the output b is generated to signify the existence of weak signal. Figure 2.5 represents the general curve of the step function. It changes between two different states when ever its input passes a well known threshold value. It is called step because it forms the shape of step when it changes. The x-axis represents "net" which is the total sum of the inputs of the transfer function. c is the threshold based on which the transfer function decides whether to trigger output a or b.



Figure 2.5: step function in ANN

Although the idea of threshold is natural and looks logical, this model of the neuron that uses the step function is irrelevant. If compared to a biological neuron structure, this structure ignores the levels of strength of the input signal. It is just working with ON and OFF notions which are not suitable enough for complex requirements of artificial neural networks. The magnitude variation of the inputs is not considered and also different output amplitudes are impossible to be considered based on the step function (Mehrotra et al., 2001). This problem can be explained easily if one considers a small variation of one of the inputs from $-\delta$ to δ . This small variation causes the connected output to flip from ON to OFF or vice versa. Such an output change is considered a point of weakness for the use of step transfer function.

2.4.4.1 Linear ramp transfer function

In this transfer function, the mathematical ramp is used to squash the input sum to a neuron in an output range. The ramp function is defined mathematically by:

$$f(net) = \begin{cases} a, net < c \\ b, net > d \\ a + \frac{(net - c)(b - a)}{d - c}, Otherwise \end{cases}$$
(2.2)



Figure 2.6: The form of ramp function

Unlike the step function, ramp function is a linear transfer function that has continuous curve. It also differs from the step function in the fact that its output can have multiple output levels with different input level combinations. Figure 2.6 presents the curve of a simple ramp function that is implemented in different neural network applications. The output of the ramp function is a saturation output that has upper and lower limits.

2.4.4.2 Sigmoid transfer function

This function type is one of the most common and famous transfer functions in the artificial neural network applications. It is famous mainly due to the simplicity of its function and the simple calculation of its derivative. For these two reasons, most of the neural network structures are implementing this type of transfer functions. This type of function is differentiable everywhere in the curve. This type of functions has saturation values defined by:

$$\lim_{\substack{net \to \infty}} f(net) = b$$

$$\lim_{net \to \infty} f(net) = a$$
(2.3)

The values of a and b can be generally set to 0, 1 and -1 based on the type of the sigmoid function used. In the artificial neural network applications, the tangent and logarithmic sigmoid functions are the most implemented. Figure 2.7 shows the general form of the sigmoid function. It shows that the function always converge to upper and lower limits when

the input signal is negative or positive high. The curve of such function is smooth and continuous as mentioned earlier.



Figure 2.7: General form of the sigmoid function

Sigmoid function can be mathematically represented using the next formula:

$$f(net) = z + \frac{1}{1 + e^{-c.net + y}}$$
(2.4)

Where; z is a bias value added to the transfer function, c and y are constant that controls the shape of the curve of the transfer function. Biological observant claim that their observations lead to a conclusion that the biological neurons firing rate behaviour is roughly sigmoid. The only drawback of the sigmoid functions resides in the heavy extensive calculation of exponential functions (Mehrotra et al., 2001). As said earlier, the sigmoid function is famous for the simple differentiation method of its curve. The derivative of a sigmoid function can be proven to be given by (Tantua, 2015):

$$f^{(net)} = f(net)(1 - f(net))$$
 (2.5)

This makes easier the process of error back propagation toward the previous layers during the training of the network.

The neural network has different structures that are suited for different applications. One can find the fully connected neural networks, the partially connected neural networks, radial basis networks, feed forward networks, and many other structures in literature. In this section, the structure of feed forward neural network is going to be studied and discussed.

2.4.5 Feed forward networks

Feed forward networks are a subclass of the acyclic networks. These are the networks in which the connection is done between each layer and the next layer. No connection or information passes from the layer to its precedent layer in a feed forward network. Such type of network is described using sequential indexing to identify each single node of each layer in the network. The network in Figure 2.8 is an example of this type of networks.



Figure 2.8: Feed forward neural network structure

2.5 Training of Neural Networks

Training or learning process of neural networks is the process in which a network is being taught the pattern in an input or group of input output combinations. This process can have different forms or notions based on the type of the network to be trained. Learning is defined

as a steady state of the behaviour after being change as a result of gaining experience. Neural networks store their knowledge in the weights of the layers. The values of the weights store information of the input output relation of a system or a process. These weights can be treated in matrix form for notion simplification. All the network weights are stored in matrix form for each layer and all mathematical operations are applied on these matrices. There are different methods implemented for the design and training of the neural network. These algorithms are all sharing the same purpose of leading the network to gain experience and learn some new information. The difference between these different learning algorithms resides in the way in which the learning is accomplished. Each type of the neural network has its own characteristics that make it suitable to be used with an algorithm and not suitable to be used with other algorithms. Mainly, the learning process of the neural network can be divided into two main categories. These are the supervised learning category and the unsupervised learning its experience through the use of previous examples for the process. On the other hand, the unsupervised learning doesn't require any examples during the training process.

2.5.1 Unsupervised learning of a network

In an unsupervised learning of a network, no need for teacher during the training is expressed. The training is done based on local information available in the system. This type of network is also known as a self organizational network because the network works autonomously and tries to detect the relation between the data fed to it. Main examples of this type of learning are the Hebbian learning algorithm in addition to the competitive learning structure. Unsupervised learning is also referred to by the name of online learning neural network. This is because the network learns while it is in process and there is no need to stop the process of the network during the learning.

2.5.2 Supervised learning neural network

The supervised learning process is referred to as learning by example. This process needs the existence of an external teacher to supervise the learning process. The main problem or concern of the supervised learning is the convergence of the network toward a desired target. The ANN is supplied with data about the inputs and targets of the process. The learning is then

acquired by updating the weights by applying suitable learning algorithm dependent on the error of the output of the process. The learning process continues until reaching a point of convergence of the system. The back propagation algorithm is one of the most important supervised training algorithms in the ANN applications.

2.6 Back propagation learning algorithm

The back propagation learning process of neural network implies the use of multiple steps in order to accomplish the training task. These steps start by the forward feeding of the systems inputs through the network, and passing by all the layers till finding system outputs. These outputs are then compared with the target to generate the error of the system. This error is the signal that to be back propagated following the same track until reaching the first layer of the network. This algorithm is an efficient and convergent learning algorithm although it consumes a lot of memory and processing during the learning. A back propagation trained artificial neural network is powerful enough to model accurately different types of functions (Gupta, 2006).

The back propagation training process is simple in its formulation as it is an application of the well known gradient descent algorithm. The gradient descent algorithm is an algorithm used to ensure the convergence of an error signal toward zero under all conditions. In the back propagation, the inputs of the system or the training data is shown successively to the input layer of the network, this, in turn reflects the inputs to the hidden layer with weights. At the end of the network, the final results are evaluated with the target group. An error signal is generated to be used in the update process of the weights. After the end of weights adjustments, the inputs are fed again to the neural network and a second iteration is applied. This process goes in a continuous mode until finding the best error results that meets the application requirements.



Figure 2.9: Error back propagation algorithm

There are two important parameters that are able to affect the training and the efficiency of the artificial neural networks. These are the learning rate and the momentum factor. These two factors are used in the weight updating during the training of the ANN.

2.6.1 Formulation of the back propagation

The back propagation algorithm invests in the gradient descent theory in the minimization of error. The minimization of error implies the calculation of the gradient of the error. For that reason the mean squared error is found and used as a training measurement tool. The need for calculation of the gradient implies that the transfer functions must be derivable all the time. The tangent and logarithmic sigmoid functions are of the most used types of transfer functions. The formula of the logarithmic sigmoid transfer function is given by:

$$o(x) = z + \frac{1}{1 + e^{-c \cdot x + y}}$$
(2.6)

Where; the variable z, c, and y are constant values that control the behavior of the squashing transfer function. The derivative of the sigmoid function of the previous equation is given by:

$$o'(x) = f(x)(1 - f(x))$$
(2.7)

In the first step of training a neural network, the feed forward pass process is applied. The total strength of the neuron can be given by:

$$TP = \sum x_n \omega_n + b \tag{2.8}$$

After finding the total potential of the network neurons, this potential should be squashed through a transfer function to generate the neuron output. In most of back propagation applications, linear transfer functions are used like sigmoid function. Another type of transfer functions is the tangent sigmoid function that is given by:

$$o(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$
(2.9)

The derivative of such function is written as:

$$o^{\vee}(x) = 1 - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2}$$
(2.10)

The output generated from the output layer's activation function is then considered as the neural network's output. At this stage the error of the training iteration can be calculated in order to be used in the back propagation. As mentioned earlier in this chapter, this error represents the deviation of the outputs from their targeted outputs. The next equation represents the error calculation formula where the training aim is to make this error as small as possible by doing successive iterations.

$$E = \sum (D - o)^2$$
 (2.11)

Where, D is the desired output. The derivative of the error function is then given by:

$$\Delta_{j} = (D_{j} - o_{j})o_{j}(1 - o_{j})$$
(2.12)

This error derivative is then spread back into the previous layer to update the weight values. The new weight value is then given by:

$$\omega_{jhnew} = \omega_{jhold} + \eta \Delta_j o_h + \alpha (\delta \omega_{jhold})$$
(2.13)

Where; ω_{jhnew} , ω_{jhold} are the new and old values of the weight j of the hidden layer h; μ is the learning rate; α is the momentum factor; o_h is the output of the hidden layer number h; Δ_j is the error of the weight j while $\delta \omega_{jhold}$ is the variation of the hidden weight. At each layer of the network, the error of the next layer is used to update the weights of the actual layer thus; for the hidden layers, the error is found by:

$$\Delta_h = o_h (1 - o_h) \sum \omega_{jh} \Delta_j \tag{2.14}$$

And the new weights are expressed by:

$$\omega_{hinew} = \omega_{hiold} + \mu \Delta_h o_i + \alpha (\delta \omega_{hiold})$$
(2.15)

Where; ω_{hinew} ; ω_{hiold} are the new and old weights of the first hidden layer; Δ_h is the error of the first hidden layer; and $\delta\omega_{hiold}$ is the variation of the hidden weight. and After updating all the network weights from the last layer back to the first layer, it is time to do another forward iteration of the inputs. The previous process is repeated at each iteration until an acceptable value of error is reached.

CHAPTER 3

ALZHEIMER DISEASE AND IMAGE PROCESSING

3.1 Alzheimer Disease

Alzheimer is a loss of the performance or abilities like thinking and logical reasoning. It is very dangerous in some case to the point that it can affect the health and daily life activities of a person. Alzheimer is also very dangerous as it cause the person to behave unusually in different cases and situations (Dahab, Ghoniemy, & Selim, 2012; Gharaibeh & Kheshman, 2013). Alzheimer is the basic reason for many mental troubles and dementia for elder people. It is considered very aggressive on the general mode and alters the intellectual behaviours of the person. The Alzheimer represents 3% for people whose age is between 65 and 74 years old. This percentage increases to 19% for people between 75 and 84 years old. The number jumps to 47% for those who are elder than 84 years old. This fact makes an important alert especially for countries that have higher number of elder people.

3.2 Magnetic Resonance Imaging MRI

MRI was first used in the 1977 to obtain two and three dimensional images of the body. It was developed to assist doctors in injuries diagnoses and detection. The MRI system produces strong and stable magnetic field in addition to smaller fields that allow the scan of different parts of the human body. The system emits a pulse that causes some molecules in the body to change place and direction. When these molecules stop moving, they emit their energy which will then be captured by the MRI system and translated to images. The created images can be very useful in detection of the dead cells in the brain and thus detecting the Alzheimer disease.

3.3 Image Processing

The image processing is used to describe all the processes and mathematical functions applied to the images. These mathematical functions are applied to adapt the construction or appearance of the digital images. Image processing is one of the applications of the digital signal processing techniques applied to two dimensional image arrays. There are many reasons for using image processing; it can be used for enhancing the visual appearance and the visualisation of the image when it becomes mess clear or noisy. In this case, the image processing is very useful to make images clearer and visibly more comfortable. The new images after treatment using image processing techniques will have better characteristics and appearance.

Image processing is also used to remove some types of additives that can appear in the images due to different factors like environment, image capturing devices, and many other sources of noise. It's also used to extract some useful information from the images that can be used later for different purposes. Whatever the purpose of image processing techniques; there is no doubt that it is one of the most important techniques that gave other dimension for our daily life. Our modern TVs, smart phones, cameras and all other devices are using the last image processing technologies to present us the best of their functions.

3.3.1 Digital Image Representation

The image is from the mathematical point of view a two or multidimensional matrix whose elements describe the density of light or colour at each point. All kinds of digital images are collections of pixels that define the light intensity in each small part of image. These small parts are known by the notion of pixel. The higher the number of pixels the better the resolution of that image is in an image. An image of 6000 pixels contains 6000 different light intensity spots within it while an image with 10000 pixels contains 10000 distinct values within it. Modern cameras are being evaluated based on the resolution or the number of pixels they can contain in their images. The video stream is nothing but a stream of images that are being presented successively with speed that can't be noticed with human eye. Figure 3.1 below explains better the difference between two identical images with only difference in the pixel resolution of the image. The figure shows that the pixel with 200*200 pixels resolution is clearer than the image that has 100*100 pixels although both are referring to the same image.



(b) 100*100 pixels

Figure 3.1: High resolution vs. low resolution image

Pixels in the digital imaging systems are the structural elements of the image. Mathematically, the pixel can be represented using binary system, 8bit system or 16 bit system. The range of values of an image pixel is a function of the bit resolution. In binary system the pixel range is between 0 and 1; in an 8bit system it is between 0 and 255; in the 16 bit system the range is from 0 to 65535. The 8 bit system is used widely in digital imaging for its small size and good resolution. The colour intensity in this system can have 256 different levels for each pixel in the image.

3.3.2 RGB image format or coloured image

RGB image is an image that has separate intensity matrix for each one of the three main colours (red, green, and blue). These three colours are characterized by their wavelength and frequency. Each one of these colours has a range of frequencies in the spectrum of visual light frequencies. The RGB image is visually more meaningful for human eye that differentiate colours and extract more information from coloured images. The computers are less affected by the colours of an image.

3.3.3 RGB to grey scale image

The gray scale image is an image that contains no idea about the light frequency in the image. The only information contained in the gray scale image is the intensity of white colour. The light intensity from white (0) to black (255) is presented in the image pixels. Gray scale images are used for computer aided imaging systems because they are more suitable and low cost for computer processing. For the computer, the most important is the brightness at each point of the image rather than the colour or spectrum of light densities. For that reason, all RGB images are being converted to gray scale images before applying the processing using computer software. This increases the processing speed without affecting the efficiency of the process. The gray scale level of each pixel in the image can be given in function of the three colours concentration by the next formula. The formula uses special weight for each one of the colours in order to consider the sensitivity of the eye. It was found that the best formula is the one given below as it ensures that the eye can identify the detail of the image correctly (Zollitch, 2016).

$$Gs = 0.587G + 0.299R + 0.114B \tag{3.1}$$

All image conversion is applied using MATLAB instructions that simplifies the image processing and makes it very easy to be carried out. Figure 3.2 below shows the visual difference between the coloured RGB image and the gray image after conversion.



(a) RGB image



(b) Gray image

Figure 3.2: Conversion of RGB image to gray image

3.3.4 Median filtering of the images

Median filter is a type of non linear statistic order filters that are based on a type of ranking of the elements of a kernel in an image. That means the elements contained in a window of the image that is defined by the user are arranged incrementally. One of the elements of the ordered vector is then chosen to replace the centre element of the window. In the median filter, this element is the median or the middle value of the vector. In some other types of filters, this element is chosen arbitrarily or sometimes the max and min values are used. Generally, Median filter like all types of filters are implemented to decrease or reject all types of noise from images. The most type of noise that can be treated using the Median is the impulsive noise that appears in the form of strange values in the middle of symmetrical or similar values. By replacing all pixels with their neighbouring Median values, all impulsive changes can be easily eliminated creating new more homogeneous image content (Gonzalez & Woods, 2001). The application of median filter on a noisy image is proven to be very useful in different cases like the one presented in Figure 3.3 below. The figure of green vegetables has become clearer than the noisy version after the application of Median filter.





(a) Noisy image(b) Filtered imageFigure 3.3: Effect of applying median filter on noisy image

3.3.5 Image adjustment

The image adjustment is a way of mapping the intensity image values of gray image and increasing the contrast of image. It is very well known by converting the image intensity values into another mapping of the intensity such that the contrast in the image is increased. The adjustment is done by scaling linearly the pixel values between two limits (generally the upper and lower 1% values). Pixels values that are lower or upper than these limits are saturated. The next formula is used to adjust a grey scale image intensity contrast:

$$O(i,j) = \begin{cases} 0; I(i,j) \le L_{in} \\ (I(i,j) - L_{in}) \frac{255}{H_{in} - L_{in}}; L_{in} < I(i,j) < H_{in} \\ 255; I(i,j) > H_{in} \end{cases}$$
(3.2)

Where;

O is the output or adjusted image; I is the input image that is to be adjusted.

 L_{in} , H_{in} are the pixel value that separates the lowest and highest 1% of pixels in the original image respectively

3.4 Work Structure

As it was explained earlier in the first chapter, the work of this thesis will be divided into two important parts. In the first part, the brain MRI images will be processed using simple image processing techniques. This processing is indispensable for the success of the proposed system. The image processing phase passes by different steps in order to simplify the application of neural network function. In this work, these steps were arranged in the next order:

- Reading the RGB image
- Converting the RGB image to gray scale image
- Application of median filter
- Adjusting the image to enhance contrast
- Resizing the images to small size that can be fed to ANN with minimal used computer memory
- Converting image to vectors
- Normalizing the images



Figure 3.4: Flowchart of the processing of images

After the end of all these steps of image processing, the database images are now ready to be used with the artificial neural network. The flowchart of the different steps of the image processing until the application of neural network on the system is presented in Figure 3.4 above.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter of the work is going to be sacrificed for the discussion of the applied techniques and the results obtained during the work. It will discuss the implementation of ANN for developing a classification system of brain MRI images based on the extracted features. The MRI images used in this system are going to be classified based on the neural network structure into Alzheimer Disease (AD) images or not Alzheimer disease (NAD) images. The AD images are those images in which the shape and size of the brain are classified as eligible to have or already have Alzheimer disease symptoms. The NAD images are the images that are classified as normal images whose shape and size of the brain are normal within the Circumference of the head.

All the 243 images of the MRI database are going to be processed in MATLAB and treated using different techniques before applying the neural network for classification. MATLAB was installed on a Core i7 laptop with 2.5 GHz processor speed and 8 GHz RAM speed.

4.2 Processing phase of MRI images

MRI images were all collected and separated into two folders named AD images and NAD images. It is important to mention that these images were all classified by the experience of medical experts with respect of the medical reports of all the patients. The processing phase of images included reading the coloured images of the brain MRI, converting them into gray scale images, filtering them using appropriate filter, extracting the edges and background of the images, image adjustment and size reduction of images. The final images are then going to be processed into normalized vector format that is the easiest way to be fed to the neural network. At the end of this phase the images are going to be subdivided into two parts to be used in the training and test of the network. Targets are going to be created for each individual image (in this case AD and NAD) so that the network is supposed to generate the desired

target whenever the image is provided. Figure 4.1 presents a sample of the training database used in the proposed neural network system.



Figure 4.1: Sample data base of the training images

4.3 Part 1: Without Image Processing

In this part, the artificial neural network will be applied on the original resized images without applying any image processing techniques to evaluate the effect of image processing techniques on the performance of the classification process.

Parameter	Value	Parameter	Value
No. of layers	4	Learning rate	0.2
Input transfer function	No	Momentum factor	0.5
Output transfer function	tangent	Layer sizes	2500, 50, 100, 2
Hidden1 transfer function	tangent	Training images	340
Hidden2 transfer function	tangent	Test images	146

 Table 4.1: Parameters of the used neural network

The parameters presented in Table 4.1 were applied during the training of the network and the results were obtained and evaluated.

After training the network for more than 276 seconds and during 5223 iterations, the mean squared error reached a value of 0.0650 which was enough for the generalization of the network. The MSE curve is presented in the next figure, it is decreasing slowly but the final MSE is acceptable. After checking the number of correctly classified images, it was found that 93% of the training images were recognized correctly. 80.25% of the test images were also classified correctly in the results of this experiment. The MSE curve of the training is presented in the Figure 4.2 below while the results of the network are tabulated in the Table 4.2. The table shows that the difference between the real output and the desired output is considerable although it is acceptable.



Figure 4.2: MSE curve of the training process

	NAD	AD	NAD	AD	NAD	AD	NAD
0,658	0,187	0,633	0,239	0,774	0,129	0,532	0,216
0,331	0,822	0,356	0,772	0,233	0,872	0,467	0,788
0,536	0,272	0,663	0,063	0,668	0,073	0,670	0,062
0,437	0,739	0,319	0,933	0,337	0,933	0,320	0,942
0,854	0,091	0,756	0,135	0,598	0,294	0,933	0,446
0,141	0,907	0,254	0,858	0,405	0,732	0,072	0,561
0,929	0,076	0,855	0,900	0,497	0,934	0,457	0,246
0,074	0,937	0,135	0,104	0,525	0,074	0,562	0,777
0,913	0,118	0,857	0,134	0,875	0,255	0,682	0,423
0,091	0,894	0,133	0,879	0,123	0,748	0,294	0,591

 Table 4.2: Results obtained from ANN (without DIP)

4.4 Part 2: With Image Processing

4.4.1 ANN system using tangent sigmoid functions and

In this part the images were divided into two groups, one for the training of the network whose parameters are given in Table 4.3. These parameters were applied during the training and test of the network. The training was started until the MSE was small enough for the generalisation of the network on the system.

Parameter	Value	Parameter	Value
No. of layers	4	Learning rate	0.2
Input transfer function	No	Momentum factor	0.5
Output transfer function	tangent	Layer sizes	2500, 50, 100, 2
Hidden1 transfer function	tangent	Training images	340
Hidden2 transfer function	tangent	Test images	146
AD output target	[1;0]	MSE	0.0058
NAD output target	[0; 1]	Iterations	1680
Test time	18 (ms)	Training time	99 (s)
Training performance	100%	Test performance	92.42%
Total performance	97.94%		

Table 4.3: Parameters of the ANN (with DIP)



Figure 4.3: MATLAB interface of ANN training tool

At the end of the training process of the network structure shown in Figure 4.3; the parameters were saved for the generalisation of the results. All test data was passed through the network

to check the validity of the resultant network. The results obtained showed that all the training images were classified correctly. 340 images out of 340 training images were classified with no error. Hence; the training performance is 100%. After the application of the test images on the network, 135 images of 146 test images were classified correctly. The test performance of the test reached the value of 92.4%. As an overall evaluation of the network performance, the network's efficiency was 97.94%. Sample of the output of the neural network is presented in

Table 4.4.

AD NAD AD NAD AD NAD AD NAD 0,99 0,03 0,99 0,04 0,62 0,03 0,89 0,01 0,97 0,01 0,01 0,95 0,36 0,97 0,11 0,99 0,89 0,02 0,95 0,02 0,89 0,00 0,94 0,01 0,98 0,06 0.12 0.98 0.13 1,00 0,05 0.99 0,95 0,05 0,96 0,00 1,00 0,04 0,95 0,03 0,05 0,96 0,05 0,95 0,04 1,00 0,00 0,97 0,99 1,00 0,16 1,00 0,12 0,18 0,99 0,00 0,00 0,85 0,00 0,88 0,01 0,82 0,01 1,00

Table 4.4: Output of the neural network (with DIP)



Figure 4.4: Curve of the mean squared error during training



Figure 4.5: Curve of test data mean squared error

4.4.2 ANN using logarithmic transfer function

In order to examine the behaviour of the neural network with different types of transfer functions, different types of transfer functions were used in different training configurations. As a first attempt, all the transfer functions were changed as shown in Table 4.5. After continuous training for approximately 20 minutes and 20000 iterations, it was found that the system is unable to converge toward a minimum MSE. The MSE stopped at a value of 0.5 with no any advance in performance. None of the training or test images were classified correctly. It was concluded that the use of logarithmic functions is unsatisfactory for the training of the system.

Parameter	Value	Parameter Value	
No. of layers	4	Learning rate 0.2	
Input transfer function	No	Momentum factor	0.5
Output transfer function	Logarithm	Layer sizes	2500, 50, 100, 2
Hidden1 transfer function	Logarithm	Training images	340
Hidden2 transfer function	Logarithm	Test images	146
AD output target	[1; 0]	MSE	0.5
NAD output target	[0; 1]	Iterations	20000
Test time	0 (s)	Training time	1200 (s)

Table 4.5: Parameters of the system (with DIP)



Figure 4.6: Mean Squared Error curve (Logarithmic sigmoid)

Figure 4.6 presents the MSE curve of the training process. From the figure it can be seen clearly that the MSE was fixed with no change during more than 19000 iterations. This means there was no hope of system to converge toward the desired target output using logarithmic transfer functions.

In order to overcome the convergence problem presented in the last experiment, many experiments were applied with changing the transfer functions. It was found that the system will not converge while using logarithmic transfer function in the output layer. For that, in the next experiment the parameters of the Table 4.5 were applied for the training of the network except for the output layer transfer function. Instead of the logarithmic function, a tangent transfer function was used and the training was started. After 3793 iterations the system converged to an MSE value of 0.00029 during 4 minutes of training. In this system, the training performance of 100% was achieved while performance of 93.9% was obtained in the test. The details of the results are presented in Table 4.6. The evolution of the MSE curve is presented in Figure 4.7. This curve shows that the MSE is decreasing during the training

process until it reaches high performance. The shape of the MSE curve is perfectly shaped with good decreasing slope. Table 4.7 resumes a part of the neural network output as given after the end of training process.

Parameter	Value	Parameter	Value
MSE	0.000287	Iterations	3793
Test time	18 (ms)	Training time	240 (s)
Training performance	100%	Test performance	93.94%
Total performance	98.35%		

Table 4.6 :	Results	of the	training	(with	DIP)
I HOIC HO.	itebuite	or the	uumms	(" " "	$\nu_{\rm H}$



Figure 4.7: Curve of the training MSE

AD	NAD	AD	NAD	AD	NAD	AD	NAD
0,999	0,009	0,998	0,014	0,973	0,003	0,986	0,006
0,001	0,992	0,002	0,988	0,030	0,997	0,014	0,994
0,289	0,012	0,524	0,002	0,971	0,000	0,522	0,001
0,714	0,988	0,470	0,999	0,031	1,000	0,447	0,999
0,962	0,004	0,979	0,032	0,995	0,001	1,000	0,008
0,043	0,996	0,021	0,972	0,005	0,999	0,000	0,991
1,000	0,040	1,000	0,024	0,999	0,041	0,998	0,001
0,000	0,959	0,000	0,975	0,001	0,953	0,002	0,999

1,000	0,000	1,000	0,000	1,000	0,009	0,998	0,311
0,000	1,000	0,000	1,000	0,000	0,991	0,002	0,679
1,000	0,000	0,996	0,000	0,990	0,049	1,000	0,009
0,000	1,000	0,004	1,000	0,010	0,944	0,000	0,990

4.4.3 ANN using linear ramp transfer function

In this part, the system is trained using linear ramp transfer functions for all the three layers. Linear ramp functions have different limits from the sigmoid functions. They are also not smooth on the entire curve. The parameters of this experiment are presented in Table 4.8.

Parameter	Value	Parameter	Value
No. of layers	4	Learning rate	0.02
Input transfer function	No	Momentum factor	0.5
Output transfer function	ramp	Layer sizes	2500, 50, 100, 2
Hidden1 transfer function	ramp	Training images	340
Hidden2 transfer function	ramp	Test images	146
AD output target	[1; 0]	MSE	0.045
NAD output target	[0; 1]	Iterations	2296
Test time	0.02 (s)	Training time	135 (s)
Training performance	96.9%	Test performance	94.7%
Total performance	96.3%		

Table 4.8: Parameters of the system, linear ramp (with DIP)



Figure 4.8: Mean Squared Error curve (ramp function)

It was noticed that the MSE in this case stopped decreasing at the amount of 0.045. It can be also seen from Figure 4.8 that the MSE has null slope near to this value. The training performance of this case was less than the performance of the previous experiments. However, the test results have shown slightly better performance in this experiment compared to the previously obtained results. The Figure 4.8 is showing also the test performance function (in blue colour). It is noticed that the test performance function is performing better in this case if compared with the previous cases. It is obvious from the Table 4.9 that the output of the network contains different values with more individual error from the expected target data. However, these values are still satisfactory to be considered as true. This can explain the high MSE of the results although the performance of classification is higher than previous experiments.

AD	NAD	AD	NAD	AD	NAD	AD	NAD
0,988	0,111	0,870	0,110	0,646	0,342	0,862	-0,030
-0,011	0,876	0,094	0,890	0,387	0,819	0,141	1,010
0,897	0,036	0,975	0,195	0,532	-0,179	0,740	-0,101
0,280	1,077	0,258	0,816	0,424	1,239	0,323	1,218
0,908	0,211	0,754	0,270	0,745	0,123	1,090	0,252
0,188	0,901	0,346	0,692	0,226	0,876	-0,119	0,826
0,927	0,274	1,130	0,234	1,131	0,448	1,105	0,018

Table 4.9: Output results of the network (ramp function with DIP)

-0,052	0,693	-0,046	0,707	0,001	0,498	-0,013	1,177
1,159	0,099	1,050	0,061	1,061	0,132	0,960	0,239
-0,182	0,785	-0,075	0,813	-0,115	1,048	0,023	0,625

For the rest of the work, the system is going to be tested using ramp functions with increasing the number of test images while decreasing the training images.

4.5 Part 3: Two Equal Parts of Database with Image Processing

In this part, the database was divided into two equal parts with 243 images each. Each part contained half of the AD classified images and half of those classified as normal images. The first part was used for the training while the second part was used for the test of the network. The training results are shown in the next part of this chapter. Other training parameters of the system were not modified during this part of the work.

4.5.1 Training using tangent sigmoid functions

In order to check the validity of the neural network classification system with larger group of test images, the training is applied in this experiment using tangent sigmoid functions. After the end of the training it was found that all the 243 training images were correctly classified with 100% performance. Among the 243 test images; just 214 images were classified as expected with performance of 88%. The system total performance was 94.3%. The training and test results of this structure are resumed in Table 4.10. It took the system 2 and half minutes of training to achieve a MSE of 0.013 after 2623 iterations. The curve of the MSE during the training is presented in Figure 4.9.

Parameter	Value	Parameter	Value
Training images	243	Test images	243
MSE	0.013	Iterations	2623
Test time	0.02 (s)	Training time	152 (s)
Training performance	100%	Test performance	88%
Total performance	94%		

 Table 4.10: Obtained results of test (tangent sigmoid with DIP)



Figure 4.9: MSE curve during the training (tangent sigmoid)

4.5.2 Training using ramp functions

The training of the same system using the ramp transfer function has given the results shown in Table 4.11. the output of the neural network is presented in Table 4.12 below. The results show that the training time was 136 seconds to reach a MSE of 0.05. The overall performance of the system was reported to be 88%. This was divided between the training that has a performance of 96% and the test that has performance of 80.25%. Compared to the results obtained in the previous experiment, it shows that the use of ramp function is less efficient than the use of the tangent type transfer functions. Figure 4.10 presents the MSE curve during the training of the neural network. It is decreasing all over the training period with perfect curve.

Parameter	Value	Parameter	Value
Training images	243	Test images	243
MSE	0.05	Iterations	2297
Test time	0.017 (s)	Training time	136 (s)
Training performance	96%	Test performance	80.25%
Total performance	88%		

Table 4.11: Ramp function training results with DIP



Figure 4.10: MSE curve of the training network

AD	NAD	AD	NAD	AD	NAD	AD	NAD
0,832	0,326	0,770	0,380	0,185	0,485	0,444	0,054
0,076	0,715	0,170	0,569	0,673	0,782	0,385	0,747
0,683	0,072	0,863	0,276	0,559	0,087	0,643	0,084
0,373	0,690	0,228	0,810	0,578	0,810	0,554	0,733
0,742	0,308	1,073	0,079	0,682	0,067	0,962	-0,030
0,265	0,698	0,331	0,776	0,226	0,966	-0,206	0,818
0,964	0,462	1,150	0,740	0,874	0,488	0,795	0,070
-0,049	0,579	-0,255	0,575	0,173	0,419	0,196	0,861
0,896	0,035	0,911	0,109	0,972	0,022	1,003	0,197
-0,412	0,992	-0,021	1,160	-0,041	0,930	0,176	0,702

Table 4.12: Network results (ramp with DIP)

4.5.3 Training of different parameters

In this part, different training parameters were implemented and their results were collected and tabulated in Table 4.13. In the first lines of the table, just single hidden layer network was implemented. In the rest of the table, 2 hidden layers were used in our network. The table illustrates the effect of changing the parameters on the obtained results. It is obvious that the results of any configuration are unpredictable prior to the practical implementation. The choice of the best configuration must be done with special care about the results, obtained error and the time consumed during the training.

Hidden	MSE	Training	Learning	Momentum	Training	Test
neurons		time	rate	factor	performance	performance
50	0.0016	100 s	0.1	0.5	94%	82%
75	0.0008	120 s	0.2	0.5	97%	83%
100	0.002	160 s	0.3	0.4	97%	85%
200	0.005	170 s	0.3	0.2	98%	89%
[50 50]	0.003	50 s	0.3	0.2	99%	91%
[50 75]	0.0028	55 s	0.3	0.2	99%	93%
[50 150]	0.0018	74 s	0.3	0.2	100%	94.33%

 Table 4.13: Comparison of different training results

4.6 Comparison with Previous Works

In order to provide an idea about the success of our work, a comparison with previous works is conducted. In (Ramirez et al., 2009), the proposed method gave a classification accuracy of the Alzheimer that reached 96.2%. In (Long & Wyatt, 2010) the classification rates of 88%, 94.7% and 97.3% were obtained in the AD classification system for different experiments. The research presented by (Padilla et al., 2010) has given a classification accuracy of Alzheimer disease between 87-97.9% based on different detection methods. (Davatzikos, 2011) have reported a sensitivity of 94.1% for their research about semi automatic classification system of Alzheimer disease. Accuracy of 86.7%, 87.2% and 85.8% were reported by the work presented in (Agarwal & Mostafa, 2011). The same work reported an accuracy of 99% based on the use of support vector machine used to classify AD cognitive scores. In (Saraswathi et al., 2013), three different machine learning algorithms were used. The three methods have given training accuracies of 96, 91, and 97%. The obtained test accuracies were 85, 81, and 86%. In (Mahmood & Ghimire, 2013) a system accuracy of 89.22% was obtained after training the system using 230 MRI images. In (Mahanand, Babu, Suresh, & Sundararajan, 2015) the paper presented the work based on cognitive radial basis artificial network for the detection of Alzheimer disease. The paper reported a training accuracy of 100% and test performance of 89.4% as their best achievements. In this work, accuracies of 88%, 94% and 96% were obtained in different training configurations of the system. Table 4.14 and Table 4.15 present a comparison of the results of different works that uses OASIS and other database.

Work	DIP?	Accuracy	Table number
Our work	NO DIP	88.8%	
Our work	With DIP	97.5%	
(Long & Wyatt, 2010)	With DIP	94.7%-97%	Table 2
(Saraswathi et al., 2013)	With DIP	81%-97%	Table 2
(Mahmood & Ghimire, 2013)	NO DIP	89.22%	Figure 6 & 7
(Mahanand et al., 2015)	With DIP	89.8%	Table IV

Table 4.14: Comparison of results of OASIS database

Table 4.15: Comparison of results of other database

Work	DIP?	Accuracy	Table number
(Ramirez et al., 2009)	With DIP	96.2%	No table
(Padilla et al., 2010)	With DIP	96-97.9%	Table 2
(Agarwal & Mostafa, 2011)	With DIP	89.5%	Table 1
(Margarida Silveira, 2010)	NO DIP	90.97%	Table II

CHAPTER 5

CONCLUSIONS AND PROPOSED WORKS

In this thesis, a computer system based on the artificial neural networks approach was developed for the classification of MRI brain images. The system's main goal was to classify the MRI images into images that have symptoms of Alzheimer disease and images that don't have these symptoms. Back propagation learning algorithm of the artificial neural network was implemented in this system to train the neural network structure. 486 brain MRI images were used in the classification system among which 243 of AD and 243 of normal brain images. These images were implemented for the training and validation of the neural network structure to accomplish the required task.

MRI images were treated using different image processing techniques prior to the application of ANN to increases the success chances of the proposed system. Images processing techniques are used to extract the features and reduce the processing cost and complexity of the system. During the assessment of neural network structure; the system was applied in to distinctive phases of function. The first phase is the training phase in which the ANN is trained to classify the images into two different categories. The second phase includes the test of the previously trained network to evaluate its performance with new unknown images.

After carrying out different experiments the neural network structure has proven high performance and capability to classify the data base images during the training and the test processes. The obtained results have shown that the proposed network is capable of learning all the training database images and classifying them to 100% performance. This conclusion was true for all tested types of transfer functions except the logarithmic sigmoid which produced an efficiency of 0% during the training and test of network. For this reason, the logarithmic transfer function was excluded from the proposed network. Tangent sigmoid transfer function based ANN has proven high training and test performance under different conditions of the database.

Training accuracy of the proposed system ranged between 96-100% for all the accepted cases while test performance range from 80-97% was obtained. High training and test performance

rates were obtained during the experiments of this work proving the capability of artificial neural networks to be used for the classification and identification of brain MRI images.

The work of this thesis is a motivation for more expanded studies in the medical applications of artificial neural network. The collected results during this work encourage for deeper and extended works on the detection of Alzheimer disease for fully automated classification system.

REFERENCES

- Agarwal, M., & Mostafa, J. (2011). Content-based image retrieval for Alzheimer's disease detection. In 2011 9th International Workshop on Content-Based Multimedia Indexing (CBMI) (pp. 13–18). https://doi.org/10.1109/CBMI.2011.5972513
- Anirban, S., & Kumar, P. s. (2015). Improved Digital Image Processing based Detection for Alzheimer disease. International Journal of Advances in Arts, Sciences and Engineering, 3(6), 1–7.
- Buckner, R. (2012). What is OASIS? Retrieved February 20, 2017, from http://www.oasisbrains.org/app/template/Index.vm
- Dahab, D. A., Ghoniemy, S. S. A., & Selim, G. M. (2012). Automated Brain Tumor Detection and Identification Using Image Processing and Probabilistic Neural Network Techniques, 1(2), 1–8.
- Davatzikos, D. H. Y. K. M. P. C. (2011). Semi-Supervised Pattern Classification: Application to Structural MRI of Alzheimer's Disease. In 2011 International Workshop on Pattern Recognition in NeuroImaging (pp. 1–4). https://doi.org/10.1109/PRNI.2011.12
- deFigueiredo, R. J., Shankle, W. R., Maccato, A., Dick, M. B., Mundkur, P., Mena, I., & Cotman, C. W. (1995). Neural-network-based classification of cognitively normal, demented, Alzheimer disease and vascular dementia from single photon emission with computed tomography image data from brain. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(12), 5530–4. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=41729&tool=pmcentrez&ren dertype=abstract
- Eskildsen, S. F., Coupé, P., Fonov, V., & Collins, D. L. (2014). Detecting Alzheimer's disease by morphological MRI using hippocampal grading and cortical thickness. *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, 38–47.
- Gharaibeh, N., & Kheshman, A. A. (2013). Automated Detection of Alzheimer Disease Using Region Growing technique and Artificial Neural Network. *Journal of World Academy of Science, Engineering and Technology*, 7(5), 12–16.
- Gonzalez, R. C., & Woods, R. E. (2001). Digital Image Processing (2nd Editio). New Jersey:

Prentice-Hall.

- Gupta, C. (2006). Implementation of Back Propagation Algorithm (of Neural Networks) In VHDL.
- Hudson, D., & Cohen, M. (1999). Neural Networks and Artificial Intelligence for Biomedical Engineering. (M. Akay, Ed.). San Francisco: IEEE.
- Katherine, R. (2012). *Machine learning for image-based classification of Alzheimer's disease*. Retrieved from http://www.doc.ic.ac.uk/~krg03/krgray_03102012_phdthesis.pdf
- Kiran, K. (2009). Parallelized Backpropagation Neural Network Algorithm using DistributedSystem.ThaparUniversity.Retrievedfromhttp://dspace.thapar.edu:8080/jspui/bitstream/10266/760/1/ME THESIS.pdf
- Lahmiri, S., & Boukadoum, M. (2013). Alzheimer's disease detection in brain magnetic resonance images using multiscale fractal analysis. *ISRN Radiol*, 2013, 627303. https://doi.org/10.5402/2013/627303
- Long, X., & Wyatt, C. (2010). An Automatic Unsupervised Classification of MR Images in Alzheimer's Disease. In *Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2910–2917). IEEE. https://doi.org/10.1109/CVPR.2010.5540031
- Mahanand, B. S., Babu, G. S., Suresh, S., & Sundararajan, N. (2015). Identification of imaging biomarkers responsible for Alzheimer's Disease using a McRBFN classifier. In *International Conference on Cognitive Computing and Information* (pp. 1–4). https://doi.org/10.1109/CCIP.2015.7100723
- Mahmood, R., & Ghimire, B. (2013). Automatic detection and classification of Alzheimer's Disease from MRI scans using principal component analysis and artificial neural networks. In 20th International Conference on Systems, Signals and Image Processing (pp. 133–137). https://doi.org/10.1109/IWSSIP.2013.6623471
- Margarida Silveira, J. M. (2010). Boosting Alzheimer Disease Diagnosis using PET images. In 2010 International Conference on Pattern Recognition (pp. 2556–2559).
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., ...
 Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease:
 Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, 7(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Mehrotra, K., Mohan, C., & Ranka, S. (2001). Elements of Artificial Neural networks.

- Padilla, P., Gorriz, J. M., Ramirez, J., Chaves, R., Segovia, F., Alvarez, I., ... Puntonet, C. (2010). Alzheimer's disease detection in functional images using 2D Gabor wavelet analysis. *IEEE Electronics Letters*, 46(8), 556–558. https://doi.org/10.1049/el.2010.0219
- Patil, K., & Kale, G. (2016). International Journal of Advanced Research in Study on Dignosis of Alzheimer 's by Using Neural Networks, 6(3), 37–40.
- Ramirez, J., Chaves, R., Gorriz, J. M., Lopez, M., Lvarez, I. A., Salas-Gonzalez, D., ... Padilla, P. (2009). Computer aided diagnosis of the Alzheimer's Disease combining SPECT-based feature selection and Random forest classifiers. In 2009 IEEE Nuclear Science Symposium Conference (pp. 2738–2742). IEEE.
- Saraswathi, S., Mahanand, B. S., Kloczkowski, A., Suresh, S., & Sundararajan, N. (2013). Detection of onset of Alzheimer's disease from MRI images using a GA-ELM-PSO classifier. In *Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI)* (pp. 42–48).
- Savio, A., Garcia-Sebastian, M., Hernandez, C., Grana, M., & Villanua, J. (2009). Classification Results of Artificial Neural Networks for Alzheimer's Disease Detection. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 641–648). Springer. https://doi.org/10.1007/978-3-642-04394-9_78
- Tantua, P. D. (2015). Intelligent Knee Rheumatoid Arthritis Identification System Based on Image Processing and Neural Classification. NEU.
- Wagner, A. D. (2000). Early detection of Alzheimer's disease: An fMRI marker for people at risk? *Nature Neuroscience*, 3(10), 973–974. https://doi.org/10.1038/79904
- You, J. (2007). Advances in Visual Information Systems, 4781(October). https://doi.org/10.1007/978-3-540-76414-4
- Zollitch, C. (2016). Grey scale image. Retrieved April 2, 2017, from http://www.stemmerimaging.co.uk/en/knowledge-base/grey-level-grey-value/