Early Recognition of Alzheimer's Disease using Brain MRI

Rania Ahmed Kadry Abdel GawadBirry
College of Engineering and Technology, Computer Engineering Department
Arab Academy for Science, Technology and Maritime Transport
Alexandria, Egypt
rania_kadry2012@yahoo.com

***Abstract*—Alzheimer's disease (AD) is a brain disease that causes a slow decline in memory, thinking and reasoning skills. It represents a major public health problem. Magnetic Resonance Imaging (MRI) have shown that the brains of people with (AD) shrink significantly as the disease progresses. This shrinkage appears in specific brain regions such as the hippocampus which is a small, curved formation in the brain that plays an important role in the limbic system also involved in the formation of new memories and is also associated with learning and emotions. Medical information on brain MRI is used in detecting the abnormalities in physiological structures. Structural MRI measurements can detect and follow the evolution of brain atrophy which is a marker of the disease progression; therefore, it allows diagnosis and prediction of AD. The research's main target is the early recognition of Alzheimer's disease automatically, which will thereby avoid deterioration of the case resulting in complete brain damage stage. Alzheimer's disease yields visible changes in the brain structures. The aim is to recognize if the patient belongs to Alzheimer's disease category or a normal healthy person at an early stage.**

Initially, image pre-processing and features extraction techniques are applied including data reduction using Discrete Cosine Transform (DCT) and Cropping, then traditional classification techniques like Euclidean Distance, Chebyshev Distance, Cosine Distance, City Block Distance, and Black pixel counter, were applied on the resulting vectors for classification. Image pre-processing includes noise reduction, Gray-scale conversion and binary scale conversion were applied for the MRI images. Feature extraction techniques follow including cropping and low spatial frequency components (DCT). This paper aims to automatically recognize and detect Alzheimer's infected brain using MRI, without the need of clinical expert. This early recognition would be helpful to postpone the disease progression and maintain it at an almost steady stage. It was concluded after collecting a dataset of 50 MRI , 25 for normal MRI and 25 for AD MRI that Chebyshev Distance classifier yielded the highest success rate in the recognition of Alzheimer's disease with accuracy 94% compared to other classification techniques used where, Euclidean Distance is 91.6%, Cosine Distance is 86.8%, City block Distance is 89.6%, Correlation Distance is 86.4% and Black pixels counter is 90%.

***Keywords*: AD, MRI, DCT, Chebyshev Distance classifier, Euclidean Distance, Correlation Distance and feature extraction.**

I. INTRODUCTION

AD is the most common form of dementia. While many strides have been made in elucidating the underlying causes of AD, studying the disorder has faced several hurdles: First, the structures affected by AD lie deep within the brain where biopsy is not practical. Second, animal models do not develop AD naturally, and genetically engineered models designed to mimic AD do not fully reproduce the human phenotypes [1]-[2]. Third, while studies using Positron Emission Tomography (PET) have been very useful for examining plaques and metabolic changes, they involve the injection of radioactive contrast agents. Many of these materials have short half-lives and must be created on-site, making PET is very expensive and difficult to be performed at nonspecialized centers.

Finally, studies which examine cerebrospinal fluid (CSF) require participants to undergo an invasive and sometimes painful *lumbar puncture* (the procedure of taking fluid from the spine in the lower backthrough a hollow needle, usually done for diagnostic purposes), potentially on multiple occasions [3]-[4]. In contrast to other techniques, Magnetic Resonance (MR) offers a non-invasive method for analyzing structural and functional brain characteristics without the need for ionizing radiation.It is impossible to predict with certainty the patient who is more likely to develop AD, but there are several factors that increase the risk. These at-risk populations include individuals in the prodromal stage of AD, termed amnestic Mild Cognitive Impairment (a MCI), and people at a genetic risk for developing AD.

A diagnosis of a MCIindicates that there is more memory decline than would be expected based on the person's age and education level, however memory impairment is not interfering with daily activities. It is estimated that 10-20% of people of 65 years old and up have a MCI, and out of those 10-15% will progress to develop AD in 3-4 years. [5]-[6], Because approximately 30% of people diagnosed with a MCI will remain stable or improve over time, it is important to find *biomarkers* (a naturally occurring molecule, gene, or characteristic by which a particularpathological or physiological process, disease, etc.) can be identified.

II. ALZHEIMER'S DISEASE BACKGROUND

Linear support vector machines were used to classify the grey matter segment of T1-weighted MR scans from pathologically proven AD patients and cognitively normal elderly individuals obtained from two centers with different scanning equipment. Because the clinical diagnosis of mild AD is difficult, the ability of support vector machines was tested to differentiate control scans from patients without post-mortem confirmation. Finally these methods were used to differentiate scans between patients suffering from AD from those with *fronto-temporal lobar degeneration (FTLD)* [7].

Up to 96% of pathologically verified AD patients were correctly classified using whole brain images. Patients with mild, clinically probable AD and age/sex matched controls were correctly separated in 89% of the cases which was compatible with published diagnosis rates in the best clinical centers. This method correctly assigned 89% of the patients with post-mortem confirmed diagnosis with either AD or *fronto-temporal lobar degeneration* to their respective group.

This study led to three conclusions: Firstly, support vector machines successfully separated patients with AD from healthy aging subjects. Secondly, they performed well in the differential diagnosis of two different forms of dementia. Thirdly, the method was robust and has proven that it could be generalized across different centers. This suggested an important role for computer based diagnostic image analysis for clinical practice. A support vector machine (SVM) is an example of a supervised, multivariate classification method.

SVMs are supervised in the sense that they include a training step to learn about differences between groups to be classified. It was reported that a similar distribution was found for all the classifications of AD from normal scans. Classification of FTLD from AD depended on voxels in frontal as well as parietal areas. In figure 1, lower panel for group assignation also shows cortical voxels scattered throughout the brain without any regionally specific pattern. They also contributed to a differentiation between two groups.

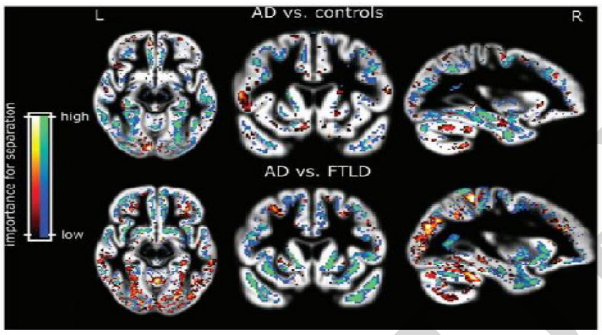


Fig. 1: Voxels most relevant for classification of patients from group I after SVM training with the data from group I (upper panel) [7].

The blue and green areas indicate higher grey matter volumeincreasing the likelihood of classification into normal. Red and yellow show regions where higher grey matter volume indicates the opposite. The lower panel depicts relevant areas for the separation of AD from FTLD. Blue and green indicate areas where lower grey matter volume indicates FTLD. Results are overlaid on the mean grey matter compartment image from all subjects [7]. Table 1 shows the results achieved using SVMs.

Table I. Results of SVM classification using grey matter from the whole brain for image analysis [7].

Group	Correctly classified (%)	Sensitivity (%)*	Specificity (%)*
AD and controls Group I	95.0	95.0	95.0
AD and controls Group II	92.9	100	85.7
AD and controls Group III	81	60.6	93.0
Dataset I for training, set II for testing	96.4	100	92.9
Dataset II for training, set I for testing	87.5	95.0	80.0
Group I + II	95.6	97.1	94.1
AD from Dataset II and FTLD Group IV	89.2	83.3	94.7

*Considering a correctly identified AD case as a true positive.

Subjects from group I were correctly assigned to the appropriate diagnostic category in 95.0% of trials with the leave-one out method using whole brain images (sensitivity 95.0%, specificity 95.0%). One 89 year old AD patient and one 86 year old control were misclassified. Figure 1Displays the regions that were most influential in making the binary classification between AD and normal when grey matter of the whole brain was used for analysis. Subjects from group IIwere correctly assigned 92.9% of trials with the leave-one out method using whole brain images and so on Finally, when group I was used to train the data and group II was used to test, 96.4% of patients were correctly assigned to the appropriate group. Conversely, when group II was used to train and group I to test, 87.5% of patients were correctly assigned to the appropriate group [7].

III. PROCESSING ALGORITHMS

The Image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set ofcharacteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques [8]. In this paper stages of AD MRI recognition were done as shown in figure 2, starting by image pre-processing stage where it is converted to gray scale then to binary image.Feature extraction techniques are then applied to the image by edge detection which is used for areacalculation.

Discrete Cosine Transform (DCT) is used as another feature extraction technique and zigzag scan was applied for image reduction to achieve feature vector which stores the most important features of the image required for classification. Vector resulted from zigzag scan goes through classification phase where traditionalclassification techniques such as Euclidean Distance, Chebyshev Distance, Cosine Distance, City BlockDistance and Correlation Distance are applied. Image cropping is used as a feature extraction technique that isapplied to the image after being converted to binary and black pixel counter is used for classification.

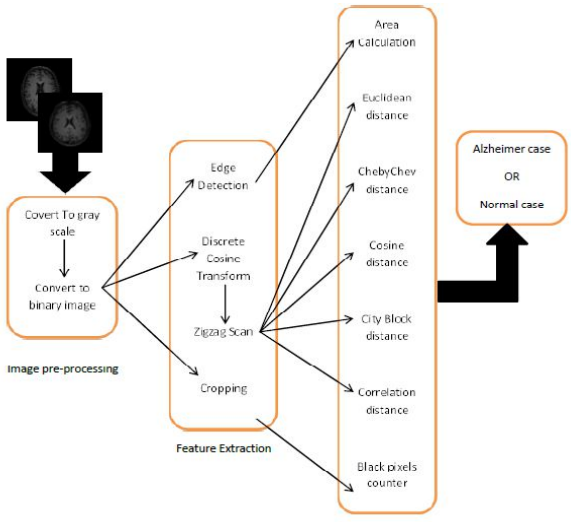


Fig. 2: Stages of Alzheimer MRI recognition

IV. IMAGE PRE-PROCESSING

Image pre-processing is useful for enhancing the visual appearance of the image and improving the manipulation of the datasets, but it also can emphasize image artifacts or even lead to a loss of information if not correctly used [9]-[10]. In this paper Gray scale conversion and Binary scale conversion were used as image pre-processing.

- *Gray Scale Conversion:*

In photography and computing, a gray scale digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information [11].

Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black (0) at the weakest intensity to white (255) at the strongest [12]. There are three methods that convert a RGB image to a gray scaled image [13] which are shown as follows by Eq. 1, 2 and 3.

Lightness method:

The lightness method averages the most prominent and least prominent colors:

$$\frac{\max(R, G, B) + \min(R, G, B)}{2} \quad (1)$$

Average method:

The average method simply averages the values:

$$\frac{R + G, + B}{3} \quad (2)$$

Luminosity method:

It averages the values, but it forms a weighted average to account for human perception. It is more sensitive to green than other colors, so green is weighted most heavily.

$$x_1R + x_2G + x_3B \quad (3)$$

Where x_1, x_2, x_3 represents the weights.

- *Binary Scale Conversion:*

A binary image is a digital image that has only two possible values for each pixel. Typically the two colors used for a binary image are black and white though any two colors can be used. The color [White] used for the object(s) in the image is the foreground color while the rest of the image is the background color [Black]. Before converting a Gray scaled image to a binary image the threshold of the gray scaled image must be first obtained. Otsu's methodis used to automatically perform clustering-based image thresholding, or, the reduction of a gray level image to a binary image. The algorithm assumes that the image to be thresholded contains two classes of pixels or bi-modal histogram (e.g. foreground and background) then calculates the optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal [12] as shown in figure 3.

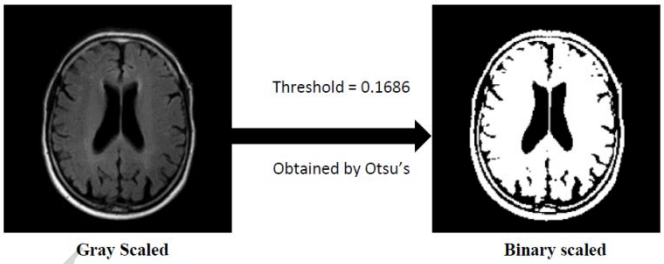


Fig. 3: The conversion from gray scaled to binary scaled

V. FEATURE EXTRACTION

In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction.When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features (features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen, it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a method of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy [14].Feature extraction techniques used in this research are:

- *Edge Detection:*

Edge detection is a set of mathematical methods which aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities.The result of applying an edge detector to an image may lead to a set of connected curves that indicate the boundaries of objects, the boundaries of surface markings as

IJHM

well as curves that correspond to discontinuities in surface orientation as shown in figure 4. Applying an edge detection algorithm to an image may significantly reduce the amount of data to be processed, therefore filter out information that may be regarded as less relevant.Edge detection methods are: Sobel, Canny, Prewitt andRoberts method [15]-[16].Before applying edge detection on a gray scaled image the threshold of the gray scaled image must be first obtained to get accurate edge detection, if applying the edge detection on the binary image then threshold is not needed [17].

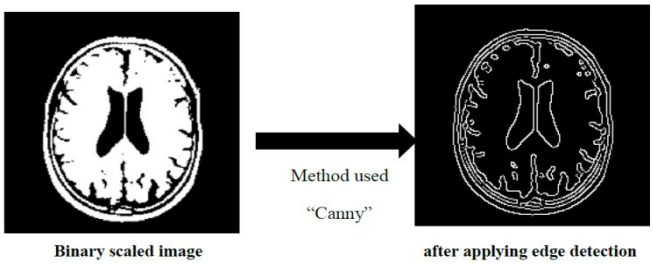


Fig. 4: The appliance of edge detection technique

- *Discrete Cosine Transform (DCT)*

A discrete cosine transform (DCT) expresses a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies [18]. DCTs are important to numerous applications in science and engineering, from lossy compression of audio (e.g. MP3) and images (e.g. JPEG) (where small high frequency components can be discarded), to spectral methods for the numerical solution of partial differential equations. DCT is a Fourier-related transform similar to the discrete Fourier transform (DFT), but using only real numbers. DCTs are equivalent to DFTs of roughly twice the length, operating on real data with even symmetry, where in some variants the input and/or output data are shifted by half a sample [19]. The most common variant of discrete cosine transform is the type-II DCT, and it's inverse, the type-III IDCT" [20]-[21]. The definition of the two-dimensional DCT for an input A and output image B is as follows:

$$B_{pq} = a_p a_q \sum_{n=0}^{N-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi (2m+1)p}{2M} \cos \frac{\pi (2n+1)q}{2N}, \quad (4)$$

$$0 \le p \le M-1 \;\&\; 0 \le q \le N-1$$

Where

$$a_p = \begin{cases} \frac{1}{\sqrt{M}}, & p = 0 \\ \sqrt{\frac{2}{M}}, & 1 \le p \le M-1 \end{cases} \quad (5)$$

and

$$a_q = \begin{cases} \frac{1}{\sqrt{N}}, & q = 0 \\ \sqrt{\frac{2}{N}}, & 1 \le p \le N-1 \end{cases} \quad (6)$$

M and N are the row and column size of A, respectively. If DCT is applied onto real data, the result is also real. The DCT tends to concentrate information, making it useful for image compression applications using Matlab toolkit [22]-[23] to obtain the following results as shown in figure 5.

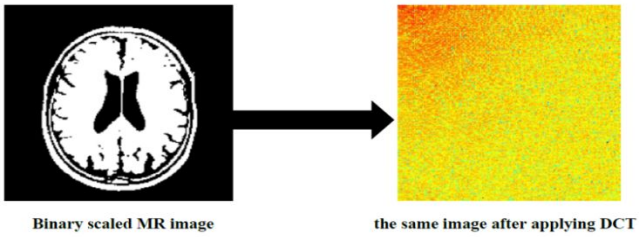


Fig. 5: The appliance of DCT technique on a binary scaled MRI.

After Applying DCT on the binary scaled MRI, zigzag scan is then applied on 10x10 part of this image. The main purpose of the zigzag scan is to group low frequency coefficients presented at the top of the vector. Although zigzag scan is under the title of Feature extraction, it simplifies the amount of resources required to describe a large set of data which reduces the complexity and computation power for the next stage which is Classification. Figure 6 shows 10x10 matrixes representing the cropped part of the image after applying DCT and zigzag scan results that are stored in a vector.

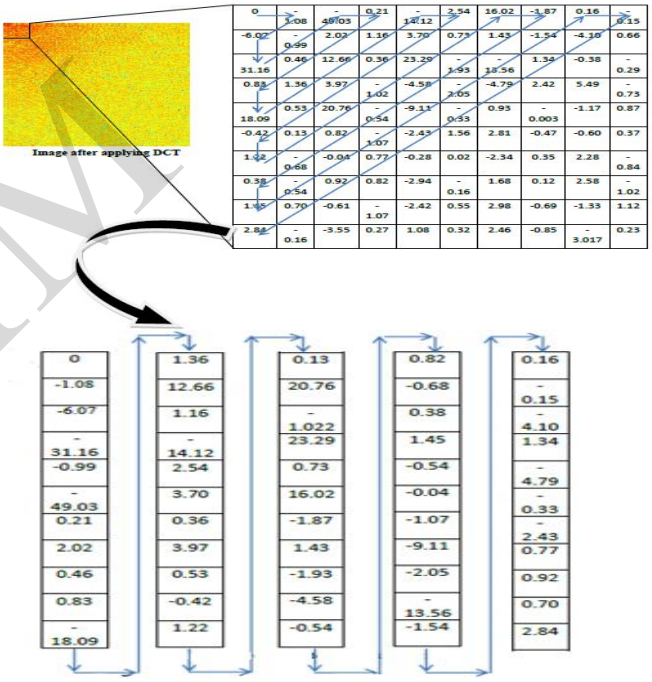


Fig. 6: 10x10 matrix representing the cropped partof the image after applying DCT in the first image and the second is the appliance of the final vector after applying zigzag scan on a DCT image.

- *Cropping*

Cropping refers to the removal of the outer parts of an image to improve framing, accentuate subject matter or change aspect ratio. In this research cropping is used as a simple feature extraction technique which extracts the most important features from the MRI for using it later in the classification stage as in figure 7. [24]

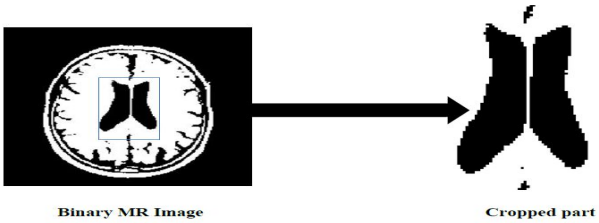


Fig. 7: The cropping technique applied on Binary MRI.

VI. CLASSIFICATION

Identifying to which of a set of categories a new observation belongs, on the basis of a training set of data is called Classification.Classification is supervised learning[25] and has a wide variety of techniques, in this research two techniques are used, which are:

- *Minimum Distance Classifiers*

It is used to classify unknown image data to classes which minimize the distance between the image data and the class in multi-feature space. The distance is defined as an index of similarity so that the minimum distance is identical to the maximum similarity. The most often used distances classifiers in this procedure are as follows [26]:

1. Euclidean Distance:

It is the distance between two points that one would measure with a ruler, and is given by the Pythagorean formula. The Euclidean distance between point's p and q is the length of the line segment connecting them is given as shown:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (pn - qn)^2} \quad (7)$$

$$d(p,q) = \sqrt{\sum_{i=n}^{i=n} (pi - qi)^2} \quad (8)$$

Where p, q represents the vectors and n represents the vector length.

2. Chebyshev Distance (Maximum Coordinate Difference):

Chebyshev distance (or Tchebyshevdistance), maximum metric, defined on a vector space where the distance between two vectors is the greatest of their differences along any coordinate dimension.It is also known as chessboard distance, since in the game of chess the minimum number of moves needed by a king to go from one square on a chessboard to another equals the Chebyshev distance between the centers of the squares, if the squares have side length one, as represented in 2-D spatial coordinates with axes aligned to the edges of the board [27]. The Chebyshev distance between two vectors or points p and q, with standard coordinates $p_i$and $q_i$respectively, as in Eq. 9:

$$d(p,q) := \max_i(|p_i - q_i|) \quad (9)$$

3. Correlation Distance:

Correlation distance is a measure of statistical dependence between two random variables or two random vectors of arbitrary, not necessarily equal dimension. An important property is that this measure of dependence is zero if and only if the random variables are statistically independent. This measure is derived from a number of other quantities that are used in its specification, specifically, distance variance, distance standard deviation and distance covariance [28].

When two sets of data are strongly linked together it is said that they have a high Correlation. When correlation is "1" means that the 2 sets of data have perfect similarity (perfect correlation), where if the correlation values is zero this means there is no similarity between these 2 sets of data.The mathematical formula of the correlation distance is as follows:

$$d(p,q) = \frac{n\sum_{i=1}^{i=n} p_i q_i - \sum_{i=1}^{i=n} p_i \sum_{i=1}^{i=n} q_i}{\sqrt{n\sum_{i=1}^{i=n} p_i^2 - (\sum_{i=1}^{i=n} p_i)^2}\sqrt{n\sum_{i=1}^{i=n} q_i^2 - (\sum_{i=1}^{i=n} q_i)^2}} \quad (10)$$

Where p, q represents the vectors, n represents the vector length.

4. Cosine Distance:

Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The cosine of "0°" is "1", and it is less than "1" for any other angle. The resulting similarity ranges from −1 meaning exactly opposite, to 1 meaning exactly the same, with 0 usually indicating independence, and in-between values indicating intermediate similarity or dissimilarity.In the case of information retrieval, the cosine similarity of two documents will range from 0 to 1. The angle between two term frequency vectors cannot be greater than 90°[29].The mathematical formula for cosine distance is shown in Eq. 11:

$$d(p,q) = \frac{\sum_{i=1}^{i=n} p_i \times q_i}{\sqrt{\sum_{i=1}^{i=n} p_i^2} \times \sqrt{\sum_{i=1}^{i=n} q_i^2}} \quad (11)$$

p, q represents the vectors and n represents the vector length.

5. City Block Distance:

It is considered having two points in the xy-plane. The shortest distance between the two points is alongthe hypotenuse, which is the Euclidean distance. The City block distance is instead calculated as the distance in x plus the distance in y. In most cases, this distance measure yields results similar to the Euclidean distance. However, with *City block distance*, the effect of a large difference in a single dimension is dampened (since the distances are not squared) [30].The mathematical formula of city block distance as follows:

$$d(p,q) = \sum_{i=1}^{i=n} |p_i - q_i| \quad (12)$$

p, q represents the vectors and n represents the vector length.

IJHM

- *Simple Classifiers:*
  1) Area Calculation.
  2) Black pixels counter.

1. Area Calculation:

The area of objects in binary images is mainly used in this research is applied on the MRI image after applying edge detection. Where the area is estimated for all of the "on" pixels in an image by summing the areas of each pixel in the image. The area of an individual pixel is determined by looking at its 2-by-2 neighborhood. There are six different patterns [31], each representing a different area where patterns with zero on pixels (area = 0), with one on pixel (area = 1/4), with two adjacent on pixels (area = 1/2), and so on. This technique of classification was not effective in the recognition of AD because the obtained area for most of the MRI images where nearly the same.

2. Black pixels counter:

This technique simply counts the number of black pixels inside a binary image [31] as shown in figure 8.

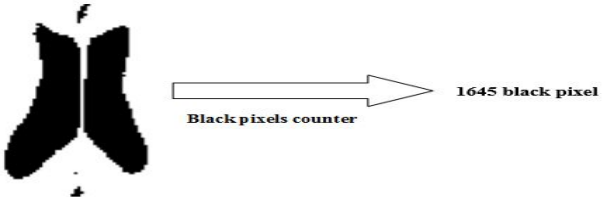


Fig. 8: Black pixel counter for a cropped part of Alzheimer binary scaled image

VII. RESULTS AND DISSCION

The results of the work done after image pre-processing and enhancement starting with feature extraction using DCT, then applying classification techniques: Euclidean distance, Chebyshev distance, Cosine distance, City block distance and correlation distance in order to differentiate images whether normal or AD infected, providing the details of each depicted technique where Feature extraction using DCT includes **c**onverting gray scaled image to binary image as shown figure 9, then applying Discrete Cosine Transform (DCT) for both normal and AD MRI binary images as shown in figure 10 and the results stored in vector for both images after zigzag scan as discussed.

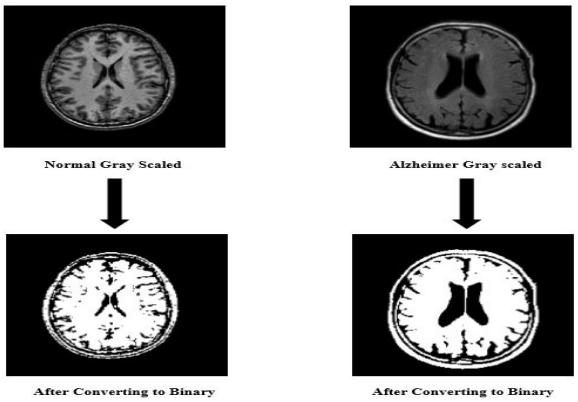


Fig. 9: The output after converting normal and AD MRI to binary.

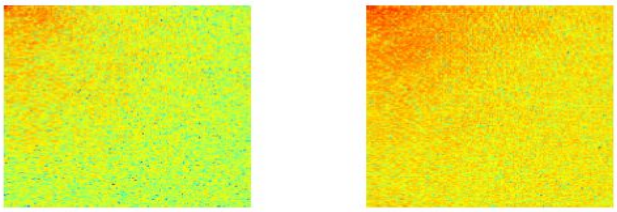


Fig. 10: The output after applying DCT for normal and AD MRI.

Collecting a dataset of 50 MRI images, 25 for normal MRI and 25 for AD MRI, taking each normal MRI as a reference image in every running process. Some performance measures including precision, sensitivity, F-measures, specificity and accuracy were calculated for each classifier by using the equations below [32]-[33] and the results are show in table 2.

Precision= $t_p/ t_p+ f_p$(13)

Sensitivity = $t_p/ t_p+ f_n$(14)

F-measure=*2*Precision*sensitivity/ Precision +Sensitivity*(15)

Specificity = $t_n/ t_n+ f_p$(16)

Accuracy= $(t_p+ t_n)/(t_p+ f_p+ f_n+ t_n)$ (17)

Where, $\mathbf{t_p}$ "True positives", $\mathbf{f_p}$"False positives", $\mathbf{t_n}$ "True negatives" and $\mathbf{f_n}$"False negatives"

Table II. The Results Of Each Classification Techniques

Classification techniques	Performance Measures				
	precision	sensitivity	F-measure	specificity	accuracy
Euclidean distance	87%	96%	91%	88%	91.6%
ChebyChev distance	90%	97%	93%	91%	94%
Cosine distance	83%	90%	86%	84%	86.8%
City Block distance	81%	97%	88%	84%	89.6%
Correlation distance	84%	88%	86%	85%	86.4%
Black pixels counter	100%	83%	90%	100%	90%

According to the success percentages that have been calculated, it was concluded that Chebyshev Distance classifier yielded the highest success rate in the recognition of Alzheimer's disease with accuracy 94%, compared to the other classification techniques used that shows lower values.

VIII. CONCLUSIONS

The research's main target is the early recognition of Alzheimer's disease automatically, which will therebyavoid deterioration of the case resulting in complete brain damage.

Alzheimer's disease yields visible changes in the brain structures. The aim is to recognize if the patient belongs to Alzheimer's disease category or a normal person at an early stage. Researches proved that there's no absolute treatment that treats Alzheimer's disease, but early recognition can be useful to postpone symptoms of dementia. This research aims to automatically recognize and detect Alzheimer's infected brain using MRI, without the need of clinical expert, to control AD patient at an early tolerable stage. This early recognition would be helpful to postpone the disease progression and maintain it at an almost steady stage.

IJHM

The proposed algorithm is done by using Matlab, it is so simple that the amount of computations required is very small compared to other algorithms surveyed in the paper. Initially, collecting MRI images for both normal and AD, where image pre-processing and features extraction techniques are applied including data reduction using DCT and Cropping, then traditional classification techniques like Euclidean Distance, Chebyshev Distance, Cosine Distance, City Block Distance, and Black pixel counter, are applied on the resulting vectors for classification.It is concluded that Chebyshev Distance classifier yielded the highest success rate in the recognition of Alzheimer's disease with accuracy 94% compared to other used classification techniques. The next challenge would then be to identify different stages of AD to monitor the disease progression. Developing another algorithm that can recognize AD using other imaging techniques such as Computed Tomography (CT) and applying other classification techniques such as Artificial Neural Networks..

REFERENCES

[1] Emily J. Mason, Manus J. Donahue and Brandon A. Ally, " Using Magnetic Resonance Imaging in the Early Detection of Alzheimer's Disease ". Feb. 2013, pp.225-227

[2] Persson J, Lind J, Larsson A, Ingvar M, Sleegers K, Van Broeckhoven C, et al. "Altered deactivation in individuals with genetic risk for Alzheimer's disease". Neuropsychology, 2008.

[3] Sinha G. "Peering inside Alzheimer's brains". Nat. Biotechnol. May 2011,:384–7.

[4] Filippini N, Macintosh BJ, Hough MG, Goodwin GM, Frisoni GB, Smith SM, et al. "Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele". Academy Sci. USA. 2009 Apr. 2009, 28;106(17):7209–14.

[5] Greicius MD, Srivastava G, Reiss AL, Menon V. "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI". Proc Natl Acad. Sci. USA. Mar 2004. 30;101(13):4637–42.

[6] Alzheimer's Association. "Alzheimer's disease facts and figures".Alzheimers Dement. 8th ed. Mar.2012. 5:131–68

[7] Stefan Klöppel Cynthia M. Stonnington, Carlton Chu, Bogdan Draganski, Rachael I. Scahill,Jonathan D. Rohrer,Nick C. Fox, Clifford R. Jack, Jr,John Ashburner,and Richard S. J. Frackowiak." Automatic classification of MR scans in Alzheimer's disease", Jan. 2008.

[8] IPOL journal,"Image processing", 2016

[9] Vincent Luboz, "Image Segmentation and pre-processing", department of surgery and cancer, Imperial College London.[10] BandyopadhyayS. K, "Pre-processing of Mammogram Images", International Journal of Engineering Science and Technology, vol. 2: pp. 6753-6758, 2010.

[10] PonrajD. N, M. E. Jenifer, P. Poongodi, J. S. Manoharan, "A Survey on the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer", Journal of Emerging Trends in Computing and Information Sciences. vol. 2: pp.656-664, 2011.

[11] J .Stephen "Stephen Johnson on Digital Photography. O'Reilly". ISBN 0- 596-52370-X, 2006

[12] Anke Meyer-Bäse,"Pattern Recognition for Medical Imaging" Chapter1: Introduction: pp. 1.Available at: http://books.google.com.eg/.

[13] John D. Cook, "Three algorithms for converting color to grayscale", Aug. 2009.

[14] Rafael C. Gonzalez, Richard E. Woods, ―Chapter 1, "Digital Image Processing", Introduction to Digital Image Processing‖. Third Edition. pp. 1-3.

[15] KaganamiH. G., BeijiZ., Region-Based Segmentation versus Edge Detection, 5 th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kyoto, Japan, 2009.

[16] Kohonen, "Self-organizing Map (KSOM) Extracted Features for Enhancing MLP-ANN Prediction Models of BODS", vol 18, 2008. [18] Wallace G. K, "Overview of the JPEG still Image Compression standard" SPIE, vol. 1244, pp. 220-233, 1990.

[17] Zhu O., AlwanA., "An Efficient and Scalable 2D DCT-Based Feature Coding Scheme For Remote Speech Recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, 2001.

[18] Umbaugh, Scott E ." Digital image processing and analysis : human and computer vision applications with CVIP tools", second edition, Boca Raton, 2010, FL: CRC Press. ISBN 97814-3980-2052

[19] KhayamSA., "The Discrete Cosine Transform (DCT): Theory and Application", Department of Electrical & Computer Engineering, Michigan State University, 2003.

[20] Matlab Student version, The Math Works kit 2015.

[21] "The Math Works Manual ", 2016.

[22] Makhoul, J., "A fast cosine transform in one and two dimensions". IEEE Transactions on Acoustics, Speech, and Signal Processing 28 (1): 27–34, Feb 1980, doi:10.1109/TASSP.1980.1163351

[23] "Cropping images Encyclopedia", 2008.

[24] Alpaydin, Ethem. "Introduction to Machine Learning", MIT Press, 2010, p. 9. ISBN 9780-262-01243-0

[25] Elena Deza& Michel Marie Deza, "Encyclopedia of Distances", page 94,Springer,2009.

[26] Cyrus. D. Cantrell ."Modern Mathematical Methods for Physicists and Engineers".Cambridge University Press. 2000, ISBN 0-521-59827-3

[27] Bickel, P.J. and Xu, Y. "Discussion of: Brownian distance covariance", Annals of Applied Statistics, 2009, 3 (4), 1266–1269. doi:10.1214/09-AOAS312A.

[28] Singhal, Amit, "Modern Information Retrieval: A Brief Overview". Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2001, 24 (4): 35–43.

[29] "Image processing Toolbox", 2016.[32] David M, "Evaluation: from Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation", 2011 [33] Kannikar I., Mihaela S., Serge M."Computer Analysis of Images and Patterns", Aug.2013, pp 188-195.