Predicting and understanding Alzheimer's using Machine Learning

Aswathy Thankachan[1], Bino Thomas[2]
APJ Abdul Kalam Technological University, India

***Abstract*—**: Alzheimer's disease (AD) is a type of dementia where no treatment is available. The progression of these diseases is harmful and eventually leads to death. There are currently no specific techniques that can confirm with a 100% certainty AD diagnosis. In recent years, Neuro-imaging combined with machine learning techniques have been studied for the detection of Alzheimer's disease. The accuracy of most of the existing methods is not up to the mark and tumor brain images are similar to AD brain images.MRI Image of the brain is one of the best biomedical instrumentation used to discover the AD existence. Around 300 axial view brain MRI is used for training and testing purposes, collected from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. In this paper, there is a fusion based feature extraction approach combined with GLCM, SURF, PCA coupled with SVM for classifying scans into categories of Normal Control (NC), Mild Cognitive Impairment (MCI) and Alzheimer's disease (AD). The results show that the highest classification performance obtained by using the proposed model.
.

Keywords*—*: Alzheimer's disease (AD), Mild Cognitive Impairment (MCI), Mild Cognitive Impairment (MCI), Gray-Level Co-Occurrence Matrix (GLCM), Speeded-Up Robust Features (SURF), support vector machine

I. INTRODUCTION

Alzheimer's disease (AD) is an progressive & irreversible neuro-degenerative disorder that causes a decline in memory and thinking skill . This is a type of dementia that begins with mild deterioration and further it may worsen progressively . Alzheimer's disease is the 6 th leading death cause in older people in the United States . More than 5 million US are living with this. It is estimated that, there will be 16 million Americans living having this disease in the year 2050. Damage to the brain cells can cause dementia and as a result brain cells aren't able to communicate with each other properly, which causes the loss of cognitive functioning and behavioral abilities of the person. Dementia has different stages, ranging from the mildest stage when the person starts to lose some cognitive functioning, to the most worsen stage when the person is completely dependent on other people for living. The major symptoms are memory loss, poor judgment, taking longer time for routine task, emotional unpredictability, disorientation of social skill etc. The stages are normal control which is a non Alzheimer's stage, cognitive impairment (MCI),and severe Alzheimer disease(AD).

The commonly used Medical imaging techniques like magnetic resonance imaging (MRI), are Computed tomography (CT), positron emission tomography (PET) to detect atrophy of this disease. Among the neuro imaging techniques PET scan uses some radioactive components which are harmful for patients and its is much more expensive. CT scans ,high dose of radiations involved. But there is no risk of exposure to radiations involved in MRI Scans. For Alzheimer's patents the radiations caused to make them mentally uncomfortable .MRIs can be acquired in multiple planes without repositioning the patients. They didn't use ionizing radiations like for imaging. They demonstrate superior soft tissue contrast than CT Scans. So this is one of the best biomedical instrumentation used to discover the disease existence..For conducting research experiments there exist image database includes Alzheimer's Disease Neuro-imaging Initiative (ADNI), The Open Access Series of Imaging Studies (OASIS) etc. These database are publically available database that unites researchers with study data as they work to define the progression of Alzheimer's disease (AD).

In the proposed approach ,various shape and texture features are extracted from axial regions of brain for the disease detection. Texture features are extracted using Gray level co-occurrence matrix. Surf algorithm is used to find the local, similarity invariant representation and comparison of images.Finally principle compound analysis is used for feature reduction. To classify the images into different disease categories we use SVM for classification.
The paper is organized as follows .Section 2 describes literature review for automatic AD detection of MRIs.section 3 proposed approach. and finally section 4 concludes the result of proposed model

II. LITERATURE REVIEW

MRI Image Analysis of Alzheimer

The general approach of AD classification using MR Images is described in Fig. 2.1. The MR Images were collected from the database. After selecting images from the dataset, features are first extracted and then chosen.

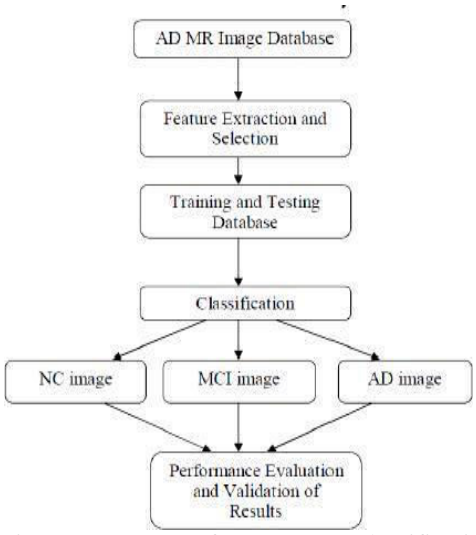

Fig. 2.1: Steps for Image Classification of MRI of Alzheimer's

Feature extraction helps to obtain a meaningful representation of the observations by removing repetitive and insignificant information from the observations. The classifier utilize this features to classify the data. The general assumption is that a classifier classify smaller & relevant features provides better accuracy .and utilizes less memory. After feature extraction, features are selected in which only some of the features from the dataset are selected and used in the training process of the learning algorithm. In this process the aims to find an optimal subset that increases the efficiency of the learning algorithm. Feature extraction & selection helps to attain a
compact pattern representation thereby reduce measurement cost and the improves the accuracy of classification. Feature selection algorithms deals with dimension reduction without transforming to a new set of dimensions. It enables by combining features from different data models. Features were used as an inputs to classifiers ,that assign them to specific class which represent and result are validated.

Feature Extraction Techniques

Feature extraction enable to minimize the original data by analysis certain characteristic of images which have relevant data, or features, that distinguish one pattern from another pattern. Features like shape, color, texture, wavelet based , region, histogram, GLCM based etc are extracted from the MRI images for the disease diagnosis.[7]

Yulin Ge et [4] al introduced a method by separating the human brain into total gray matter (GM) & white matter (WM). The studies shows that the volume of Grey matter loss seems to be a steady and a linear functionof age, but Loss of WM expected to be retard until middle adult life. Quantitative analysis of percentage volumes of grey matter and white matter can be used for the understanding the atrophy of brain due to the normal aging. Serge

A. Mitelman [5] analyzed the MR images and portioned into CSF, gray matter, and white matter tissue. 3 subjected groups were measured and found that compared to the healthy subjects, the overall patient group had a significantly smaller mean cortical gray matter volume and remarkably larger mean CSF volume

Al Naami [6] proposed a fusion based method to distinguish between the normal and (AD) MRIs. In this combined method MRIs collected and analyzed based on the use of Low pass -morphological filters to get the extracted statistical outputs through intensity histogram. The information that obtains from histogram are Mean Value, Standard Deviation and number ofpixels.Also, the artificial neural network (ANN) is used to evaluate their performance . Finally, the obtained result o with confidence accuracy (95%) has com-pared with classification accuracy of ANN (100 %). The robust of the developed method can be considered effectively to diagnose and determine the type of AD image. but the sample size is small

Namitha et.al [8] proposed a model that helpful in assisting the diagnosis of Alzhemer's using T1 weighted MRI images. It involves the construction of statistical features from MRI Images .The regions include multiple trans-axial slicesfromhippocampus&amygdalaregions,that playsanimportantrolein AD diagnosis. Ir-relevant tissues are removed using brain extraction tool (BET).Although first-order statistics based features are translation as well as rotation invariant which captures significant information about gray levels, It doesn't deals with information about the relative positions of the various gray levels within that image. GLCM defines quantify coarseness, smoothness and texture related information ,angular second moment (ASM),contrast, correlation, homogeneity and entropy. Features from multiple slices are then averaged, resulted into a smaller set of relevant features. The reduced set of features enhances the performance of decision learning system, and takes less memory and computation time. Drawback is it require priorknowledge ofROI.

Laila khader .et.al [10] 2015 proposed novel method for automatic classification ofMRIbasedonindependentcomponentanalysis(ICA) basedonADNIdatabase .ICA is a probabilistic and multivariate method for learning a linear transform of random vectors.The basic goal of ICA is to search for the components which are maximally as independent and non-Gaussian as possible. Its fundamental difference to classical multivariate statistical methods such as PCA and linear discriminate analysis (LDA) is in the assumption of non-gaussianity, which ensures the identification of original components, in comparison with these classical methods. Result: classification accuracy is still not optimal,the factors of age and gender have not been taken into account, mean accuracy, specificity, and sensitivity is less than 85

III. PROPOSED APPROACH

The proposed approach consist of training phase and testing phase is described in fig 3.1 fig3.2

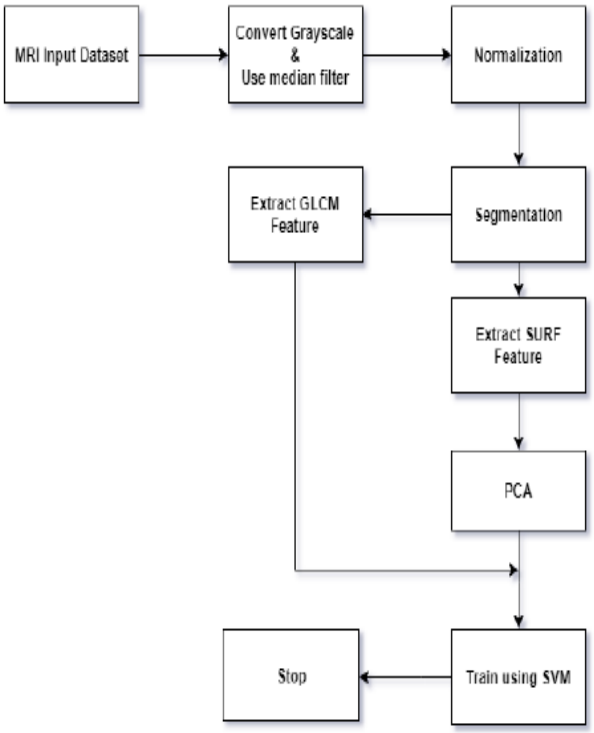

Fig3.1 Training Phase

Training Phase: There is a set of labeled MRI scan is pre processed for feature extraction and classification. Preprocessing steps include conversion of input image into grey scale image. Gray scale removes all color information, leaving only the luminance of each pixel. Since the image contains lot of noises like salt and pepper noise,spekle noise, Median filter is used to remove it. Then the image is converted into normalization purpose to achieve consistency in dynamic range for a set of input images Extract features such as Texture, Shape from the pre-processed scans and train them using an SVM classifier using this feature set. The output of the training phase is a trained classifier that is capable of predicting classification label based on features of MRI scan.

Classification: This phase can be summarized as follows:
a) Take as input, MRI scan of a patient.
b) Pre-process the MRI scan.
c) Extract the required features from the patient's MRI scan Using the cascaded model
d) Use the trained classifier to predict the categories of different stages.

Dataset

For conducting research experiments ,there are publically available dataset which can be accessed upon request. Here we use ANDI dataset that unites researchers with study data as they work to define the progression of Alzheimer's disease (AD).For the experiment around 300 MRI images were used in both training and testing purpose for the 3 different categories. The images are taken with an age group ranging from 45 to 95 from the database. The images are in bmp format .The axial view of hippocampus region of MRI images of 1.5 T is taken for experiments.

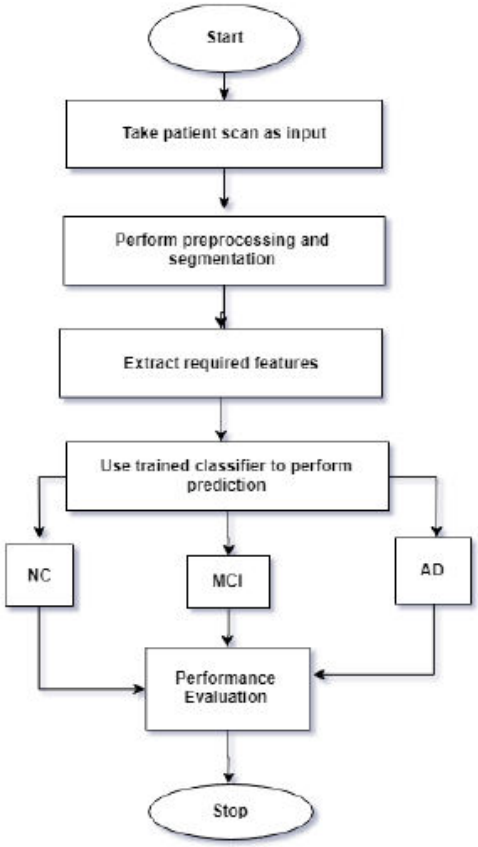

Fig3.2 Testing phase

Feature Extraction

GRAY LEVEL CO-OCCURRENCE MATRIX

Also referred as co-occurrencedistribution. It is one amongst the classical second-order statistical method for texture analysis. An image is composed of pixelsassociated with an intensity (a specific gray level), the GLCM is the tabulation of how often different combinations of gray levels co-occur in an image section image. Texture feature calculations use the contents of the GLCM to give a measure of the variation in intensity at the pixel of interest.GLCM texture considers the relation between 2 pixels at a time, called the reference and the neighbor pixel. It represents the distance and angular spatial relationship over an image sub-region of specific region of specific size. It is a matrix in which the number of rows and columns are equal to the number of gray levels in the image.The GLCM calculates, how often a pixel with gray-level value i occurs either horizontally, vertically, or diagonally to adjacent pixels with the value j.GLCM of an image is computed using a

displacement vector d, defined by its radius δ and orientation θ.

Common statistics derived from co-occurrence probabilities

- *Energy*

Also called Uniformity or Angular second moment.Measuresthe textural uniformity that is pixel pair repetitions. Detects disorders in textures. Energy reaches a maximum value equal to one.

- Contrast

Measures the spatial frequency of an image and is difference moment of GLCM.It is the difference between highest and lowest values of a contiguous set of pixels.It measures the amount of local variation present in the image

- Coorelation

It is a measure of linear dependencybetween gray levels

- Homogeneity

It is also called Inverse difference moment. It measures image homogeneity as it assumes larger values for smaller gray tone difference in pair element It is more sensitive to the presence of near diagonal elements in the GLCM.It has the maximum value when all elements in the image are same.Homogeneity decrease if constant increase while energy is kept constant

SURF

SURF (Speeded UP Robust Feature) is a local robust feature detector. The standard version of SURF is several times faster than SIFT and it is also more robust against SIFT. They are invariant tocommon image transformations like image rotation, scale changes, illumination changes, small changes in the view point.detectSURFFeatures(I) where I is the preprocessed image that returns a SURFPoints object, points, containing information about SURF features detected in the 2-D processed image I. The detectSURFFeatures function implements the Speeded-Up Robust Features (SURF) algorithm to find blob features. extractFeatures(I,points) returns extracted feature vectors, also known as descriptors, and their corresponding locations, from a binary or intensity image.The function derives the descriptors from pixels surrounding an interest point. The pixels represent and match features specified by a single-point location. Each single-point specifies the center location of a neighborhood. The method used for descriptor extraction depends on the class of the input points.

PCA

It is commonly used to decrease the dimensionality of images and get most of information. The central idea behind PCA is to find an orthonormal set of axes pointing at the direction of maximum covariance in the data.. The idea is to find the orthonormal basis vectors, or the eigenvectors, of the covariance matrix of a set of images, with each image treated as a single point in a high-dimensional space. It is supposed that the images form a connected sub region in the image space. The eigenvectors map the most significant variations between images and are preferred over other correlation techniques that assume that every pixel in an image is of equal importance . PCA is a powerful tool for analyzing data and once you have found these patterns in the data, and you compress the data by reducing the number of dimensions, without much loss of information.
.It calculates the covariance matrix by using the features values and Calculate the eigenvectors and eigen values of the covariance matrix.: By default eig does not always return the eigen values and eigenvectors in sorted order. Use the sort function to put the eigen values in ascending order and reorder the corresponding eigenvectors.Extract the eigen values from the diagonal of matrix using diag function, then sort the resulting vector in ascending order to form a feature vector. The Resulted feature vector combined to GLCM features to derive the new feature vector which is capable for extracting the relevant features.

SVM

Support Vector Machine is used as a classifier as it has gained in popularity in recent years because of its superior performance in practical applications, especially in the field of bioinformatics. A two-class support vector machine (SVM) classifier aims to do a hyper plane that maximizes the margin, which is the distance between the closest points on either side of the boundary. These points are known as the support vectors, and their role in the construction of a maximum-margin hyperplane.From a machine learning perspective ,for binary classification problem, the patients classified with Alzheimer's are modeled as positive class and the normal control corresponds to negative class. The patients with Alzheimer can be either of 2 stages whether it is a severe stage or mild cognitive stage .An SVM is modeled inside the above positive class is modeled. The inner class with positive class is modeled as a severe stage and negative class is labeled as an MCI stage.

Result

For evaluating classification results, the simplest measurements would be the classification accuracy rate, which is calculated from the number of correctly predicted samples divided by the total number of predicted samples. To test the results we used true positive, true negative, false positive and false negative.

True Positive (TP): positive samples correctly classified as positive.

False Positive (FP): positive samples incorrectly classified as negative.

True Negative (TN): negative samples correctly classified as negative.

False Negative (FN): negative samples incorrectly classified as positive.

	Predicted outcome	
True label	Positive (patient)	Negative (normal)
Positive (patient)	True positive (Tp)	False negative (Fn)
Negative (normal)	False positive (Fp)	True negative (Tn)

The Sensitivity (SEN) of the classifier is the number of true positives (TP) divided by the total number of real positives. In the example, it will be the number of patients. The Specificity (SPE) of the classifier is the number of true negatives (TN) divided by the total number of real negatives (controls). The Accuracy (ACC) number of true positive (TP) plus number of true negative (TN) divided by the sum of true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The Accuracy is calculated when number of positive samples equal to number of negative samples,

$$SEN = TP/((TP+FN))$$

$$SPE = TN/((TN+FP))$$

$$ACC = ((TN+TP))/((TP+TN+FP+FN))$$

AD VS NC

True Label	Predicted Outcome	
	Positive(AD)	Negative(NC)
Positive(AD)	8	2
Negative(NC)	2	8

According to above equation, Accuracy = 80%
Specificity=80%
sensitivity =80%

AD VS MCI

True Label	Predicted Outcome	
	Positive(AD)	Negative(MCI)
Positive(AD)	8	3
Negative(MCI)	2	7

According to above equation, Accuracy = 75%
Specificity=77.8%
sensitivity =88.89%

Conclusion

To develop a Alzheimer's disease detection procedure with better accuracy is the problem statement.So here developed a procedure which includes feature extraction and classification, pro- vides better accuracy that will support the cliniciansfor accurate diagnosis in classify the stages of MRIs.The obtained accuracy is 75%(for AD vs MCI) and 80%(for AD vs HC) for this particular dataset.The obtained training and testing time is less. Training and testing time is less.The future scope of the method is to incorporating MSME values and clinical data can improve the prediction.

Reference

1 Rajesh Sharma R, Marikkannu P,"A Survey on Image Classification Approaches and Techniques",International Journal of Advanced Research in Computer & Communication Engineering Vol. 2, Issue 1, January 2013

2 Ariruna Dasgupta ,Asoke Nath, Classification of Machine Learning Algorithms, Interna- tional Journal of Innovative Research in Advanced Engineering (IJIRAE) (March 2016)

3 Pooja Kamavisdar , Sonam Saluja , Sonu Agrawal , A Survey on Image Classification Approaches and Techniques"International Journal of Advanced Research in Computer and Communication Engineering January 2013

4 Yulin Ge, Robert I. Grossman, Age-Related Total Gray Matter and White Matter Changes in Normal Adult Brain, American Journal of Neuroradiology September 2002

5 Serge A. Mitelman," A comprehensive assessment of gray and white matter volumes and their relationship to outcome and severity in schizophrenia",Neuroimage. 2007 Aug 15

6 B. Al-Naami, N. Gharaibeh,"Automated Detection of Alzheimer Disease Using Region Growing technique and Artificial Neural Network", International Journal of Biomedical and Biological Engineering Vol:7, No:5, 2013

7 Yang Han, Xing-Ming Zhao A hybrid sequential feature selection approach for the diagnosis of Alzheimer's Disease, 2016 International Joint Conference on Neural Networks (IJCNN).

8 Namita Aggarwal, Bharti, and R.K. Agrawal, Computer Aided Diagnosis of Alzheimers Disease from MRI Brain Images International Conference Image Analysis and Recognition ICIAR 2012: Image Analysis and Recognition pp 259-267

9Arpita Raut "Detection of alzhemers disease via Statistical Features from Brain slice",Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference,2013 38

10 Laila khader Wenlu Yanga,f , Ronald L.M. Luib,Independent Component Analysis- Based Classification of Alzheimer's Disease MRI Data, Journal of Alzheimer's Disease 24 (2011)

11 Mohammad Reza Daliri," Automated Diagnosis of Alzheimer Disease using the Scale- Invariant Feature Transforms in Magnetic Resonance Images" J Med Syst (2012)