Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 1): pp. 4902-4924
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.01.525

A Hybrid Machine Learning Framework For Biomarkers Based ADNI Disease Prediction

I.Murali Krishna Research Scholar, JNTU, Kakinada, Andhra Pradesh, India

Challa Narsimham Principal, Dadi Institute of Engineering and Technology (DIET), Visakhapatnam

A. S. N. Chakravarthy Professor & HoD, Department of Computer Science and Engineering, JNTUK, University College of Engineering, Vizianagaram, Andhra Pradesh, India
imuralik2015@gmail.com

Abstract

In most of the real-time applications, machine learning algorithms are used to predict the Alzheimer's disease on high dimensional feature space. However, the condition of Alzheimer Dementia (AD) exponentially progresses due to lack of early intervention. Most of the traditional ADNI models are independent of image feature space and biomarkers due to high computational time and memory. In order to improve the disease prediction rate, this research work use multiple biomarkers for disease prediction on the ADNI training data. In this work, an improved CNN based feature selection method, a segmentation model and classification model are implemented on the large number of feature space and biomarkers. Current algorithms are tested and evaluated; an improved set feature selection method is proposed with re-sampling strategies. Experimental results proved that the present CNN feature selection-based segmentation and classification model has better prediction rate than the conventional models on high dimensional features.

1. Introduction

The most frequent form of dementia is Alzheimer's disease (AD), which is usually associated with the elderly (over 65). In the early stages of the disease, postulated treatments are more likely to be successful. Several studies have concluded and confirmed over recent years that 90 % of cases with validation according to neuro pathological standards can be diagnosed with exact diagnosis by clinical evaluation alone [1]. However, after a patient has been diagnosed, it can may have substantial loss of quality of life and the possibility of change or even deceleration of the progression of the disease. This is why it's important to diagnose dementia very early. Currently, several medications are

licensed by the United State Administration of Food and Drug (FDA) for the treatment of persons with an AD diagnosis. In earlier stages of the disease, treatment that modifies the disease is more likely to have a significant effect. Population stratification is necessary if sufficient subject matter is to be recruited and the outcomes of experimental treatments are to be tested in subjects where it is expected that the effects are most successful, thus minimizing the total cost of the trial through removing false positives in a previous step. Recent research concentrates at a much earlier stage on the identification of risk topics. Imaging MR tests are often part of clinical assessment protocols of MCI patients. Neurodegeneration designs, however, may not always be compatible with specific anatomical structure or functional areas standard descriptions. Therefore the study will theoretically decrease the biomarker capacity to detect variations or improvements over time if it is limited to predefined regions. Alzheimer's disease (AD) is common among the ageing population and this has also been on the rise to a significant degree each year in its death-related proportion (Alzheimer's Association 2012)[2]. It has also been reported that individuals with mild cognitive impairment (MCI) condition who are also referred to as the predecessor of dementia in AD may develop to an AD condition at about 10 percent each year. Although there are no pharmaceutical medicines to recover from AD or MCI to a Cognitive Normal (CN) condition, it remains important to be able to perceive the ailment at its initial stage to delay its progression. Due to its effective interpretation of the results, the techniques of selecting features have gained much attention today in the analysis of neuroimaging. Following the AD diagnosis, the patient and their family become anxious and go through feelings of frustration, uncertainty, anger, loss etc. MRI diagnostic values were magnified using automated classification of the MR images. In the case of the MRI technique, there is a strong magnetic field with the radio waves used to produce brain images. All of these were used to study the brain where even the slightest change is noticeable. They may also be used to cross out other brain-related conditions and cognitive symptoms. They were also used to measure brain and its shrinkage. This helps to diagnose an AD. The Filter methods were normally used as a preprocessing step. Feature selection was independent of any particular machine learning algorithm. In contrast to this, the choice of features was made with a resulting variable based on the scores obtained in the various statistical trials made for the correlation. The rank of each feature will be in line with some of the univariate metrics and to select the features with the best ranking[3].

Classification is a supervised form of learning used in biomedical science. Training examples are required in the classification task to predict a target class from an unseen example. However, sometimes imbalanced class distribution in the training data is one of major issue in classification problem. In real-time medical data , the number of disease features available to a classifier for training and testing are insufficient.

Alzheimer's is a neurodegenerative type of Dementia which accounting for 60-90% of all Dementia. In many countries, AD is found to be one of many striking causes of death. Initially Dementia symptoms start with a memory decline. Patients lose their cognitive and motor skills which make their daily lives more difficult. Ultimately they need the help of others to look after themselves. Care for Dementia is a major socio-economic challenge facing us today (Alzheimer's 2019 report). Accurate diagnosis of critical factors causing the disease is vital for prompt treatment. Several techniques of classification of high- dimensional patterns were based on computational anatomy, functional neuroimaging, and neuropsychological analysis. These techniques demonstrate that classifications of individuals may lead to higher accuracy rates in diagnosis of disease, as opposed to group analysis. Early diagnosis helps to prevent progression of illness. It is estimated that Dementia will increase by three hundred per cent over the next forty to fifty years. A preceding stage of Dementia is mild cognitive impairment (MCI)[4]. Treatment can be planned if the disease is diagnosed at this stage and thus can prevent progression to Dementia. Consequently, MCI diagnosis is vital in preventing progression to AD and death due to Dementia. The principal symptom of this prodromal stage is the decline in cognitive abilities. Patients will be able to perform daily tasks but there will be loss of cognitive abilities. Mild cognitive impairment is noted by memory loss and other cognitive disturbances that can progress to Dementia when untreated. Estimates of MCI prevalence among older people range from 2-10% at age 65 to 5-25% by age 85. The rate of conversion from MCI to AD is found to be 5
% per annum and 80 % per 10 years[5].

MCI exhibits positive amyloid and negative amyloid types. Patients with amyloid-positive MCI, episodic memory loss, association with CSF tau protein markers if the rate of progression to Dementia is increased present. If at this stage patients are identified, modification of the lifestyle may prevent progression to Dementia. Neuropsychological (NP) evaluation is a comprehensive evaluation of the cognitive processes. The role of NP evaluations has advanced to that of evaluating the cognitive and psychosocial effects of brain damage that is often well-located. NP can be very useful for early diagnosis and differentiation between NC, MCI, AD and also to determine if a patient is responding to treatment. The reliable and standardized neuropsychological batteries used which are easier to administer, cheaper than neuroimaging, less time consuming, and early accuracy picks up brain damage. Dementia diagnosis from Alzheimer requires data from multiple modalities of neuroimaging and clinical data obtained through neuropsychological tests[6]. Although several machine learning algorithms are used for disease diagnosis, statistical methods, ANN (Artificial Neural Networks), SVM, Random Forest, and decision tree classifier are the frequently used techniques. For image analysis and classification, many fully automated and semi-automated patter recognition

methods were used. The combination of MRI morphometric analysis based on FDG-PET and Voxel makes it possible to detect mild AD. Support vector machine is capable of classifying high-dimensional data and therefore been used to classify T1-weighted MRI[7]. Numerous studies have conducted automatic classification for early detection of MCI with structural MRI .On the basis of the type of feature to be obtained from MRI, three different approaches such as voxel-based, vertex-based techniques and region-based methods of interest are employed. In this voxel before feature selection was smoothed or down-sampled. Anatomical Parceling by registration with an atlas to group voxels of the different anatomical regions was another method used in this analysis. Hippocampus region has also been evaluated as hippocampus is the main region showing significant atrophy due to AD and MCI . Early detection by machine learning with relevant biomarkers helps to treat and prevent disease progression in a timely manner[8]. Application of machine learning methods in the classification of Normal Control, MCI and AD using Neuroimaging (NIMG) data from MRI, PET, Single Photon Emission Computed Tomography (SPECT), Neuropsychological (NP) scores and Genetic Biomarkers are major growing research areas. Some studies used only structural biomarker MRI, PET, Cerebrospinal Fluid (CSF), and genetic information to discriminate against AD and MCI while some studies used neuropsychological data alone to diagnose disease.

2.Literature survey

Alzheimer's disease (AD) is a neurodegenerative disorder which affects brain functionality, an important part of the central nervous system. AD normally affects people aged 65 and above, resulting in brain region atrophy that causes memory , thinking, language, and learning ability to decline[9] The main cause of memory loss is neuron death that causes brain atrophy due to the formation of plagues and tangles in the brain region. After the signs of AD have been identified a patient will live for an 8-20 year span depending on the severity of the disease. The drugs available for AD can control AD progression or slow it down. Oxidative stress was found to be an early occurrence in AD before other cytopathological symptoms, and thus plays a key pathogenic role in AD. The macromolecules in prone AD neurons are primarily affected by oxidative stress. For example, in AD patients a marked increase was observed in the levels of 8-hydroxyl-2-deoxyguanosine, produced as an oxidative product of DNA and RNA. It has also been shown that oxidative protein modification results in weak ATP synthesis, and strong protein cross-linking provides resistance to the removal of abnormal proteins when they are extensively ubiquitinated. In addition, abnormal increase in levels of reactive thiobarbituric acid (TBARS), malondialdehyde (MDA), 4-hydroxy-2-transnonenal (4-HNE) was also observed as a result of lipid peroxidation in

AD brain . Evidences from the past decades suggest that antioxidant therapy can be useful in the fight against many neurodegenerative diseases like AD. Antioxidants play a significant role in reducing the free radical mediated damage done by the toxic reactions in neuronal cells. For example, alcohol tocopherol has been shown to be a strong, lipid-soluble chain-breaking antioxidant to attenuate the toxic effects of alcohol and improve rodent cognitive functions. In addition, it has been found that vitamin E protects membrane fatty acids from lipid peroxidation and is more effective against peroxyl radical cell viability loss. Based on their mode of action the neuroprotective antioxidants can be divided into three different classes. Direct antioxidants are the first and major type of antioxidants, which chemically interferes with the free radicals formed. To this category belong aryl amines, indoles, lycopenes, retinol, flavonoids, polyphenols, tocopherols, and other monophenols. The second type of antioxidant compounds are called indirect antioxidants which are not consumed and have long half-lives and prevent free radicals from forming. The third group of antioxidants is called the metabolic antioxidants. Despite some clinical trial pitfalls, active or passive immunization holds better hope for treatment and even prevention of AD . It has been found that the APOE4 isoform is associated with increased serum cholesterol, which in turn may play a direct role in aggregating and depositing A. The exact mechanism by which the regulation of cholesterol influences the production of A remains elusive, however. Cholesterol was reported to adversely regulate the activity of al-secretase and positively regulate the activities of al- and al-secretase. These reports suggest that a net decrease in cholesterol levels increases the mediated APP cleavage of some secretase and decreases the mediated APP processing of the amyloidogenic pathway. As a result, compounds that inhibit cholesterol biosynthesis can also be a potential therapeutic agent against AD which is used DTIs to identify changes in the structure of white matter in MCIs as well as its sub-types and concentrated on exploring whether DTIs can be used as MCI imaging markers as possible. DTI 's ability to discern CNs MCIs has been tested through binary models of logistic regression[10].

They suggested a straightforward statistical test for the presence of distributed spatial patterns for fMRI datasets. On the multivariate model , various maximum estimates of different restrictions were made to estimate the extent of spatial heterogeneity in data from fMRIs. Simulation time courses were used to evaluate the test statistics derived from actual fMRI data, and analyzed from a real fMRI experiment. Measuring spatial variability in fMRI has significant theoretical consequences of better characterizing neurological disorders, such as stroke and Alzheimer's disease. They also suggested SVM approach for the classification of medical images in two groups, which would create an ideal hypersphere. The outcome achieves decreased false positive and false negative error levels to produce very good results in the classification.

Khan et al . introduced a new CAD approach for earlier diagnosis of Alzheimer's on the basis of Non-negative Matrix Factorization (NMF) as well as confidence-bounded SVM. The CAD method is designed for analyzing functional brain scans as well as classifying them. The suggested technique of NMF- SVM results in a maximum classification accuracy of 91 per cent with high sensitivity and specificity levels (higher than 90 per cent). The NMF-SVM CAD method is an effective technique for both SPECT and PETAD image classification[11]. They implemented a new irregular brain detection system. For extracting features from the MR images a multi-resolution technique was used. Kernel Principal Component Analysis (KPCA) has been harnessed to reduce the size of the features, with the objective of obtaining the discriminating features. To identify brain MR images as regular or abnormal, Least Squares SVM (LS-SVM), a new version of SVM with low computational costs, was used. For experiments, a 6-fold stratified cross-validation procedure was implemented which indicates that the proposed scheme surpassed other competent schemes for predicting accuracy with relatively few features[12].

With the non-linear Radial Basis Function (RBF) kernel, LS-SVM increases performance. The proposed algorithm specifies optimized values for the RBF kernel hyperparameters and implemented k-fold stratified cross validation in order to boost device generalization. Results show that the experimental machine learning program has the ability to make accurate brain abnormalities predictions from the subjects themselves[13]. The input image is processed using the Standard Median (SM) filter and the Center Weighted Median (CWM) filter and, consequently, the output noise is detected using an impulse detector that compares the original value for a tri-state decision. Tri-state decision technique effectively reduces the noise of impulses without losing information of the picture. The methodology proposed provides consistent output for a large

variety of images[13].

Habes et al[14] implemented an approach that uses deep learning to segment the digital image. The segmentation is done using a technique called DnCNN (Denoising Convolutionary Neural Network) that is a pertained or integrated system. The experimental results show that few methods of segmentation increase the computational cost and manual settings are required to enhance the efficiency and procedures implemented to solve the problem of segmentation.

Liu et al[15] proposed an early diagnosis of Alzheimer's disease approach for region-based segmentation using Voxel Based Morphometry (VBM) applied on T1 MRI image. They took the MRI images randomly for confirmation of VBM accuracy. The VBM Z-Score chart is used to assess the MCI and normal healthy images and is 87.8 percent accurate.

C. Ludwig et al[16] proposed a technique for automatic brain MRI image segmentation

using adaptive histogram analysis, morphological operations and knowledge-based rules to precisely distinguish different regions. Morphological operators are used to isolate the area of the cerebrum, adaptive thresholding applied to mark ventricular and extra ventricular areas, eventually classifying the MRI into gray and white matter to detect the anomalies. Martí et al.[17] suggested a novel approach that would use voxel-based methods to remove irregular regions of the brain using MRI images. The pathological characteristics are extracted using the field of spatial filtration. The collected data are compared with manually segmented images and the results are produced efficiently. Rallabandi et al.[18] presented a paper describing a median filtering switching scheme that is ideal for data compression and edge detection and is used to extract impulse noise from digital images with less signal distortion, as well as applying specific noise removal operations to obtain filtered images and finding that the accuracy is lower as compared to median filtering. Raza et al[19] introduced several thresholding methods for segmentation and recorded various thresholding methods such as Global Thresholding, Local Thresholding, Adaptive Thresholding, and selection such as Iterative Based, Histogram Based, and Otsu 's method and clustering method. Just two classes are created with this approach and can't be used for multi-channel image. Paul M. They suggested an approach which would map patients affected by the hippocampus and ventricular area atrophy. They developed an anatomical surface-based modeling approach to map the complex changes that occurred in hippocampus and ventricular regions. Brain maps describe the anomalies that disassociate with the healthy. The suggested approach assists in recognizing the earlier changes in brain images and detecting anomalies. Stamate et al[20] had published a paper on the segmentation of the brain tissue ventricular area using iterative thresholding and the active counter model. Iterative thresholding is initially used to identify the endocardial boundary and to estimate blood and myocardial signals, and the algorithm proposed improves precision and specifically segments the tissue's left ventricular region.

The Hough transform algorithm is a probabilistic fiber tracking method based on a voting system. The algorithm tests candidate 3D polynomial curves by assigning a score to a single curve that passes a seed point in d-dimensional space in a diffuse imaging volume. The aim of the algorithm is to find all possible curves that go through selected seed points while calculating the values and, finally, to select the curve that has the highest value. Curves with the highest scores are stored in a range called the Hough transform in d dimensions and can represent potential fiber-tracts in the brain. The results are achieved by means of a voting process in which the candidates' score is defined through real valued local votes in the case of curves based on diffusion data. This ODF model does, however, best use this limiting angular resolution (although the protocol is not ideal for fiber crossing resolution). Elastic deformations from echo-planar image

distortions were then applied to the 3D coorders of each retrieved fibre in order that the anatomy would be more accurately aligned, mapping the average b0-image to the T1-weighted image (there is an uninterrupted anatomical reference, assumes). For tracts to cross marked core areas, labels have been dilated with 5 voxel width isotropic box kernel [21]. They are flexible, nonlinear models that model complex applications in the real world. They are adaptable to individual data without the functional or distributive form of the underlying model being explicitly specified. The later likelihoods required to establish classification rules and statistical analysis can be estimated. However, the learning time for large ANNs is high, and it requires a lot of calculation to adjust parameters.

- k-closest neighbours: k-NNs are proposed to be classified as the cluster most frequents among the k-closely represented samples from numerous neuroimaging studies[22]. In two or more classes, the minimum average distance is classified according to the class. This is a powerful and easy to apply classifier. It provides accurate distance and weighted average pixel information. However, due to its "placious" learning algorithm, its efficiency degrades for large and large-scale data. The choice of k affects slow and high memory costs in the classification performance. Gaussian mixing model: it is easy to implement GMM proposed by numerous researchers on neuroimaging [23]. Thanks to its probabilistic basis, efficient and robust. For large data sets, it does not take much time. The exponential functions of this classification are not excluded but are slow to follow trends over the course of time.

Support of vector machinery, especially when a space for operations is very large, SVMs used in several works have high performance. These machines offer training in large areas with small learning sets for generalizable nonlinear classifications. The number of classification errors for all samples is minimized. Learning SVM is however slow and takes time to calculate it. The cost of storing data is high. The best kernel for a particular task is not approved a priori for a method. The optimal solution can therefore depend on the kernel selected. Methods of Atlas segmentation in medical image analysis are commonly used.

3.PROPOSED BIOMARKERS BASED ADNI DISEASE PREDICTION MODEL

Figure 1, describes the overall framework of the proposed model on the ADNI biomarker disease prediction. In this framework, initially ADNI slices are extracted from the input .nii image by using the Freesurfer toolbox. These slices are used to find the essential key features for disease prediction. In this model, a new feature extraction method is proposed to extract the key features in the CNN framework. A TADPOLE biomarkers training dataset is used to find the severe disease patterns in the ADNI image features. Biomarker

segmentation algorithm is used to group the biomarker features based on the ADNI features. An integrated biomarker based ADNI training dataset is prepared in order to predict the best disease patterns by using the classification model as shown in figure 2. The gray matrix and run length matrix are common features of the second-order statistics. Statistical parameters pioneers for the use of texture analysis. The method was later used in medical image studies for texture analysis, medical ADNI images and many other disease prediction entities. GLCM measures the number of equivalent value pixels in the same direction. Runtime frequency in several directions is taken into consideration in the matrices. MRI images are pre- trained and divided into statistical mapping tools. The goal is to apply Gabor filters and retain the image information for the direction of the spatial frequency. The 2D Gabor Wavelet enmeshes each part of the 3D brain..MRI brains made up of a number of transaxial cuts. Every piece is attached to the wavelet in various angles and orientation. The 2D Gabor filter Characteristics are removed with each slice in the 3D MRI. Characteristics. The "g" spatial aspect ratio represents the orientation angle of the brain MRI . Now input image "I" Forms a vector of 48 dimensions Each of the vectors of each filtered image represented by the concatenation. For each part of the MRI 3D this vector is generated. The ADNI was launched in 2003, and its main objective is to assess the progression of Alzheimer's disease and MCI through different biological markers such as MRI, PET, clinical and neuropsychological evaluation. For more info, we used the data collection of the Open Access Series of Imaging Studies (OASIS), a research community that seeks to provide freely available neuroimaging datasets. Evaluating the performance of any computational algorithm is an important factor in order to prove its effectiveness, generalization potential and the consistency of the results by choosing a proper metric evaluation. To assess how often segmentation algorithms are effective for various datasets, the assessment criteria used in this study. Such segmentation algorithms are usually validated using MR brain images of both normal and pathological brains from publicly accessible databases (benchmark datasets) and scanning centers (real-time photos). The contrast and strength representation of these images diverge with various modalities of tissue, pathologies, and MR imaging. There are several methods available to determine the efficiency of the segmentation algorithms based on the different characteristics of the metrics. Most commonly used methods are visual perception of human beings. One widely used approach is the supervised evaluation process, in which the results of segmentation algorithms are compared with manual segmented images or regular images. The performance of the proposed brain tissue segmentation algorithms is validated by the proposed method using the similarity measures Jaccard (J) and Dice (D) with manual segmented and resulting segmented images. Such techniques of assessment take two images as input and generate the value between 0 and 1.

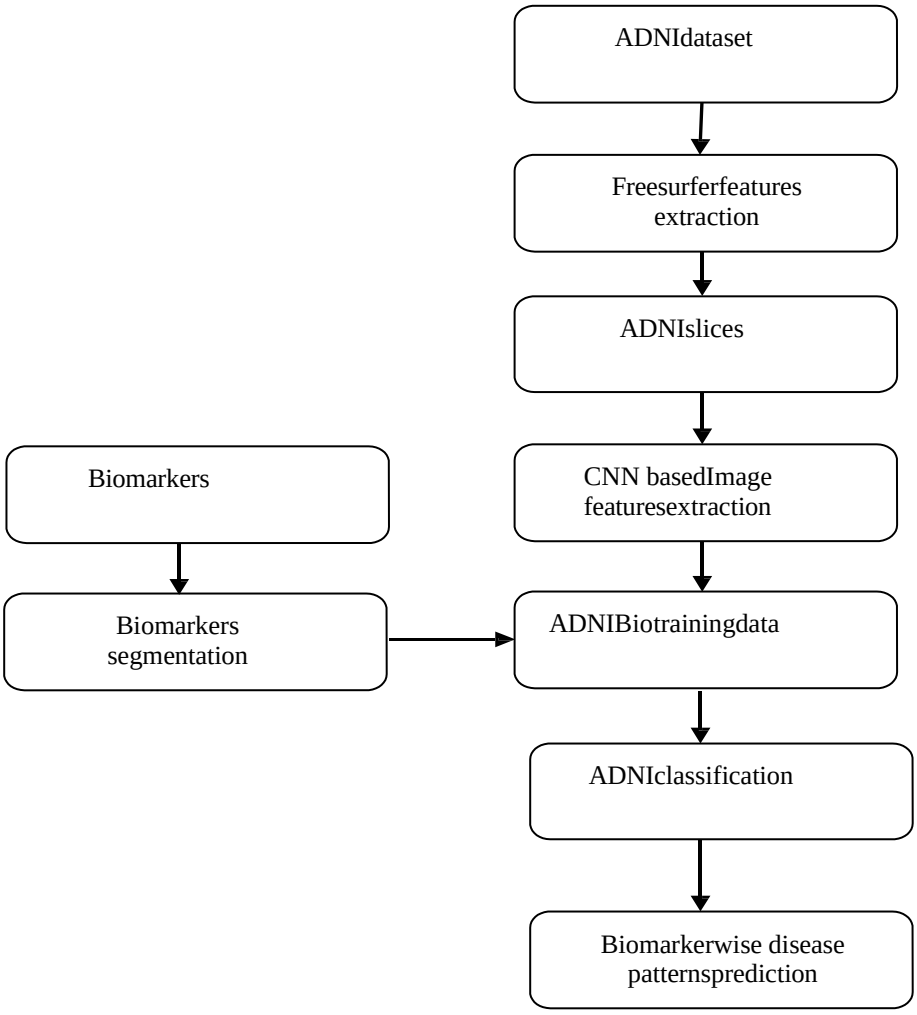

Figure 1: Overview of proposed model

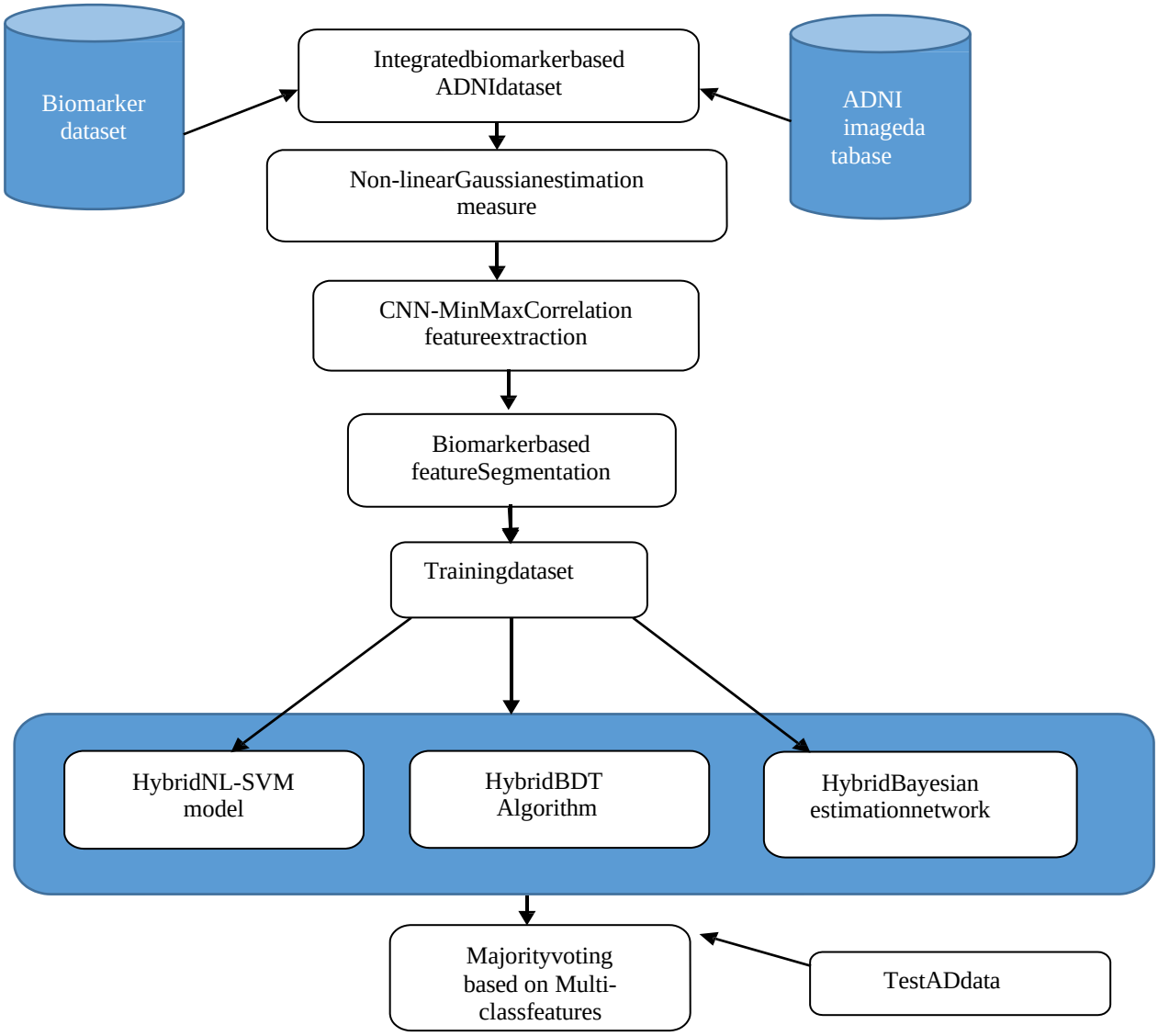

Figure2:Proposed Model of biomarkers based ADNI disease prediction

Algorithm1:A Novel feature extraction measure in CNN framework

Step 1: Input ADNI images.

Step 2: Filter the ADNI images using Subject ID.

Step 3: Non-linear Gaussian estimation using histogram H is computed as:

$$NLG = \max\{H.\log(e^{-|2\log(H)|}), \sum_{i=0}^{N} \frac{1}{\sqrt{2\pi\log(H)}}.e^{-|I[i]-H|/2}\} \quad ----(1)$$

Step 4: For each subject id in Image database

Do

For each block in image id

Do

Compute Block histogram intensities I.

Compute average block histogram value H using eq.(1).

Step 5: Apply Maximized correlation and kernel estimator measures for efficient correlated feature extraction process as

$$\text{MMC=MinMaxCorrelation} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \max \{ \frac{\{i \times j\} \times \log (P(i,j)) - \{\min\{\mu_x, \mu_{y\}}\}\}}{\min \{\sigma_x, \sigma_{y\}}},$$

$$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{i-j\}^2 \times \log (P(i,j)\}$$

Nonlinear kernel estimator for feature extraction

NLKE is used to find the non-linear structure of the image patterns using the gaussian distribution measure. kernel Gaussian estimation using co-occurrence matrix P is computed as:

$$\text{NLKE} = \sum_{i=0}^{N} \frac{1}{\sqrt{2\pi P(i,j)}} . e^{-|\log ((p(I,J))-\mu_x \times \mu_y)^2 / 2(.\sigma_x \times \sigma_{y)|}}$$

In the CNN framework, the maximization of MMC and NLKE measures are used to find the best feature selection for biomarker based ADNI disease prediction.

Done

Done

Algorithm 2: Biomarker segmentation algorithm

Expectation phase (E-phase) : In the expectation phase, model parameters are estimated using the hybrid probabilistic measure. This probabilistic measure is used to find the essential key features for the biomarker-disease clustering.

Let $\hat{\theta}$ represent the novel posterior estimation parameter used to predict the occurrence of biomarker disease pattern in the given large number of training samples.

$\text{Prob}(\hat{\theta}/c)$ is the occurrence of biomarker-disease new patterns in the large category of

disease classes.

In the expectation phase, the maximization of the biomarker -disease patterns in all the training real-time datasets are given as :

$$\Pr ob(\hat{\theta}/c_i) = \frac{\max\{\sigma_k(BD_i, C_i), \log(BD_i . C_j)\}.\sum_{i=1}^{N} P(c_i / D(BD)_i)}{|N|.P(c_i / BD)}$$

Maximization phase: In the maximization phase, model parameters are estimated using the biomarker -disease and its class patterns. In the maximization step, the probability of data occurrence in the given disease class is given as

$$\Pr ob(c / BD_i, \hat{\theta}) = \frac{P(BD_i / c_i).\log(|c|).P(\hat{\theta})}{\Pr ob(c/ D_i)}.$$

These two phases are repeated until the number of maximum iterations or no change in the error rate.

Proposed Classification Algorithm

1: Read pre-processing biomarker -disease patterns, training alzhemidatasets, biomarker database. All these input patterns are partitioned 'm' clusters based on EM approach.

2: To each clustered

3: do

4: Apply proposed ensemble decision tree model on each cluster data.

5: In the proposed classification model, a novel biomarker -disease feature selection measure is implemented on each cluster.

Proposed decision tree biomarker based disease prediction

Let D_p represents the clustered gene-disease probabilistic patterns for the decision tree classification problem. The hoeffding entropy of the biomarker -disease pattern analysis is given by

$$n = \sum D_p[i].\log(D_P[i])$$

$$Ent(D_p) = \frac{(n + D_p[i].\min\{D_p\}\{(\sum \log(D_p[i])}{\sqrt[3]{(\sum \log(D_p[i]).\mu_{D_{p[i]}})}}$$

Proposed Biomarker based Joint Bayesian Probability score

$\theta = \text{Conditional PriorProb}(s_i);$

$\phi = \text{Joint Prob}(D / s_i);$

$\text{PropBayesScore} = \log(\theta) + \log(\phi)$

where

$$\text{Joint Prob}(D / s_i) = \prod_{i=0}^{n}\prod_{j=0}^{q_i} \frac{\Gamma(\sum_{k=1}^{r} \alpha_{ijk} \tan(\alpha_{ijk}))}{\Gamma(\sum_{k=1}^{r} \alpha_{ijk} \exp(\alpha_{ijk}) + \sum \log(N_{ij}))} \prod_{k=1}^{r} \frac{\Gamma(\sum_{k=1}^{r} e^{(\alpha_{ijk})} \log(\alpha_{ijk}) + \Gamma\alpha_{ijk} \cdot \sum \log(N_{ijk}))}{\Gamma(\sum_{k=1}^{r} \alpha_{ijk} \log(\alpha_{ijk}))}$$

Input the CNN features for data classification.

For each feature set do

for each disease in GDP.

do

Apply SVM multi-class optimization models as

$$\min_{W_k, a_k} \frac{1}{2}\|W_k\|_1^2 + \tau_m + \sum_{i=1}^{l} a_i \left(y_i \left[ker < x, y > \cdot w + b \right] - 1 + \xi_i \right) - \sum_{i=1}^{l} \gamma_i \xi_i$$

$$s.t\ ker < x, y > \cdot w + b \geq 1 - \xi_i^n - \tau_m,$$

$$\xi_i^n > 0$$

$$\tau_m > 0; m = 1 \ldots classes$$

Here kernel function ker(x,y) represents the kernel functions defined from biomarker disease vector space to chemical symbol vector space.

$$Ker < x, y > = e^{-\xi_i^n \log\left(\sum \| x - y \|^2\right)} \quad \text{if x==y}$$

$$= e^{-\xi_i^n \log\left(\sum \| x - y \|^{1/2}\right)} \quad \text{if x<y}$$

$$= e^{-\xi_i^n \log\left(\sum \| y \|^2\right)} \quad \text{if x > y}$$

Step 4: Test data is predicted to the class y based on the largest decision values as

$$arg\ max\{ W_k^T D_i + b_k \}$$

4.EXPERIMENTALRESULTS

Experimental results are simulated in python and java environment. In this experimental study, different ADNI images[23] and biomarkers[24] are taken to find the disease related patterns in the large data. In this work, different biometric features and it associations are used to predict the severity in the ADNI images. Initially, ADNI images are pre-processed and sliced using the freesurfer tool. The experimental results and its performance analysis are presented in this section.

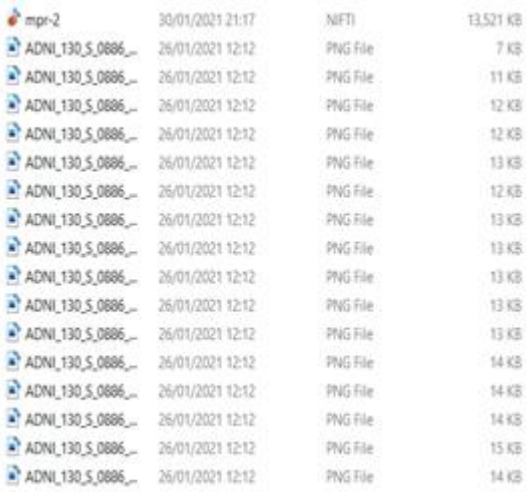

Figure 3: Sample ADNI image and its slices

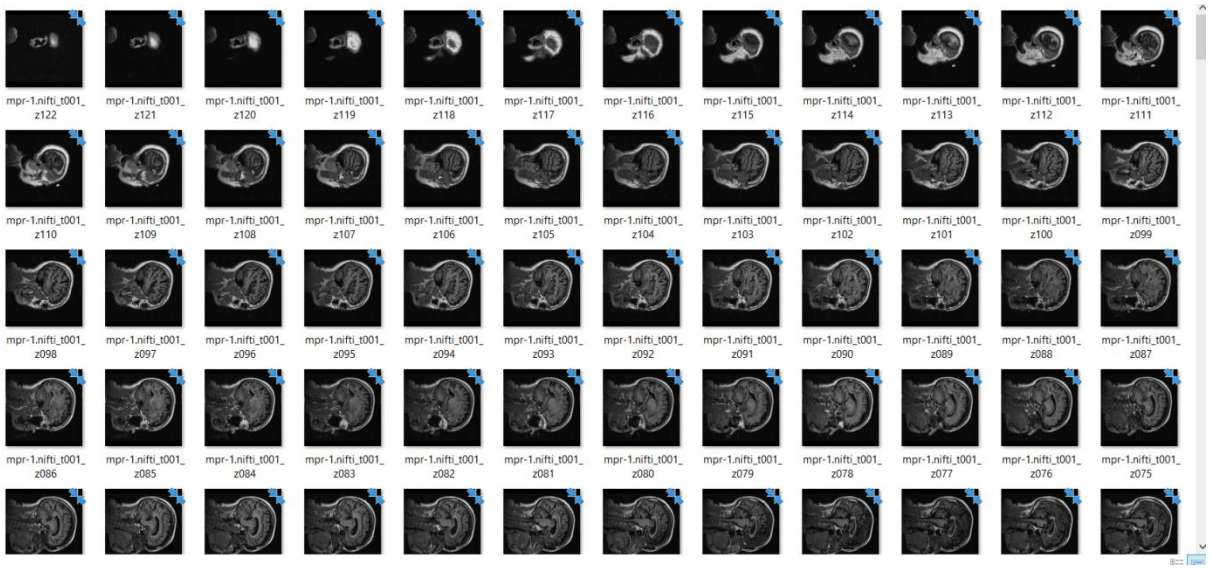

Figure 4: ADNI slices

Experimental results of biomarker based ADNI disease patterns on high dimensional databases

[APOE4=0]: 447 ==> [PTGENDER=2, PTMARRY=1]: 183
[PTGENDER=1, PTMARRY=1, PTETHCAT=1]: 234 ==> [PTRACCAT=2, FAQ=0]: 96
[APOE4=0, PTETHCAT=1]: 424 ==> [PTGENDER=2, PTMARRY=1]: 174
[PTETHCAT=1]: 744 ==> [PTGENDER=2, PTRACCAT=2, PTMARRY=1]: 306
[CDRSB=0, FAQ=0]: 233 ==> [PTMARRY=1, APOE4=0, PTETHCAT=1]: 96
[PTMARRY=1, FAQ=0, PTETHCAT=1]: 245 ==> [PTGENDER=1]: 101
[PTMARRY=1, PTETHCAT=1]: 555 ==> [PTRACCAT=2, FAQ=0]: 229
[PTMARRY=1, APOE4=0]: 312 ==> [PTGENDER=1]: 129
[PTRACCAT=2, PTMARRY=1, PTETHCAT=1]: 522 ==> [PTGENDER=1]: 216
[PTGENDER=2, PTMARRY=1, PTETHCAT=1]: 321 ==> [PTRACCAT=2, FAQ=0]: 133
[CDRSB=0]: 265 ==> [PTRACCAT=2, PTMARRY=1, APOE4=0, PTETHCAT=1]: 110
[PTMARRY=1]: 580 ==> [PTRACCAT=2, FAQ=0]: 241
[CDRSB=0, PTETHCAT=1]: 243 ==> [PTGENDER=2, PTRACCAT=2]: 101
[PTRACCAT=2, FAQ=0, PTETHCAT=1]: 335 ==> [PTMARRY=1, APOE4=0]: 140
[PTRACCAT=2]: 732 ==> [PTGENDER=2, PTMARRY=1, PTETHCAT=1]: 306
[PTMARRY=1, APOE4=0, PTETHCAT=1]: 299 ==> [PTGENDER=1]: 125
[PTGENDER=2]: 397 ==> [PTRACCAT=2, PTMARRY=1, APOE4=0, PTETHCAT=1]: 166
[APOE4=1, PTETHCAT=1]: 239 ==> [FAQ=0]: 100
[PTRACCAT=2, CDRSB=0]: 241 ==> [PTGENDER=2, PTETHCAT=1]: 101
[PTRACCAT=2, PTMARRY=1]: 544 ==> [PTGENDER=1]: 228
[PTRACCAT=2, PTMARRY=1, FAQ=0, PTETHCAT=1]: 229 ==> [PTGENDER=1]: 96
[APOE4=0]: 447 ==> [PTGENDER=1, PTRACCAT=2, PTETHCAT=1]: 188
[APOE4=1]: 259 ==> [PTGENDER=2, PTMARRY=1, PTETHCAT=1]: 109
[PTGENDER=2, PTRACCAT=2, PTMARRY=1]: 316 ==> [FAQ=0, PTETHCAT=1]: 133
[PTRACCAT=2, PTMARRY=1]: 544 ==> [FAQ=0, PTETHCAT=1]: 229
[PTGENDER=1, PTRACCAT=2, PTMARRY=1]: 228 ==> [FAQ=0, PTETHCAT=1]: 96
[PTRACCAT=2, APOE4=1]: 242 ==> [PTGENDER=2, PTMARRY=1, PTETHCAT=1]: 102
[PTRACCAT=2, APOE4=1]: 242 ==> [FAQ=0]: 102
[PTMARRY=1, PTETHCAT=1]: 555 ==> [PTGENDER=1]: 234
[APOE4=0, FAQ=0]: 251 ==> [PTGENDER=2, PTRACCAT=2]: 106
[PTRACCAT=2, APOE4=0]: 412 ==> [PTGENDER=2, PTMARRY=1]: 174
[PTMARRY=1]: 580 ==> [FAQ=0, PTETHCAT=1]: 245
[CDRSB=0]: 265 ==> [PTRACCAT=2, PTMARRY=1, APOE4=0]: 112
[CDRSB=0]: 265 ==> [PTGENDER=1, FAQ=0, PTETHCAT=1]: 112

[PTGENDER=2]: 397 ==> [PTRACCAT=2, FAQ=0, PTETHCAT=1]: 168
[PTRACCAT=2, APOE4=0, PTETHCAT=1]: 392 ==> [PTGENDER=2, PTMARRY=1]: 166
[PTRACCAT=2, APOE4=0, PTETHCAT=1]: 392 ==> [CDRSB=0]: 166
[PTRACCAT=2, APOE4=0]: 412 ==> [CDRSB=0]: 175
[CDRSB=0, FAQ=0]: 233 ==> [PTGENDER=2, PTETHCAT=1]: 99
[PTGENDER=1, APOE4=0]: 221 ==> [CDRSB=0, PTETHCAT=1]: 94
[APOE4=1, PTETHCAT=1]: 239 ==> [PTGENDER=2, PTRACCAT=2, PTMARRY=1]: 102
[FAQ=0]: 391 ==> [PTGENDER=1, PTRACCAT=2, PTETHCAT=1]: 167
[FAQ=0]: 391 ==> [APOE4=0, CDRSB=0]: 167
[PTGENDER=2, PTRACCAT=2, PTMARRY=1]: 316 ==> [FAQ=0]: 135
[PTGENDER=1, PTMARRY=1]: 248 ==> [PTRACCAT=2, FAQ=0]: 106
[PTMARRY=1]: 580 ==> [PTGENDER=1]: 248
[PTGENDER=1]: 390 ==> [PTRACCAT=2, FAQ=0, PTETHCAT=1]: 167
[APOE4=1]: 259 ==> [PTGENDER=2, PTMARRY=1]: 111
[FAQ=0, PTETHCAT=1]: 364 ==> [APOE4=0, CDRSB=0]: 156
[PTRACCAT=2, FAQ=0]: 359 ==> [APOE4=0, CDRSB=0]: 154
[CDRSB=0, FAQ=0]: 233 ==> [PTMARRY=1, APOE4=0]: 100
[APOE4=0, PTETHCAT=1]: 424 ==> [CDRSB=0]: 182
[FAQ=0]: 391 ==> [PTGENDER=2, PTRACCAT=2, PTETHCAT=1]: 168
[PTRACCAT=2, APOE4=1]: 242 ==> [PTGENDER=2, PTMARRY=1]: 104
[CDRSB=0]: 265 ==> [PTGENDER=2, PTETHCAT=1]: 114
[APOE4=0, FAQ=0]: 251 ==> [PTGENDER=2, PTETHCAT=1]: 108
[APOE4=0, FAQ=0, PTETHCAT=1]: 237 ==> [PTGENDER=2, PTRACCAT=2]: 102
[PTETHCAT=1]: 744 ==> [PTGENDER=2, PTMARRY=1]:
321
[PTGENDER=1, PTMARRY=1, PTETHCAT=1]: 234 ==> [FAQ=0]: 101
[PTRACCAT=2]: 732 ==> [PTGENDER=2, PTMARRY=1]: 316
[APOE4=0]: 447 ==> [CDRSB=0]: 193
[CDRSB=0, PTETHCAT=1]: 243 ==> [PTGENDER=1, PTRACCAT=2, FAQ=0]: 105
[PTRACCAT=2, FAQ=0, PTETHCAT=1]: 335 ==> [APOE4=0, CDRSB=0]: 145
[PTGENDER=2]: 397 ==> [PTRACCAT=2, FAQ=0]: 172
[PTGENDER=2, PTMARRY=1]: 332 ==> [FAQ=0, PTETHCAT=1]: 144
[PTRACCAT=2, APOE4=1]: 242 ==> [PTGENDER=1, PTETHCAT=1]: 105
[PTGENDER=2, PTETHCAT=1]: 382 ==> [PTRACCAT=2, PTMARRY=1, APOE4=0]: 166
[PTMARRY=1, FAQ=0]: 260 ==> [PTGENDER=1]: 113
[PTGENDER=2, PTRACCAT=2, PTMARRY=1, PTETHCAT=1]: 306 ==> [FAQ=0]: 133
[PTRACCAT=2, CDRSB=0]: 241 ==> [PTGENDER=1, FAQ=0, PTETHCAT=1]: 105
[PTRACCAT=2, CDRSB=0]: 241 ==> [PTGENDER=2]: 105

[APOE4=1]: 259 ==> [FAQ=0]: 113
[PTGENDER=2]: 397 ==> [PTMARRY=1, APOE4=0, PTETHCAT=1]: 174
[PTGENDER=2]: 397 ==> [PTRACCAT=2, PTMARRY=1, APOE4=0]: 174
[PTRACCAT=2, PTMARRY=1, PTETHCAT=1]: 522 ==> [FAQ=0]: 229
[PTGENDER=2, PTRACCAT=2]: 378 ==> [PTMARRY=1, APOE4=0, PTETHCAT=1]: 166
[PTRACCAT=2, CDRSB=0, FAQ=0]: 214 ==> [PTMARRY=1, APOE4=0]: 94
[APOE4=1, PTETHCAT=1]: 239 ==> [PTGENDER=1, PTRACCAT=2]: 105
[PTGENDER=2, PTETHCAT=1]: 382 ==> [PTRACCAT=2, FAQ=0]: 168
[PTRACCAT=2, PTMARRY=1, FAQ=0]: 241 ==> [PTGENDER=1]: 106
[FAQ=0]: 391 ==> [PTGENDER=2, PTRACCAT=2]: 172
[APOE4=1]: 259 ==> [PTGENDER=1, PTETHCAT=1]: 114
[PTMARRY=1, PTETHCAT=1]: 555 ==> [FAQ=0]: 245
[PTRACCAT=2, PTETHCAT=1]: 693 ==> [PTGENDER=2, PTMARRY=1]: 306
[CDRSB=0, FAQ=0]: 233 ==> [PTGENDER=2]: 103
[PTETHCAT=1]: 744 ==> [PTGENDER=1, PTRACCAT=2]: 329
[PTGENDER=2, PTMARRY=1]: 332 ==> [FAQ=0]: 147
[APOE4=0]: 447 ==> [PTGENDER=1, PTRACCAT=2]: 198
[PTRACCAT=2, PTMARRY=1]: 544 ==> [FAQ=0]: 241
[APOE4=0, PTETHCAT=1]: 424 ==> [PTGENDER=1, PTRACCAT=2]: 188
[PTRACCAT=2, APOE4=0, FAQ=0]: 230 ==> [PTGENDER=2, PTETHCAT=1]: 102
[PTGENDER=1, PTRACCAT=2, PTMARRY=1, PTETHCAT=1]: 216 ==> [FAQ=0]: 96
[PTGENDER=2, PTRACCAT=2]: 378 ==> [FAQ=0, PTETHCAT=1]: 168
[CDRSB=0]: 265 ==> [PTGENDER=2]: 118
[PTMARRY=1]: 580 ==> [FAQ=0]: 260
[PTGENDER=2, PTMARRY=1, PTETHCAT=1]: 321 ==> [FAQ=0]: 144
[PTMARRY=1, APOE4=0]: 312 ==> [PTRACCAT=2,
FAQ=0, PTETHCAT=1]: 140
[CDRSB=0]: 265 ==> [PTMARRY=1, APOE4=0, PTETHCAT=1]: 119
[PTRACCAT=2]: 732 ==> [PTGENDER=1, PTETHCAT=1]: 329

Table 1, describes the different combinations of biomarker-based disease patterns on large ADNI image features. These patterns are used to find the combinations of various biomarker patterns on the training dataset.

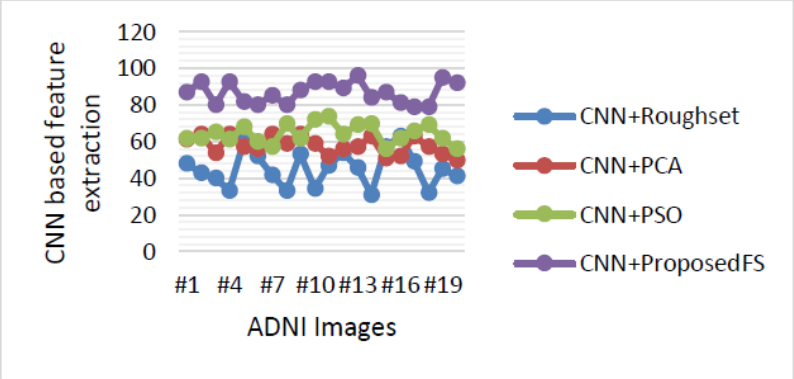

Figure 5: Performance analysis of proposed feature-based segmentation and classification model on the input biomarker based ADNI feature sets. As represented in the figure, proposed biomarker based ADNI model has better improvement in terms of feature selection on image slices.

ADNIImage	CNN+Roughset	CNN+PCA	CNN+PSO	CNN+ProposedFS
#1	5229	5879	5272	4098
#2	6363	6127	6231	3541
#3	6336	7122	7417	3358
#4	7489	5979	6228	4679
#5	7565	6730	7198	4608
#6	7505	6790	6247	3330
#7	7855	7518	6525	3894
#8	7114	6113	6045	3512
#9	5030	6853	7822	3684
#10	5869	7600	5943	3559
#11	5133	7042	7391	3686
#12	5366	7683	5197	3798
#13	5353	7849	6500	3998
#14	5027	5726	5020	3506
#15	5857	5446	5644	3566
#16	6103	5161	7769	4529
#17	6487	5652	5295	4102
#18	5970	5268	7410	3628
#19	5051	5773	5747	4686
#20	5483	6821	7066	3577

Table 2, describes the runtime computation of the proposed framework on different ADNI image slices. From the table, it is noted that the present framework has better runtime than the conventional models.

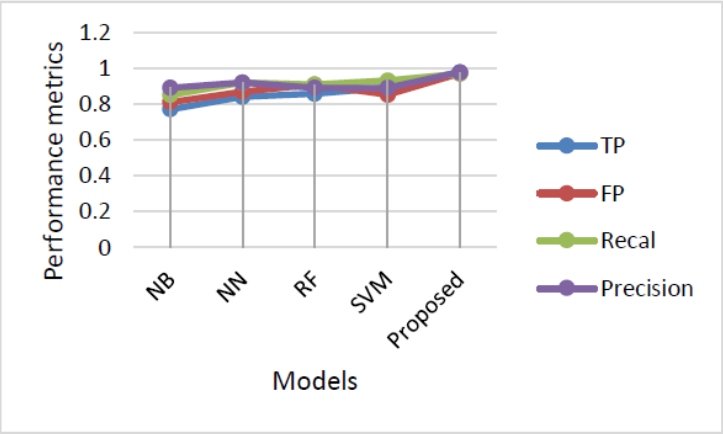

Figure 6: Comparative study of proposed biomarker based ADNI framework based on accuracy computation

CONCLUSIONS

In this work, an advanced biomarker based Alzheimer disease prediction framework is designed and implemented on the ADNI databases. In this work, a novel feature selection based segmentation and classification model is implemented on the input biomarker ADNI features. In this work, an improved CNN based feature selection method, a segmentation model and classification model are implemented on the large number of feature space and biomarkers. Current algorithms are tested and evaluated; an improved set feature selection method is proposed with re-sampling strategies. Experimental results proved that the present CNN feature selection-based segmentation and classification model has better prediction rate than the conventional models on high dimensional features.

REFERENCES

[1]A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, "Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease," Journal of Neuroscience Methods, vol. 339, p. 108701, Jun. 2020, doi: 10.1016/j.jneumeth.2020.108701.

[2]R. Achalia et al., "A proof of concept machine learning analysis using multimodal neuroimaging and neurocognitive measures as predictive biomarker in bipolar disorder," Asian Journal of Psychiatry, vol. 50, p. 101984, Apr. 2020, doi: 10.1016/j.ajp.2020.101984.

[3]O. B. Ahmed, J. Benois-Pineau, M. Allard, G. Catheline, and C. B. Amar, "Recognition of Alzheimer's disease and Mild Cognitive Impairment with multimodal image-derived biomarkers and Multiple Kernel Learning," Neurocomputing, vol. 220, pp. 98–110, Jan. 2017, doi: 10.1016/j.neucom.2016.08.041.

[4]N. An, H. Ding, J. Yang, R. Au, and T. F. A. Ang, "Deep ensemble learning for Alzheimer's disease classification," Journal of Biomedical Informatics, vol. 105, p. 103411, May 2020, doi: 10.1016/j.jbi.2020.103411.

[5]M. B. Bachli et al., "Evaluating the reliability of neurocognitive biomarkers of neurodegenerative diseases across countries: A machine learning approach," NeuroImage, vol. 208, p. 116456, Mar. 2020, doi: 10.1016/j.neuroimage.2019.116456.

[7]M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," Computer Methods and Programs in Biomedicine, vol. 187, p. 105242, Apr. 2020, doi: 10.1016/j.cmpb.2019.105242.

[8]P. D. Fransquet and J. Ryan, "Micro RNA as a potential blood-based epigenetic biomarker for Alzheimer's disease," Clinical Biochemistry, vol. 58, pp. 5–14, Aug. 2018, doi: 10.1016/j.clinbiochem.2018.05.020.

[9]R. Gaudiuso, E. Ewusi-Annan, W. Xia, and N. Melikechi, "Diagnosis of Alzheimer's disease using laser-induced breakdown spectroscopy and machine learning," Spectrochimica Acta Part B: Atomic Spectroscopy, p. 105931, Jul. 2020, doi: 10.1016/j.sab.2020.105931.

[10]D. Goyal, D. Tjandra, R. Q. Migrino, B. Giordani, Z. Syed, and J. Wiens, "Characterizing heterogeneity in the progression of Alzheimer's disease using longitudinal clinical and neuroimaging biomarkers," Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, vol. 10, pp. 629– 637, Jan. 2018, doi: 10.1016/j.dadm.2018.06.007.

[11]A. Khan and S. Zubair, "An Improved Multi-Modal based Machine Learning Approach for the Prognosis of Alzheimer's disease," Journal of King Saud University - Computer and Information Sciences, Apr. 2020, doi: 10.1016/j.jksuci.2020.04.004.

[12]S. Lahmiri and A. Shmuel, "Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease," Biomedical Signal Processing and Control, vol. 52, pp. 414–419, Jul. 2019, doi: 10.1016/j.bspc.2018.08.009.

[13]C. S. Lee and R. S. Apte, "Retinal Biomarkers of Alzheimer Disease," American Journal of Ophthalmology, May 2020, doi: 10.1016/j.ajo.2020.04.040.

[14]H. Li, M. Habes, D. A. Wolk, and Y. Fan, "A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data," Alzheimer's & Dementia, vol. 15, no. 8, pp. 1059–1070, Aug. 2019, doi: 10.1016/j.jalz.2019.02.007.

[15]L. Liu, S. Zhao, H. Chen, and A. Wang, "A new machine learning method for identifying Alzheimer's disease," Simulation Modelling Practice and Theory, vol. 99, p. 102023, Feb. 2020, doi: 10.1016/j.simpat.2019.102023.

[16]N. Ludwig et al., "Machine Learning to Detect Alzheimer's Disease from Circulating Non-coding RNAs," Genomics, Proteomics & Bioinformatics, vol. 17, no. 4, pp. 430–440, Aug. 2019, doi: 10.1016/j.gpb.2019.09.004.

[17]G. Martí-Juan, G. Sanroma-Guell, and G. Piella, "A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's disease," Computer Methods and Programs in Biomedicine, vol. 189, p. 105348, Jun. 2020, doi: 10.1016/j.cmpb.2020.105348.

[18]V. P. S. Rallabandi, K. Tulpule, and M. Gattu, "Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer's disease using structural MRI analysis," Informatics in Medicine Unlocked, vol. 18, p. 100305, Jan. 2020, doi: 10.1016/j.imu.2020.100305.

[19]M. Raza, M. Awais, W. Ellahi, N. Aslam, H. X. Nguyen, and H. Le-Minh, "Diagnosis and monitoring of Alzheimer's patients using classical and deep learning techniques," Expert Systems with Applications, vol. 136, pp. 353–364, Dec. 2019, doi: 10.1016/j.eswa.2019.06.038.

[20]D. Stamate et al., "A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: Results from the European Medical Information Framework for Alzheimer disease biomarker discovery cohort," Alzheimer's & Dementia: Translational Research & Clinical Interventions, vol. 5, pp. 933–938, Jan. 2019, doi: 10.1016/j.trci.2019.11.001.

[21]G. Uysal and M. Ozturk, "Hippocampal atrophy based Alzheimer's disease diagnosis via machine learning methods," Journal of Neuroscience Methods, vol. 337, p. 108669, May 2020, doi: 10.1016/j.jneumeth.2020.108669.

[22]J. J. van der Zande, M. Dauwan, E. Van Dellen, P. Scheltens, A. W. Lemstra, and C. J. Stam, "APPLYING RANDOM FOREST MACHINE LEARNING TO DIAGNOSE ALZHEIMER'S DISEASE AND DEMENTIA WITH LEWY BODIES: A COMBINATION OF ELECTROENCEPHALOGRAPHY

(EEG), CLINICAL PARAMETERS AND BIOMARKERS," Alzheimer's &Dementia, vol. 12, no. 7, Supplement, pp. P661–P662, Jul. 2016, doi: 10.1016/j.jalz.2016.06.1501.

[23]D. Zafeiris, S. Rutella, and G. R. Ball, "An Artificial Neural Network Integrated Pipeline for Biomarker Discovery Using Alzheimer's Disease as a Case Study," Computational and Structural Biotechnology Journal, vol. 16, pp. 77–87, Jan. 2018, doi: 10.1016/j.csbj.2018.02.001.

[24] http://adni.loni.usc.edu/

[25] https://tadpole.grand-challenge.org/Data/