Original Research Article

J. Computing & Int. Systems (2018) 52–56

SACRED HEART RESEARCH PUBLICATIONS

Journal of Computing and Intelligent Systems

Journal homepage: www.shcpub.edu.in

ISSN: 2456-9496

ALZHEIMER'S ILLNESS PROGNOSIS USING HYBRID GENETIC FEATURE SELECTION AND SUPPORT VECTOR MACHINE CLASSIFICATION

V. Nandhini[#1], R. Pugazendi [#2]

Received on 23rd Aug 2018, Accepted on 27th Nov 2018

Abstract *— Alzheimer's Illness (AI) is the general method of mental sickness, enduring a large number of aged individuals around the world. It is an age-related mind disease that gradually endures the individual memory and conceptual abilities and fitness even the aptitudes to misbehave to least demanding exercises. Be that as it may, commonly the memory and remaining idea issues are escalated and are hard to do regular assignment. Alzheimer's illness (AI) is the most stressed activity in the contemporary world their correct estimate can't be made independently. A computerized framework requires to be extended for the estimate for this disease successful and fast method. So, in this paper introduces the AI detection process using hybridized method called genetic feature selection process and support vector machine based classification algorithm. There are distinctive neuropsychological tests, the different calculations utilized with the end goal of conclusion and the apparatus that might be utilized for the investigation to discover Alzheimer's illness. NACC (National Alzheimer's Coordinating Center) having the Researchers information Dictionary – Uniform Data Set(RDD_UDS)is gives dataset for researchers. Thus the work gives better exactness and a superior outcome.*

Keywords: Alzheimer's illness, diagnosis methods, Data mining techniques, SVM, Genetic Algorithm, NAAC, RDD_UDS.

I INTRODUCTION

Data mining is a most supreme technology which is having high latent to support inventing or discovering of hidden data from large datasets. These extracted data's are load in a data warehouse which are stored and preserved into multidimensional databases. In recent days, the applications of data mining in healthcare systems play an important role because the health field comprises rich information and it developed into a necessary technology. This technology is made use to mine the information from the database at any time, which requires for processing. The comprehending of this mechanisms leads to develop the effective and improve the feature of disease analysis in the certain datasets. Dementia is a kind of disorder it contains many different set of symptoms they are functional degradation in sense of space, memory loss, decision taken, manipulating ability, abstract thinking, and attention.

the scientific community has shown increasing interest towards the prediction of (Alzheimer illness) AI and MCI due to its impact in diagnosis and treatment of this disease. This paper aims to develop a reliable automatic diagnostic tool against normal controls in the context of Multiple-Instance Learning, a flourishing new area in machine learning applied, showing improved results against the usual cerebral global mean. Most importantly, regarding classification, two MIL algorithms belonging to two different sub paradigms - Instance-space, Bag-space and Vocabulary based - were applied along with the renowned supervised Support Vector Machines (SVM) is applied to classify the Alzheimer disease with effective manner. Before processing the Alzheimer data, the dimensionality of the dataset is further reduced by applying genetic algorithm for improving the efficiency of AI detection process. In addition, a longitudinal evaluation using support vector machine is compared with the traditional forecasting algorithms such as naïve classifier, decision tree, k-nearest neighboring approach was performed, which allowed a comparison with cross-sectional analysis in the baseline data, and a simulation of data classification was performed with MATLAB.

The rest of this paper is structured as follows. Section 2 addresses literature study of the Alzheimer's disease using various data mining techniques. In Section 3, a thorough overview of applied techniques such as genetic feature selection and support vector AI classification process and its algorithms steps as Section 4, there is a minute description of the methodology and classification procedure, with a focus on the explanation of the algorithms and their underlying ideas.

* Corresponding author: E-mail: nandhudhyana14@gmail.com, pugazendi_r@gmail.com.

[1]Research Scholar, Dept. of Computer Science, Government Arts College, Periyar University, Salem – 7, Tamilnadu, India.

[2]Assistant Professor, Dept. of Computer Science, Government Arts College, Periyar University, Salem – 7, Tamilnadu, India.

II RELATED WORK

Ramesh Kumar Lama et al [1] exhibited and thought about AD forecast approaches making utilization of structural Magnetic Resonance (SMR) pictures to separate AD, Mild Cognitive Impairment (MCI), and Healthy Control (HC) subjects utilizing a Support Vector Machine (SVM), and Import Vector Machine (IVM), and a Regularized Extreme Learning Machine (RELM). The ravenous score based element determination component is utilized to choose important feature vectors. Also, a portion based discriminative approach is embraced to manage complex information dissemination. The creators thought about the execution of the classifier for volumetric sMR picture information from Alzheimer's Disease Neuroimaging activity (ADNI) datasets. Tests on the ADNI datasets exhibited that RELM with the component determination approach can significantly enhance characterization precision of an AD from MCI and HC subjects.

J.M.Gorriz et al [2] presented another arrangement strategy of SPECT pictures as indicated by on Gaussian Mixture Models (GMM) for the forecast of Alzheimer's Disease (AD). The objective of the model-based concentration for density approximation is to regular select territories of intrigue and to viably lessen the dimensionality of the issue. The result of GMM are manufactured relies upon a most customary measure conveying the Expectation Maximization (EM) calculation. By observing just the seriously heights the GMM, the result of highlight space has an imperatively lessened dimensionality concerning effortlessly centered by making utilization of the voxel powers straightly as highlights (VAF). With this feature extraction method, one remembers the result of the alleged small example measure issue and nonlinear classifiers might be connected to separate among the cerebral pictures of normal Alzheimer's patients. Thus, the paperresult showed that distinctive classifiers GMM-based procedure bringing about preferable exactness rates rather over the classification on observing all voxel esteems.

Mohit Jain et al [3] displayed information versatile control based classification framework for Alzheimer's disease grouping that creates appropriate standards by finding adaptive partitions making utilization of angle based dividing of the information. The dynamic partition are created from the histogram by examining Tuples Tests following which successful and apropos decides have found that assistance in arranging the new information precisely. The authors' proposed approach has been contrasted and other rule based and machine learning classifiers, and itemized results, dialogs of the tests uncovered to demonstrate similar investigation and effectiveness of the outcomes.

F. Segovia et al [4] examined in this paper and showed an effective concentration for Single Photon Emission Computed Tomography (SPECT) feature extraction that builds up the precision of PC helped design frameworks for Alzheimer Disease (AD).

The exhibited system makes utilize a Partial Least Squares algorithm (PLS) for mining score vectors and the Out-Of-Bag mistake for picking various scores that are connected in features. Support Vector Machine (SVM) classifier decides the basic class of the pictures, consequently accomplishing prognostics. Keeping in mind the end goal to explore this concentration, the paper have utilized animage database for an AD with 97 SPECT pictures from controls and AD patients. The pictures unmistakably depicted by developed to analyze the exhibited technique and past methodologies. The outcomes showed numerous as of late detailed PC helped frameworks for AD analysis.

M.Lopez et al [5] proposed a total Computer Aided Design Designs (CAD) framework for programme assessment of the Neuroimaging. Principle component analysis (PCA) based strategy is proposed as feature extraction methods, improved by linear direct methodologies, for example, straight discriminant analysis (LDA) or the measure of the Fisher discriminant proportion (FDR) for include determination. The last feature permit looking up the purported little example measure issue and therefore they are utilized for the investigation of neural Networks (NN) and Support vector machine (SVM) classifiers. The blend of the exhibited strategies accomplished precision results up to 96.7% and 89.52% SPECT and PET pictures, individually, which implies a critical change over the outcomes acquired by the traditional voxels as highlights (VAF) reference approach.

III RESEARCH METHODOLOGY

In this section discusses about the genetic algorithm based feature selection and support vector machine Alzheimer disease detection process. Then the processing AI disease process is shown in figure 1.

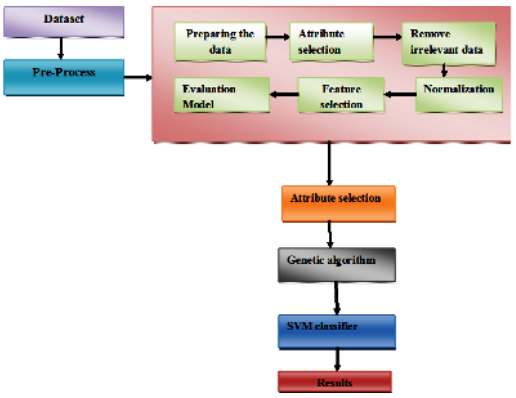

Figure 1 -Alzheimer disease Recognition System Structure

A. Data pre-processing

Alzheimer's dataset has been taken from NACC database and applied preprocessing which is followed by the data preparation, attribute selection, removing missing values, noise data, irrelevant data then the step is moving to normalization with correlation and coefficient calculations then step is moving to evaluation model.

The noise removal process is done by applying normalization process that eliminates the irrelevant information and effectively replace the missing values. Then the normalization process is done as follows.

$$N' = \left(\frac{N - \min value\ of\ N}{Max\ value\ of\ N - Min\ value\ of\ N}\right) \quad (1)$$

In eqn (1) N' is normalized data.

N is particular data in dataset.

Based on the above process, irrelevant data is removed and if the dataset having any missing values that is replaced by the computed values. After performing the preprocessing process, the dimensionality of the dataset must be reduced for improving the efficiency of Alzheimer disease recognition process. The fittest attributes selection over then applying genetic algorithm is applied the continuous process to classify AD dataset.

B. Feature Selection using Genetic Algorithm

The next step is feature selection that is done by applying genetic algorithm which utilizes three different operators such as selection, crossover and mutation. The selection process analyze the collected data set attributes and chose the good attributes from the collection of Alzheimer dataset. The selection is done by as follows.

$$P_{selection} = \frac{E_c}{\sum_{i=1}^{M} E_c} \quad (2)$$

In eqn (2), $P_{selection}$ is represented as the selected features probability value.$E_c$is represented as the particular data attributes, M is denoted as number of data attributes. After selecting particular attributes, crossover process is performed to convene the selected attributes to produces the best attributes. Finally, mutation operator is applied to generated new attributes which consists of binary numbers 1 and 0. 1 represented as the selected attributes from dataset and 0 denoted as eliminated attributes from Alzheimer dataset. For every selected features from the three operators, the fitness function is applied to the selected features for selecting optimized features effectively. Then the fitness function is applied as follows.

$$Fitness\ function\ (n) = \frac{1}{\sum_{x=\Omega} \delta(x,n)} \quad (3)$$

In eqn (3) n is the given genetic chromosome and Ω is defined as genetic process training set and the features selection is done by using following condition.

$$\delta(x,n) = \begin{cases} 1, & if\ x\ is\ selected\ as\ best\ feautres \\ 0, & if\ x\ is\ selected\ as\ worst\ features \end{cases} \quad (4)$$

Based on the above process, the optimized features are selected from Alzheimer dataset with effective manner the selected features are classified by applying the support vector machine which is explained as follows.

STEPS INVOLVED IN THE PROPOSED WORK

Step 1: Alzheimer's data set acquired from NACC data repository.

Step 2: Preprocessing is done with data set to remove noisy data, missing values and etc

Step 3: After the data preparation attributes are selected with removed irrelevant data

Step 4: Alzheimer's data set normalization.

Step 5: Feature selection is done with evaluation model and Final attributes selection using genetic algorithm.

Step 6: Applying SVM classifiers

Step 7: Results produced

Step 8: END

C. SVM CLASSIFICATION

The Support Vector Machine (SVM) is a classifier that tries to generate a linear vector that partitions the classes partially. In this case that the data is no linearly undividable, the Support Vector Machine techniques raise the volume into a higher dimensional plane till on the plane; a vector can divide the classes. It is specifically a good classifier. It supports different mathematical formulations. The SVM algorithm generates a support vector that defines that margin of largest separation between the two classes. The totality of the SVM results could not be interpreted since they represent 35 dimensions. However, the results were examined visually by plotting the SVM results of only vector of attributes. Training and testing datasets are generated by make use of obtained binary solution and NACC database. The data first divided into two groups refers to testing and training data. At initial 133 models (100 AD, 33 healthy) from NACC indiscriminately separated into a training AD dataset and testing AD dataset. The training set contains 133 samples that is $\{(X^1_{tra,}C^1_{tra})\}$...., (X^t_{tra},C^t_{tra}),...., $(X^{133}_{tra},C^{133}_{tra})\}$; testing set contains 50 samples that is (X^1_{test},C^1_{test}),...., (X^t_{tra},C^1_{tra}),...., $(X^{t50}_{tra},C^{50}_{tra})$,...., where X_{test} and X_{tra} are m-dimensional feature vector , or {affected by AI patient, ordinary individual} is class label. The coded class label is computed as follows:

$$y^t = \begin{Bmatrix} 1, & if AIPatient \\ -1, & if Ordinaryindividual \end{Bmatrix} \quad (4)$$

Then y_{tra}= $\{y^1_{tra,}y^2_{tra,}\ldots y^{133}_{tra,}$ and y_{test}=$\{y^1_{test,}y^2_{test,}..y^{50}_{test,}\}$ these vectors are coming along with coded class label and training and testing data set. After that the selected features are examined in the and divided into two groups according to the hyper plane generation which is defined as follows,

$$w.x + b = 0 \qquad (5)$$

$$y_i(w.x + b) \geq 1 \; where \; i = 1,2,3, \ldots . N \qquad (6)$$

Based on the generated hyperplane, each features is examined and the class label is predicted as follows.

Define predicted class label as follows:

$$C_{tra}^{k} = \begin{cases} 1, & if y_{tra}^{k} \geq 0 \\ 2, & if y_{tra}^{k} < 0 \end{cases} \qquad (7)$$

Testing phase

Compute predicted coded class labels $y_{tra}^{\cdot}$

$$y_{test} = \dot{a} \; .G$$

Define predicted class label as follows:

$$C_{test}^{k} = \begin{cases} 1, & if y_{test}^{k} \geq 0 \\ 2, & if y_{test}^{k} < 0 \end{cases}$$

Based on the above process, the SVM classifier was applied on the training data followed by the testing data. Each plot includes training data testing data and support vectors. In the figure 2, support patterns are indicated with blue squares surrounding one of the patterns. To assess the classification results, the previous performance classification parameters of the objects, and by forcing a linearly separable solution.

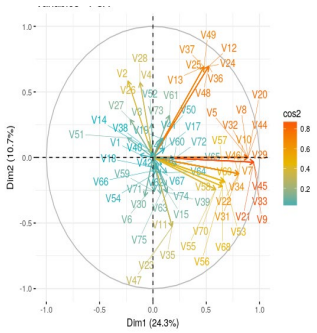

Figure 2: SVM Classifier Representation

According to the above discussions, the Alzheimer disease related effective features are successfully detected from dataset and the classification is done effectively done by using support vector machine. Then the efficiency of the system is examined using experimental results that are discussed as follows.

IV EXPERIMENTAL RESULT

The genetic features based Alzheimer disease detection process is implemented using MATLAB tool and excellence of the system is evaluated using accuracy, sensitivity, specificity, execution time, recall metrics which is discussed as follows.

1. Accuracy

Accuracy means the correct classification value of Alzheimer. The efficiency of hybridized classifier is compared with forecasting methods such as k-nearest neighbor (K-NN), Decision Tree (DT).

When compare to existing algorithms the proposed genetic algorithm based features and SVM classification produces efficient results

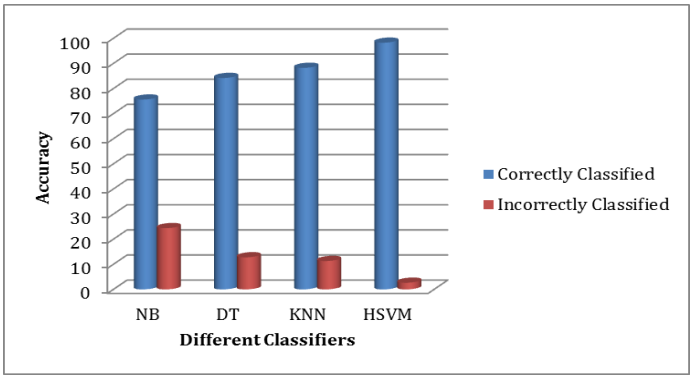

Figure 3: Accuracy

The above figure 3, indicates that accuracy of AI disease classification process. From the figure the support vector machine achieves better results while analyzing correct and incorrect features because of utilizing effective genetic related features that helps to improve the overall AI classification process. In addition to this, the system predicts the AI disease with minimum execution time that is shown in figure 4.

2. Execution time

Execution time is known as building process. When compare to existing algorithms execution time of the proposed genetic and SVM algorithms produces efficient results with less time.

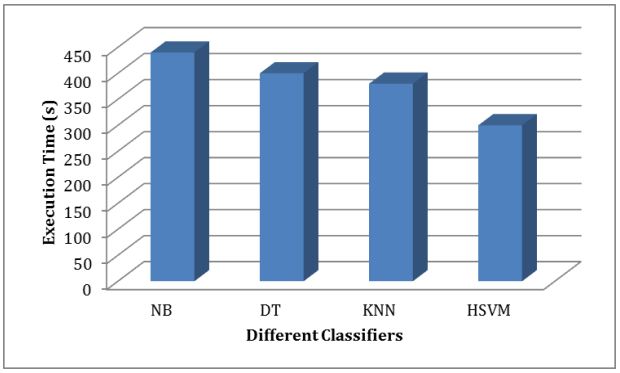

Figure 4: Execution time comparison

Not only the accuracy, the introduced support vector machine algorithm predicts the Alzheimer disease with minimum execution time. Further the excellence of system is evaluated using following metrics. 3. Sensitivity, Specificity, Accuracy, Recall

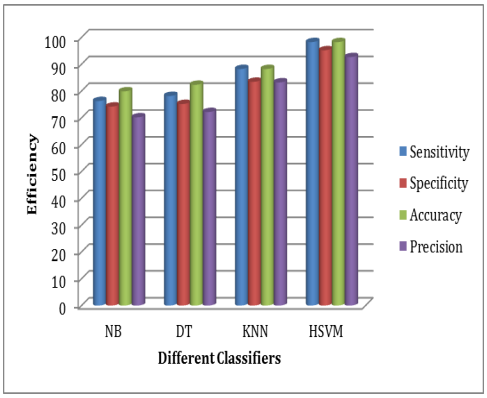

Figure 5: Sensitivity, Specificity, Precision and accuracy of existing and proposed approaches.

The above figure 5 demonstrates the difference between existing algorithms and proposed structures efficiency. Such as sensitivity, specificity, precision and accuracy percentage have been shown the figure which accuracy and other values are performing outstanding to detect and classify Alzheimer's data is support vector machine.

V CONCLUSION

Alzheimer is one of the dementia and it is very difficult to diagnose in early stage. The proposed system analyzed the use of data mining techniques for predicting Alzheimer. This study applied genetic algorithm and support vector machine on Alzheimer disease dataset and classified the data get better diagnosis. GA feature selection and SVM combination is minimizing the regularize cost function and diminish computation time. Reliability based classification where SVM is adopted if the test data is correctly classified with the selected AD dataset. The algorithm which predicts the best result in accuracy and also reduces the execution time for Alzheimer disease prediction using Genetic features and SVM classification.

REFERENCES

[1]. Ramesh Kumar Lama, JeonghwanGwak, Jeong-Seon Park, and Sang-Woong Lee, "Diagnosis of Alzheimer's Disease Based on Structural MRI Images Using a Regularized Extreme Learning Machine and PCA Features", Journal of Healthcare Engineering, Volume 2017, Article ID 5485080, 11 pages.

[2]. J.M. Górriz, F. Segovia, J. Ramírez, A. Lassl , D. Salas-Gonzalez, "GMM based SPECT image classification for the diagnosis of Alzheimer's disease", Applied Soft Computing 11 (2011) 2313–2325

[3]. Mohit Jain, PrernaDua, SumeetDua1 and Walter J Lukiw," Data Adaptive Rule-based Classification Systemfor Alzheimer Classification" Computer Science Systems Biology, Biol6:291-doi:10.4172/jcsb.1000124

[4]. F. Segovia, J.M. Gorriz, J. Ramirez, D. Salas-Gonzalez, I. Alvarez,"Early diagnosis of Alzheimer's disease based on Partial Least Squares and Support Vector Machine", Expert Systems with Applications 40 (2013) 677–683.

[5]. M. Lo´pez, J.Ramı´rez, J.M.Go´ rriz,,I.A´lvarez, D.Salas-Gonzalez, F.Segovia a, R.Chaves, P.Padilla, M.Go´mez-Rı´o, the Alzheimer's Disease Neuroimaging Initiative, "Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease", Neurocomputing 74 (2011) 1260–1271