Prediction of Alzheimer Disease Using Pearson Recursive Graph Convolutional Neural Network

M.S Roobini (✉ roobini.cse@sathyabama.ac.in)

Sathyabama Institute of Science and Technology

M. Lakshmi

SRMIST: SRM Institute of Science and Technology

Research Article

Keywords: Diabetes, Alzheimer diseases prediction, Pearson correlation, Graph convolutional neural network, ML Techniques.

Posted Date: October 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-935323/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Prediction of Alzheimer Disease Using Pearson Recursive Graph Convolutional Neural Network

M.S.Roobini,
Assistant Professor, Department of Computer Science and
Engineering,
Sathyabama Institute of Science and Technology
Chennai-600127, India
roobinims@gmail.com

M.Lakshmi,
Professor, Department of Computer Science and Engineering,
SRM Institute of Science and Technology,
SRM Nagar, Chennai-603203, India
laks@icadsindia.com

Abstract:

Alzheimer diseases are very hard to identify at beginning stage and also medication is available. So, the only way to protect those people is to predict the Alzheimer disease before it reaches the peak. More studies in diabetes show that there is a link between Diabetes and Alzheimer. Initially the prediction of diabetes is done using most relevant parameters, which detects the Diabetes. Then the severity level of diabetes is identified using some scoring levels. Based on scoring levels of diabetes, it is classified in to Type1 and Type2 using Machine Learning algorithms. When Diabetes reaches a worst case, it may affect any organs in the human body, whereas Type 2 Diabetes has associated with rare diseases which literally affects the brain and leads to cognitive impairment. After predicting the patients having cognitive impairment by applying classification algorithms are further examined to check whether it leads to Alzheimer disease. For this prediction the most relevant parameters which are common to Diabetes and Alzheimer is identified. Further identified parameters are used for prediction of Alzheimer disease with high accuracy which is helpful for taking precaution measures. In this proposed work, the most relevant features are selected using Pearson correlation-based feature elimination method and the diagnosis of the diabetes are carried using the Graph convolutional neural network (GCN). The measures of performance of the proposed work are calculated with various factors like Sensitivity measure, Recall, Precision, F-Measure. Proposed has achieved highest of 98.91%, 97.01%, 98.62%, 98.91% in above metrics.

Key words: *Diabetes, Alzheimer diseases prediction, Pearson correlation, Graph convolutional neural network, ML Techniques.*

1. INTRODUCTION:

Dangerous disease called Alzheimer (AD) is caused due NeuroBrain disfunction by memory damage. Across world 40 million people are suffered [1] due to memory loss. Dementia can be categorized as Alzheimer, dementia, temporal dementia and dementia with abnormality of deposits in brain [2]. Of all these Alzheimer diseases is the common and widely spread disease [3]. Mild cognitive impairment (MCI) and AD are considered as intermediate stage between the healthy life and dementia. There is various classification methods used in AD prediction. Liu et al., [4] stated a weighted multiple kernel learning (MKL) for classifying the AD with multi modalities. Zhu et al., [5] stated about joint regression with classification (JRC) approach for diagnosing the AD / MCI using the different modalities of the data. Xu et al., [6] proposed (mSRC) to discriminate AD. To classify the AD /MCI, Li et al., [7] proposed multimodal discriminative dictionary learning (mSCDDL) algorithm with the extension of supervised dictionary learning. Recent research proves that there is a relation between the type 2 diabetes and AD [8]. It is also observed that the persons having diabetes also have dementia than the person without diabetes [2].

Diabetes is one of most prominent among all the age groups. The diabetes patients are increasing due to lack of diet and physical activities of the human population. A person working with more stress and work pressure causes diabetes in young age. The various pathological characters compared between the islet and brain resembles both are interlinked in some circumstances. Cognitive impairment or dementias are considered as late effects of diabetes which results in Alzheimer disease. Dementias are considered as irregular brain functionality.

A machine learning algorithm proves its ability in complex medical situations. The various machines learning models such as SVM, Logistic regression technique as well as bootstrapping have been used for predicting the AD through diabetes diagnosis. Even though more researchers are contributing on predicting the AD, there is a lack of accuracy during early prediction of AD through type 2 diabetes. In our research, we chose machine learning for prediction and classification of diabetes 2. When compared to other techniques, machine learning works efficiently and its prediction time was very less. Graph convolutional neural network is 20th century neural network used in various real time classification models. Diabetes is severe diseases which leads various disorders in future like weight loss, eye dysfunction, and memory loss and heart problems. It is important to predict as early for treatment. Various researches are focused on diabetes prediction. This motivates us on to perform research in diabetes prediction to identify the Alzheimer possibilities of the patient using diabetes data.

1.1 Objectives:

There is a study in diabetes which shows that there is a link between Diabetes and Alzheimer. Hence the diabetes diagnosis is performed using our previous work methods. Once the type of diabetes disease is diagnosed as type 1 or type 2, then the parameters involved in the diagnosis are chosen by proposed machine learning based feature selection model. Based on the

significant features, the Alzheimer disease is classified using the proposed deep graph convolutional method.

1.2 Paper Contribution:

In this research manifesto, we explore the early treatment of AD based on the treatment for diabetes by following steps given:

- (i) Preprocessing to remove noise and missing data using the method called binning,
- (ii) Proposed a feature selection approach uses Pearson Correlation based feature elimination with recursive nature (PCFER)
- (iii) Classification of AD from diabetes diagnosis is done using deep learning algorithm called Deep Graph convolutional neural network (DGCN).

Rest of the paper is ordered with five chapters. Chapter 2 describes about the related research study and techniques. Chapter 3 describes about proposed methodology and its working principal. Chapter 4 explains about result achieved and the evaluation made. At last, Chapter 5 gives the conclusion with the work with future more possibilities.

2. LITERATURE REVIEW

M.S.Roobini and Lakshmi [9] analyzed the various classification algorithms for Alzheimer disease diagnosis. They analyzed the ML techniques like Support vector machine (SVM), logistic regression, decision tree algorithm, random forest methodology, and neighbor classifier. Patient's visits to diagnosis of disease are counted. Later, performance of the classification algorithms is analyzed using performance measures. Their analyzed results, shows that the logistic regression classification algorithms obtained high accuracy of 73%. Khan and Usman [10] reviewed about the recent research on early treatment of AD using ML algorithms. They compared approaches used in terms of different modalities and also, they compared factors such as preprocessing, number of features using feature selection and accuracy on predicting the disease. Shahbaz et al.,[11] analyzed six data mining methods such as k nearest neighbor (knn), Rule of induction, Decision making tree, generalized linear model (GLM) , Naïve Bayes techniques and deep learning techniques. They used ADNI medical information for analysis. They reported that among the algorithms, GLM performs better and classify the AD with the accuracy of 88.24%.

Farhan et al., [12] described the ensemble method of diagnosis of Alzheimer using the brain images. This approach can detect the disease status at earliest. They experimented dataset called OASIS which consists of the patient's detail with age of 85. The selected features are GM, CSF, size of hippocampus and WM. They used three techniques such as Support Vector Machine, Multi-Layer perceptron and J48. This ensemble technique is called as majority voting. Individual data features and combined data parameters are input for the classifier with tenfold cross validation. The accuracy of 93.75% was obtained with left hippocampus feature and the sensitivity and specificity were 100% and 87.5% respectively.

Taeho et al. [13] analyzed the deep learning and neuro imaging data for the classification of Alzheimer disease. They reviewed the articles from 2013 to 2018 and also evaluated the algorithms. Stacked auto encoder (SAE) based feature selection obtained the accuracy of 98.8%, MCI (mild cognitive impairment) obtained 83.8% of accuracy. Machine learning techniques such as convolutional and recurrent neural networks give accuracy as 84.2%. While combining the multi- model Neuro imaging and biomarkers, the best performance of the prediction has been obtained. Kruthika et al., [14] proposed multi stage classifiers for identification of dementia. They used ML techniques such as Naïve Bayes technique, KNN, and SVM for analysis. Further, MRI was used as powerful tool of analyzing this MRI images. To find best features of AD, the feature selection method called Particle Swarm Optimization is used. Here, dataset is taken from Alzheimer's Disease Neuroimaging Institute (ADNI).

Naganandhini et al., [15] proposed a decision tree-based AD prediction scheme. The proposed decision tree with hyper parameters is evaluated with OASIS dataset. They identified the brain abnormality attributes of the AD patients. Further, evaluation is done using various method of metrics such as recall etc. Their proposed approach obtained 99.10% of accuracy on AD prediction. The medical image analysis with respect to AD prediction through the structural MRI has been studied in Trojancanec et al., [16]. They extracted the smaller number of features from MRI measures. These features consist of the volume and thickness of brain. The feature subset was selected using the Correlation Selection of Features (CFS) methodology. They used the OASIS Imaging Studies dataset Adaszewski et al., [17], Demirhan [18]. ML methods like BP-NN are used in the prediction. Kalbkhani et al., [19] predicted the AD using detail coefficients of DWT with two level and also used GARCH model (generalizing the auto-regressive conditional Heteroscedasticity) for the parameter selection. They used KNN and SVM for classification. Güzel et al.,[20] used Naïve bayes and KNN for the detection and obtained 96% of accuracy. Some of the recent research on AD prediction and its metrics are listed in Table 1.

Paper	Methods used	Description
Park et al., 2020 [21]	AD prediction on large scale health data using various learning methodologies like random algorithm, SVM, and bootstrapping.	To diagnosis the AD, they used two operations such as definite AD and probable AD. They used Korean national health insurance large scale data for processing. Using Random forest, relevant features such as hemoglobin level, age and urine protein level were selected for early diagnosis. The model has been evaluated with various metrics such as AUC, sensitivity and specificity.
Wu et al., 2018 [22]	Type 2 diabetes prediction called T2DM using improved version of K	Type 2 DM has been predicted using the preprocessing steps and data mining and

	means and logistic regression techniques.	ML algorithms. The evaluated database was (PIMA) with WEKA tool. They obtained 3% of improved accuracy than other existing algorithms.
Luo et al., 2016 [23]	DM screening is done using pair-wises and size-constrained based K-means (PSCK means) algorithm.	They developed a tool in medical field for risk stratification in clinical diseases.
Tigga et al., 2020 [24]	Type 2 DM prediction using KNN, Logistic regression, SVM, Decision Tree, Random forest	They collected real time data of 952 instances with 18 questions related to health, family background and lifestyle. And they also used Pima India dataset and concluded that random forest classifier obtained high accuracy than existing model.
Swapna et. Al., 2018 [25]	Detection of DM is done by deep learning techniques. (CNN) and (LSTM) used in classification. Feature selection is done using Support vector machines (SVM).	They use electrocardiogram (ECG) signals to detect the diabetes. Their evaluated result was 95.7% accuracy.
Choubey et. Al., 2017 [26]	Hybrid algorithm of genetic based algorithm with involves radial basis NN function was used in identification of DM.	Model implements using GA in features extraction and RBFNN in data classification. The hybrid approach performs better than RBFNN alone used.
Zou et al., 2018 [27]	They used PCA and mRMR (minimum redundancy minimum relevance) for feature extraction and Random forest, J48 and ANN for classification.	Their evaluated results concluded that mRMR based feature selection performs better than PCA with high accuracy.
Kopitar et al., 2020 [28]	Prediction of undiagnosed T2DM using Glmnet, RF, XGBoost, LightGBM	They predicted fasting plasma glucose level with 100 bootstrap iterations. The model obtained low RMSE with high accuracy of prediction.
Li et al., 2018 [29]	Alzheimer's disease prediction using (MKSCDDL)	Using the proposed method, they differentiate AD and MCI from cognitively unimpaired (CU). From (ADNI) database, proposed work obtained high quality on early diagnosis using neuroimaging data using tomography and sMRI.

Chung et al., 2019 [30]	CNN applied on four classification such as AD vs NC, pMCI vs. NC, stable sMCI vs. NC and pMCI vs. sMCI h	For AD vs NC, unsupervised method called Convolutional auto encoder (CAE) used. For pMCI vs. sMCI, supervised method with transfer learning used. For AD and pMCI, gradient based visualization method used They used ANDI database for evaluation. they identified temporal and parietal lobes are the source for classification
-------------------------	--	--

Table 1: Recent research on AD and Diabetes prediction

3. PROPOSED METHODOLOGY

Diagnosing the Alzheimer disease (AD) is done based on the parameters involved in the diabetes disease. Those parameters / features are needed to be selected with efficient method to improve the prediction accuracy. The efficient feature selection and classification will improve the clinical validation, cost of disease diagnosis. Feature Selection (FS) is a process that could for selecting the subset of relevant parameters from Diabetes data which improve the classification performance by reducing feature set. To select a relevant characteristic of these features for improving the system performance, feature engineering used. The main goal of feature engineering is to prepare the input dataset that is compatible with deep learning algorithms in aiming to improve the machine learning performance and classification. The proposed architecture as given in Fig 1. The information's is separated into training and testing information based on the cross-fold validation with the ratio of 80:20. The raw input data set are preprocessed to remove the redundant and noise using the normalization technique. In this paper, feature engineering process divided into three categories such as (i) preprocessing to remove noise and missing data using the method called binning, (ii) proposed feature selection approach based on Pearson Correlation based feature elimination with recursive nature (PCFER) and (iii) classification using deep learning algorithm called Deep Graph convolutional neural network (DGCN). Compare to the existing diagnosis system, this proposed approach will reduce the feature set which will be improving the classification and minimize computation cost and rate of error.

Fig 1: Overview of proposed AD diagnosis

3.1 Preprocessing:

The raw input data set are complex to process due to noise and missing values. It is also difficult to process the dataset with whole features of numeric and non-numeric data. Hence, the raw input data are to be preprocessed before proceeding for further analysis improvement. Initially the symbols in the data are to replace with the unique numeric value using the mapping function available in python library which is mathematically represented using the logarithmic Equation [29] of the dataset X which is represented in Eqn (1).

$$X_{normalized} = \log(X_i + 1) \quad (1)$$

The data features having missing values may also have some useful information. Avoiding the missing values data may affect the performance of the classification. Those missing values are handled using the binning method. In this method, the data are smoothened and the missing values

are handled. Initially, the data are sorted and distributed to the number of buckets called bin. Later, small intervals replaced with the common value computed for those respective bins. In this work, equal width binning is used to smoothen the data. The width of bin is equal to range of bin as defined in Eqn (2)

$$[\min + 1v], [\min + 2v] \dots [\min + nv] \quad (2)$$

where,

$$v = \frac{(\max - \min)}{\text{no of bins}} \quad (3)$$

3.2 Pearson Correlation based Feature Selection in Recursive form (PCFSR)

The normalized data are set as input information for feature selecting. This phase identifies relevant parameter data subset using proposed feature selection algorithm. In this proposed system, we use hybrid feature selection approach called Pearson Correlation based FSR. It is a combination of filter with wrapper methods. Compare to the normal feature selection algorithms, optimization-based feature selection algorithms are proven to be best to improve the classification accuracy with reduced number of feature sets. Pearson Correlation is the relationship between the data in the range of [-1,1] i.e, 1 represents as positive in correlation, 0 means without correlation and - 1 indicates negative correlations of the data. Compare to the other feature selection machine learning models remove features at each step, PCFSR removes the irrelevant data at once. Due to this factor of hybrid approach, it is faster than filter, wrapper and embedded FS methods. The Correlation Coefficient of the features are calculated using the Eqn (3).

$$PCorr_{xi,yi} = \frac{\sum_{i=1}^m (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^n (p_i - \bar{p})^2} \sqrt{\sum_{i=1}^n (q_i - \bar{q})^2}} \quad (3)$$

Where here, xi, yi – features for correlation consideration. The value of this falls in the close interval [-1,1]. Then value close to -1 or 1 represents the strong feature sets and 0 indicates weak relationship of the two features. Once the features are correlated, a threshold value is used to rank the features. The features obtain minimum rank will be removed. The feature elimination process is represented as Eqn (4)

$$PCRFE(x_i) = \sum_{i=1}^n \left(y_{i,j} - \sum_{j=1}^d PCorr_{xi,yj} \times x_j \right)^2 \quad (4)$$

Algorithm 1:

Input: Normalized data set D, size of dataset N and number of features n,

Output: Optimized feature subset

Step 1: preprocessing D using Eqn (1) and Eqn (2)

Step 2: For i =1 to N

Step 3: For j = 1 to n

Step 4: Compute the correlation coefficient of the feature using the Eqn (3)

Step 5: Eliminate the features using the PCFSR Eqn (4)

Step 6: If ($PCFSR(x_i) \geq threshold$) then

Step 7: Add the features into the subset

Step 8: End if;

Step 9: End for;

Step 10: End for

3.3. Classification using Deep Graph (DGCN)

Classification structure with traditional (CNN) is constructed by alternative arrangement of convolution and sampling layers. The pooling layer of this CNN will lose some feature information while the input features are not prominent which will lead to reduce the learning capability of the network. Hence, the traditional CNN will not sometimes be suited for accurate classification.

In contrast, the graph CNN is type of learning technique established for accurate categorization of area in image processing, data mining and cloud computing. It will efficiently extract the data from sample training set which can increase accuracy of data classification. It can also add more image processing-based CNN [30]. It will be represented in the form of graphs and process the data that cannot process by the CNN. This will transfer the convolution data into the graph. The available nodes present here in this graph linked with hierarchical structure with spatial data domain which is called as convolution learning. Laplacian matrix is formed using the spectral decomposition method that can convert the parameter vectors into the data domain. The convolution of the graph is formed with the spectral domain point multiplied with inverse Fourier transform. GCN can also allow learning on structured information and parameters. This learning helps to extract features of the graph network.

The relationship of spatial domains is represented in adjacency matrix $A \in R^{N \times N}$. Further, spatial relationship between the central and neighbor nodes in the graph are determined through the

spectral similarity and spatial domain. If a_i node is adjacent to a_j node, then edge that connecting these nodes have the weight $a_{ij} \in A$. These adjacency between the nodes are declared as in Eqn (5),

$$\begin{cases} a_{ij} = 0 & \text{if } a_i \text{ is not adjacent to } a_j \\ a_{ij} = \exp(-\frac{\|a_i - a_j\|}{\sigma_w^2}) \end{cases} \quad (5)$$

Where” σ_w^2 “represents range of weight. The graph convolution consists of the three phases: (i) characteristic details of the node are extracted, (ii) Each node collects this information from the next neighbor node and forms the local structure, and (iii) to perform non-linear transformation from the previous node information to increase the ability of the expressive model. This graph convolution then aggregates the neighbor node data to form a new representation. To aggregate the spatial information, this graph convolution is the base object. It is represented in Eqn (6)

$$Z = g(H, A) \quad (6)$$

$$H^{(i+1)} = \sigma(L_G H^{(i)} W^{(i)}), L_G = \tilde{D}^{-\frac{1}{2}} \tilde{A}^{-\frac{1}{2}} \quad (7)$$

Where, $H^{(i)}$ - feature of the layer i from algorithm 1, σ - activation function, W - learnable parameter, L_G - Laplacian matrix, $\tilde{D}$ - degree matrix and $\tilde{A}$ - $A+I$ and I - Identity matrix. In order to reduce the loss during the training process, the W is continuously adjusted to optimize the output as in Eqn (8),

$$O = \sum_{i=1}^n \text{loss}(g(X_i^G, A), GT_i) \quad (8)$$

The overall structure of proposed work is given in Fig 2. Workflow of our proposed feature selection-based classification model is shown in Fig 3.

Algorithm: 2

Input: Dataset D, feature subset

Output: classification result of AD

Step 1: collect feature subset from algorithm 1

Step 2: the graph nodes are constructed and it is called as V. use the features X to initialize the graph node contents as $X^G = [X_1^G, X_2^G, \dots, X_k^G]^T \in R^{k \times D}$.

Step 3: Graph edges are constructed and it is called as E by considering the first order adjacency relationships. Calculate the strength of the edges as w .

Step 4: once training over, classify the graph nodes with DGCN.

Step 5: return the classification result

Fig 2: Proposed model of Deep Learning for Alzheimer disease diagnosis.

Fig 3: Workflow for proposed algorithm

4. RESULT AND DISCUSSIONS:

This portion describes about experimented results of proposed deep learning approaches. The proposed algorithm is evaluated by the terms of the evaluation metrics like accuracy, recall, precision, F1 score and recall. These metrics are evaluated using the python scikit learn library.

4.1 Evaluation Metrics:

The accuracy of classification is computed based on various metrics like True Positive, False Negative, True Negative and False positive.

$$Accuracy(\%) = \frac{TN+TP}{TN+FN+TP+FP} \quad (9)$$

Precision: The measure of data samples with diabetic is classified from total data samples in data set.

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

F - Measure: The measure of average from Precision and Recall using following Eqn (11)

$$F - Measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (11)$$

4.2 Diabetes Mellitus Dataset Description:

The dataset Pima Indians dataset of Diabetes is collected for evaluation from UCI database [31]. The ultimate goal of proposed work is predicting diabetes patient accurately with more disease features. These diabetes patients are prone to Alzheimer diseases in future. The data set contains samples of 768 with 15 attributes and it also contain 1 attribute class which has a binary value of 0's or 1's. if attribute class is 0, then diabetes is resulted as negative or else if attribute class is 1, then diabetes is resulted as positive for patient. The data set attributes with description is given in Table 2.

Features / Attributes Number	Features / Attributes
i.	Number of times the patient pregnant (NTP)
ii.	Concentration of Plasma Glucose content (PGC)
iii.	Diastolic Blood Pressure value (mmHg) (DBP)
iv.	thickness at skin-fold of Triceps (mm) (TSFT)
v.	s-h insulin at serum (mu U/mL) (H2SI)
vi.	Mass index of Body (kg/m2) (BMI)
vii.	Functionality of Pedigree for Diabetes (DPF)
viii.	Patients age limit
ix.	Class 1 and Class 0 (type 2 diabetes disease Diagnosis)

Table 2: Pima Indian Diabetes Dataset attributes

The evaluated results based on the accuracy and training time of various algorithms are shown in Table 3. For comparative analysis, the existing methodologies like naïve bayes, sSVM, SMO, J48 as well as KNN are used.

Algorithms	Classification Accuracy (%)		Training time (s)
	Correctly classified data	Incorrectly classified data	
Naïve Bayes	83.1	10.4	1.82
SVM	91.24	5.2	0.68
SMO	88.92	3.28	1.03
J48	86.43	7.62	1.84
KNN	89.953	13.02	1.65
Proposed PCFSR+DGCN	98.76	0.263	0.03

Table 3: Accuracy and Training time comparison of Existing vs Proposed method

Fig 4 illustrates the correctly and incorrectly classified data of various Existing and proposed algorithm. The evaluated results clearly show that the highest percentage of 98.76% obtained by proposed algorithm on classifying the diabetes data and low percentage of 0.263% obtained by the proposed approach on incorrectly classifying results compared to other algorithms. The lowest performance was secured by Naïve Bayes on correct classification. The Fig 5 illustrates the minimum training time obtained by proposed algorithm of 0.03s on training the diabetes data using deep neural network and maximum training time secured by J48 based training. The evaluation in terms of the evaluation metrics are given in Table 4.

Fig 4: Accuracy comparison of proposed vs existing algorithm

Fig 5: Comparison in terms of training time

Algorithms	Sensitivity	Precision	F-Measure
Naïve Bayes	92.01	82.8	82.67
SVM	89.02	86.03	86.93
SMO	88.04	88.92	89.01
J48	88.45	87.01	86.82
KNN	88.51	90.74	91.82
Proposed PCFSR+DGCN	98.91	98.62	98.91

Table 4s: Performance comparison of Existing vs proposed approaches

Fig 6: Performance evaluation of propose approaches

With the observation of the illustration and comparison from fig 6, our proposed approach obtains high precision and accuracy of 98.62% and 98.76% respectively, which is higher than other existing approaches. Based on all the kind of comparison and evaluation our proposed algorithm gives promising results on correct classification of diabetes patients with relevant parameter selection. This improves the diagnosis of the system performance by communicating patients with the correct result. The correct diagnosis of patients' diet, sport, and blood glucose level are shared to the patients for better diagnosis of their disease. Hence, our proposed algorithm is effective and efficient to classify the data with the identification of the features that will leads to Alzheimer disease classification.

5. CONCLUSION AND FUTURE WORK:

More research studies show that diabetes may leads to other diseases cognitive impairment, memory loss, vitamin deficiency, Alzheimer disease, heart attack, kidney failure etc. In this research we use the machine learning algorithm to classify the diabetes which leads to Alzheimer disease. The features in data set of patients such as NTP, PGC, DBP, Class 0 or 1, BMI, DPF helps to measure diabetes level and predict the occurrence of Alzheimer diseases in future. The Pearson correlation-based feature elimination using recursive function selects the diabetic patient with accuracy who may prone to Alzheimer diseases. Graph convolutional neural network classify the diabetic patient with high precision and recall. The accuracy of our prediction is 98.76%, which is helps medical world to prescribe proper medicine and diet for patient.

DECLARATIONS

Funding:

The authors did not receive financial support from any organization for the submitted work.

Conflicts of interest/Competing interests:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and material : 'Not applicable'

Code availability : 'Not applicable'

Authors' contributions : 'Not applicable'

Ethics approval : Compliance with Ethical Standards

Consent to participate ' : Not applicable'

Consent for publication:

Authors give consent to Soft Computing to publish their article.

References

1. Selkoe, D. J., and Hardy, J., (2016) “The amyloid hypothesis of Alzheimer’s disease at 25 years”, *EMBO Mol. Med.* 8, 595–608. doi: 10.15252/emmm.201606210
2. Kawamura T, Umemura T, Hotta N. Cognitive impairment in diabetic patients: Can diabetic control prevent cognitive decline? *J Diabetes Investig.* 2012; **3**:413–423
3. Inestrosa NC, Alvarez A, Pérez CA, Moreno RD, Vicente M, Linker C, Casanueva OI, Soto C, Garrido J., (1996) “Acetylcholinesterase accelerates assembly of amyloid-beta-peptides into Alzheimer’s fibrils: possible role of the peripheral site of the enzyme”, *Neuron*.**16**:881–891
4. Liu, F., Wee, C. Y., Chen, H., and Shen, D. (2014b). Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer’s Disease and mild cognitive impairment identification. *Neuroimage* 84, 466–475. doi: 10.1016/j.neuroimage.2013.09.015
5. Zhu, X., Suk, H., and Shen, D. (2014a). A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *Neuroimage* 100, 91–105. doi: 10.1016/j.neuroimage.2014.05.078
6. Xu, L., Wu, X., Chen, K., and Yao, L. (2015). Multi-modality sparse representation-based classification for Alzheimer’s disease and mild cognitive impairment. *Comput. Methods Programs Biomed.* 122, 182–190. doi: 10.1016/j.cmpb.2015.08.004

7. Li, Q., Wu, X., Xu, L., Chen, K., Yao, L., and Li, R. (2017). Multimodal discriminative dictionary learning for Alzheimer's disease and mild cognitive impairment. *Comput. Methods Programs Biomed.* 150, 1–8. doi: 10.1016/j.cmpb.2017.07.003
8. Sridhar GR, Lakshmi G, Nagamani G. (2015) "Emerging links between type 2 diabetes and Alzheimer's disease", *World J Diabetes*,6(5):744-751. doi:10.4239/wjd. v6.i5.744
9. M. S. Roobini and M. Lakshmi, (2020) "Diagnosis of Alzheimer Disease using Classification Algorithms," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, pp. 725-734, doi: 10.1109/ICISS49785.2020.9316028.
10. Khan, A. and Usman, M. (2015) "Early Diagnosis of Alzheimer's Disease using Machine Learning Techniques - A Review Paper", In Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 1: KDIR, pages 380-387 ISBN: 978-989-758-158-8
11. Shahbaz, M., Ali, S., Guergachi, A., Niazi, A. and Umer, A., (2019) "Classification of Alzheimer's Disease using Machine Learning Techniques", In Proceedings of the 8th International Conference on Data Science, Technology and Applications (DATA 2019), pages 296-303 ISBN: 978-989-758-377-3, DOI: 10.5220/0007949902960303
12. Saima Farhan, Muhammad AbuzarFahiem, and HumaTauseef, (2014) "An Ensemble-of-Classifiers Based Approach for Early Diagnosis of Alzheimer's Disease: Classification Using Structural Features of Brain Images", Hindawi, Computational and Mathematical Methods in Medicine, Volume 2014 |Article ID 862307 | <https://doi.org/10.1155/2014/862307>
13. Taeho Jo, Kwangsik Nho, Andrew J. Saykin,(2019) "Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data", *Front. Aging Neurosci* | <https://doi.org/10.3389/fnagi.2019.00220>
14. K.R. Kruthika, Rajeswari, H.D. Maheshappa,(2019) 'Multistage classifier-based approach for Alzheimer's disease prediction and retrieval', *Informatics in Medicine Unlocked*,Volume 14,Pages 34-42,ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2018.12.003>
15. Naganandhini, P. Shanmugavadivu,(2019) "Effective Diagnosis of Alzheimer's Disease using Modified Decision Tree Classifier", *Procedia Computer Science*,Volume 165, Pages 548-555, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.01.049>
16. Katarina Trojancanec, et al. (2017) "Image retrieval for Alzheimer's disease based on brain atrophy pattern",*International conference on ICT innovations* springer, cham , pp. 165-175
17. S. Adaszewski, J. Dukart, F. Kherif, R. Frackowiak, B. Draganski, (2013), "Alzheimer's Disease Neuroimaging Initiative How early can we predict Alzheimer's disease using computational anatomy?" *Neurobiol Aging*, 34 (12), pp. 2815-2826
18. A. Demirhan, (2016) "Classification of structural MRI for detecting Alzheimer's disease", *International Journal of Intelligent Systems and Applications in Engineering*, 4 (1), pp. 195-198

19. H. Kalbkhani, M.G. Shayesteh, B. Zali-Vargahan, (2013) "Robust algorithm for brain magnetic resonance image (MRI) classification based on GARCH variances series", *Biomed Signal Process Control*, 8 (6) , pp. 909-919, 10.1016/j.bspc.2013.09.001
20. C. Güzel, M. Mahmut Kaya, O. Yıldız, (2013) "Breast cancer diagnosis based on naïve Bayes machine learning classifier with KNN missing data imputationIn 3rd world conference on innovation and computer sciences
21. Park, J.H., Cho, H.E., Kim, J.H. *et al.* (2020), "Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data", *npj Digit. Med.* **3**, 46.<https://doi.org/10.1038/s41746-020-0256-0>
22. Han Wu, Shengqi Yang, Zhangqin Huang, Jian He, Xiaoyi Wang, (2018) "Type 2 diabetes mellitus prediction model based on data mining", *Informatics in Medicine Unlocked*, Volume 10, Pages 100-107,ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2017.12.006>
23. Longfei Han, Senlin Luo, (2016) "An intelligible risk stratification model based on pairwise and size constrained Kmeans", *IEEE J Biomed Health Inf*, 21 (5), pp. 1288-1296.
24. Neha Perna Tigga, Shruti Garg, (2020) "Prediction of Type 2 Diabetes using Machine Learning Classification Methods", *Procedia Computer Science*, Volume 167, Pages 706-716, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.03.336>
25. Swapna, G., Vinayakumar R., Soman K. P. (2018) "Diabetes detection using deep learning algorithms." *ICT Express* 4 (4): 243-246
26. Choubey, D.K., Paul, S. (2017) "GA_RBF NN: a classification system for diabetes." *International Journal of Biomedical Engineering and Technology* 23 (1): 71-93.
27. Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., Tang, H. (2018) "Predicting diabetes mellitus with machine learning techniques." *Frontiers in genetics* 9: 515. <https://doi.org/10.3389/fgene.2018.0051>
28. Kopitar, L., Kocbek, P., Cilar, L. et al. (2020) "Early detection of type 2 diabetes mellitus using machine learning-based prediction models", *Sci Rep* 10, 11981. <https://doi.org/10.1038/s41598-020-68771-z>.
29. R. Eberhart and J. Kennedy, (1995) "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 4, pp. 1942–1948
30. Zhou, J.; Cui, G.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. (2018), "Graph neural networks: A review of methods and applications", *arXiv* 2018, arXiv:1812.08434
31. V. Sigillito, (1990), "Data set Pima Indians Diabetes collected", <https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>, Accessed dec 8, 2015.
32. Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., ... & Alzheimer's Disease Neuroimaging Initiative. (2019). Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical*, 21, 101645.
33. Alashwal, H., Abdalla, A., Halaby, M. E., & Moustafa, A. A. (2020, January). Feature Selection for the Classification of Alzheimer's Disease Data. In *Proceedings of the 3rd International Conference on Software Engineering and Information Management* (pp. 41-45).

Author details

M.S.Roobini received B.E degree in Computer Science and Engineering from Cape Institute of Technology and the M.E degree in Software Engineering from Noorul Islam University in 2009 and 2011,respectively.she is a Faculty Member of Department of Computer Science and Engineering,School of Computing,Sathyabama Institute of Science and Technology, Chennai,India.Her current research interests include Artificial Intelligence, Machine Learning,and Computer Vision.

Dr. M Lakshmi presently working as Professor in Department of CSE,SRM Institute of Science and Technology, Kattankulathur.She previously worked as Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences from 2019 to 2018 and she worked as Principal in Sri Krishna College of Technology, Coimbatore from Sep 2018 to April 2019. She also worked as Professor and Dean, School of Computing, Sathyabama University, Chennai, till June 2018 from Nov 1995. She has an experience of more than 24 years in Teaching. She has completed her B.E. (Computer Science and Engineering) from Bharathidasan University and M.E. (Computer Science and Engineering) from Madras University and Ph.D. from Sathyabama University in the area of Wireless Ad Hoc networks.Her current research interests include Computer Vision, Artificial Intelligence, Machine Learning, and Wireless Ad Hoc networks. She has published more than 100 research papers in peer-reviewed journals.