Computers in Biology and Medicine 141 (2022) 105056

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A deep learning framework with an embedded-based feature selection approach for the early detection of the Alzheimer's disease

Nivedhitha Mahendran, Durai Raj Vincent P M *

School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

ARTICLE INFO

Keywords:
Machine learning
Deep learning
Embedded feature selection
DNA Methylation
Alzheimer's disease
Gene expression

ABSTRACT

Ageing is associated with various ailments including Alzheimer 's disease (AD), which is a progressive form of dementia. AD symptoms develop over a period of years and, unfortunately, there is no cure. Existing AD treatments can only slow down the progression of symptoms and thus it is critical to diagnose the disease at an early stage. To help improve the early diagnosis of AD, a deep learning-based classification model with an embedded feature selection approach was used to classify AD patients. An AD DNA methylation data set (64 records with 34 cases and 34 controls) from the GEO omnibus database was used for the analysis. Before selecting the relevant features, the data were preprocessed by performing quality control, normalization and downstream analysis. As the number of associated CpG sites was huge, four embedded-based feature selection models were compared and the best method was used for the proposed classification model. An Enhanced Deep Recurrent Neural Network (EDRNN) was implemented and compared to other existing classification models, including a Convolutional Neural Network (CNN), a Recurrent Neural Network (RNN), and a Deep Recurrent Neural Network (DRNN). The results showed a significant improvement in the classification accuracy of the proposed model as compared to the other methods.

1. Introduction

Neurological disorders affect the functioning of the brain, spinal cord, or nerves [1]. Abnormalities (e.g., electrical, structural) in neural structures can cause such things as altered levels of consciousness, paralysis, muscle weakness, loss of sensation, poor coordination, seizures, pain, and confusion [2]. The causes of these disorders can include lifestyle choices, infections, genetics, or environmental factors [3]. Some neurological disorders occur congenitally or early in life, while others result from tumors, trauma, or other structural defects [3].

Many neurological disorders directly affect the brain, an integral part of the body that controls many critical functions, including coordination and movement, cognition, emotions, and learning and memory [4]. One of most critical parts of the brain is the hippocampus, which is found in the temporal lobe. This structure plays a central role in emotional control and learning and memory, and damage to the hippocampus is associated with various neurological and psychiatric disorders, such as Alzheimer's disease (AD), epilepsy, and major depressive disorder, among others. Irrespective of the cause, neurological disabilities can be associated with reversible or irreversible damage to the nervous system [5]. One such neurological disorder is AD, which is primarily characterized by cognitive issues, particularly problems with learning and memory.

In AD patients, damage typically first appears in the hippocampal region and later spreads to other parts of the brain, inducing neuronal cell death [6]. As this damage progresses, the tissues in the brain shrink significantly. Unfortunately, AD is a progressive, irreversible neurodegenerative dementia with no cure [7]. However, early diagnosis and treatment can slow the progression of this disease. Boosted by recent developments in advanced technologies, such as MRI, CT scans, and cutting-edge genetics research (DNA sequencing, gene expression, etc.), better treatments for AD are being established.

Historically, the complexity of neurological diseases had made it exceedingly difficult to translate basic scientific findings into effective treatments for these disorders [8]. However, in the current age of "Big Data," the ability to gather and manipulate large data sets had increased exponentially [9]. These complex data sets can include multi-modal and high-dimensional data, such as that obtained from medical imaging, genetics, and clinical findings. With larger data sets comes the problem of data analytics. This formidable challenge is being tackled with the aid

* Corresponding author.
E-mail addresses: nivedhitha.m2019@vitstudent.ac.in (N. Mahendran), pmvincent@vit.ac.in (D.R.V. P M).

https://doi.org/10.1016/j.compbiomed.2021.105056
Received 14 September 2021; Received in revised form 20 November 2021; Accepted 20 November 2021
Available online 22 November 2021

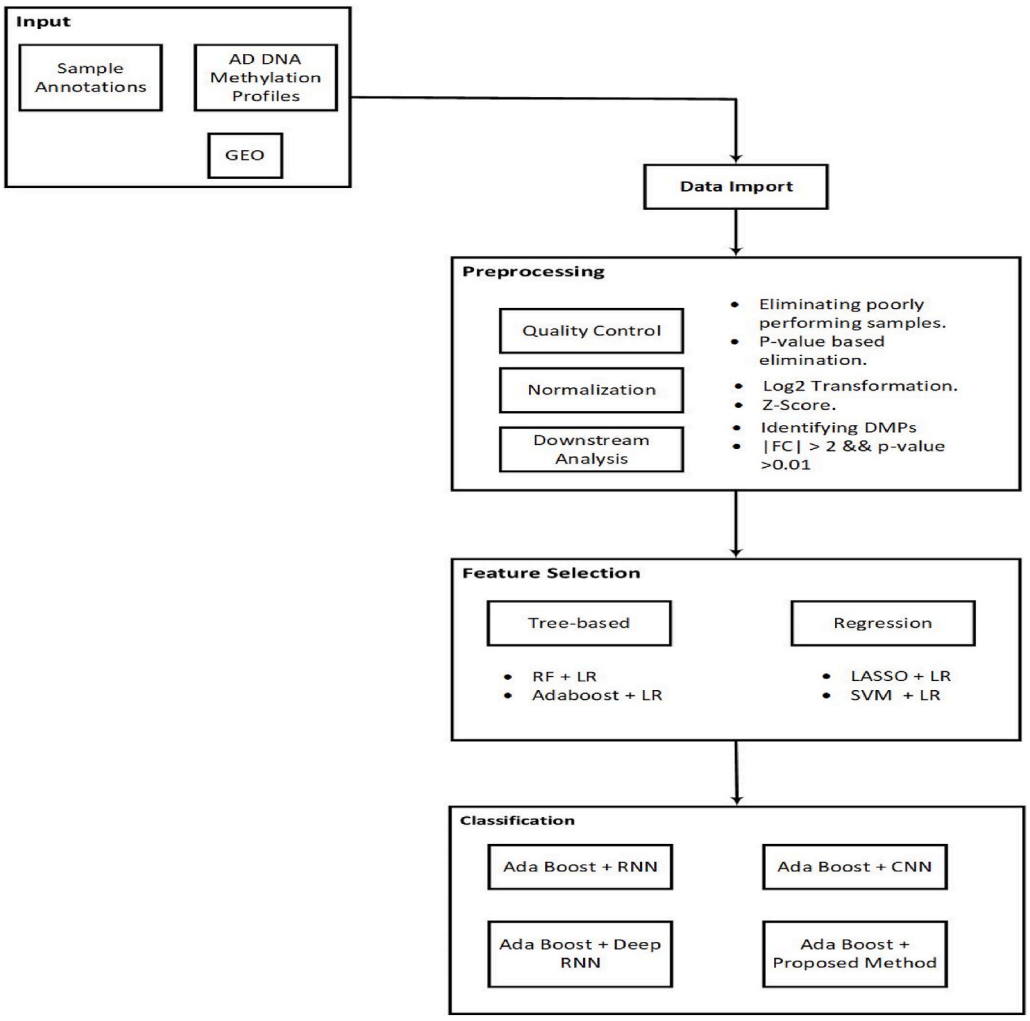

Fig. 1. The process flow for the current analysis.

of machine learning, deep learning, and advanced statistical and mathematical algorithms.

2. Background and methodology

2.1. Alzheimer's disease and artificial intelligence

As outlined above, AD is primarily associated with problems in learning and memory [10] and AD patients face a progressive and irreversible decline in their cognitive abilities. According to a 2015 survey, more than 30 million people suffer from AD worldwide, and by 2050 this number is expected to increase by 14 million [11]. Prolonged suffering from AD is associated with tissue loss in various brain regions [12]. The tissue damage begins in the grey matter and gradually progresses to the white matter (including the corpus callosum), finally reaching the hippocampus.

The most common symptoms of AD are memory loss, disruptions in verbal communication, and issues with concentration, judgment, thinking, and decision making. These symptoms start slowly and progress gradually over time, eventually preventing the patient from carrying out normal day-to-day activities [10]. As mentioned above, there is no cure for AD and available treatments only slow the progression of symptoms. Unfortunately, these treatments can burden the patients and their families financially, physically, and mentally. Recent studies have suggested that the key for the development of a cure depends on the ability to identify this disorder at an early stage.

Early on, AD manifests as a mild cognitive impairment (MCI), which then progresses to a stable MCI, followed by a progressive MCI, finally resulting in AD [13]. This progression from normal cognition to MCI to AD can take 6–12 months [11]. The most reliable methods for detecting AD at the early stage include computer-aided techniques and medical imaging. Although these clinical screening techniques are effective in some cases, they often are inaccurate and are costly. These limitations have led to the search for a molecular-based approach for early AD diagnosis. The ability to accurately identify relationships between genotypes and phenotypic symptoms makes molecular-based biomarker identification particularly attractive.

While a molecular approach may have advantages over clinical screening, the massive volume of data generated by these methods requires powerful computational techniques for analysis. The critical problem with handling molecular data is known as the high dimensionality and low sample size (HDLSS) challenge, commonly termed the "curse of dimensionality." In the context of genomic analysis, this issue is also is referred to as the "large G and small n" problem, where 'G' is the number of genes (features) and 'n' is the number of samples [14]. The HDLSS problem requires pattern recognition methods, and the solution for this lies with artificial intelligence (AI) technologies powered by deep learning algorithms [15].

A number of studies have examined the use of AI for diagnosing AD based on imaging or molecular data (e.g., gene expression and DNA methylation data). AI approaches have been applied for the detection of AD progression [15,16], early diagnosis [17–21], and biomarker

identification [15,22–26]. While many studies have focused on brain imaging data, this approach has sometimes produced inaccurate results. However molecular data can correctly identify phenotype and genotype relationships, thus making diagnosis easier. As mentioned above, the issue with using molecular data is the large G and small n problem. Thus, recent studies have focused on selecting the best gene subset that effectively assists in diagnosing AD. For instance, a wrapper based gene selection method combining genetic algorithm and support vector machine (SVM) has been used on gene expression data [27]. Five feature selection and five classification techniques have also been applied to a blood gene expression dataset to predict AD classification [28]. Other works have applied similar AI approaches to gene expression data sets to aid with AD diagnosis [28–34]. Other than gene expression, DNA methylation-based datasets have also been examined in this context. For instance, a random forest based machine learning approach has been used to identify the different methylated positions that might be used as a biomarkers for AD [35].

AI learning models learn directly from the data, and the model improves performance by exposing itself to huge volumes of data and gains experience through training over time. This experience enables the model to make predictions using previously unseen data. There are three commonly implemented categories of AI learning models: supervised for structured and labeled data, unsupervised for unlabeled and unstructured data, and semi-supervised, which combines both supervised and unsupervised approaches.

2.2. DNA methylation

In this study, a DNA methylation dataset was used to classify control and AD cases. DNA methylation is a critical epigenetic mechanism that causes chemical modifications in the DNA [36]. This process involves transferring a methyl group to a cytosine in the C5 position forming a 5-methylcytosine, and it plays a crucial role in gene expression and cell differentiation [36]. DNA methyltransferase enzymes assist in the DNA methylation process. DNA methylation is also important for normal cognitive function. When there are alterations in DNA methylation resulting from environmental risk factors or developmental mutations, side effects like mental impairment can occur [37].

2.3. Preprocessing

Preprocessing was used to improve the ability to classify the AD and control cases. A careful preprocessing of the data should reduce the variance and enhance statistical power. These changes can aid in the detection of small changes in DNA methylation that are associated with complex phenotypes. Fig. 1 shows the process flow of the implemented approach, which is described further below. The critical preprocessing steps undertaken for the DNA methylation data were quality control, normalization, and downstream analysis.

- **Quality Control:** Quality control detects poorly performing samples in the dataset [38]. It uses the p-values to eliminate poor samples. The detection p-value is generated for every CpG in the samples, which determines the reliability of the signal. If the p-value is large (>0.01), then it is considered a poor-quality signal.
- **Normalization:** Normalization removes random noise, experimental artifacts, and systematic and sometimes technical variations. These issues when left unaddressed can hide important biological differences [39]. This is especially true for DNA methylation data, where there are irregularities in the methylation levels throughout, which can skew the distribution [40]. The degree of skewness depends on the methylation levels in the sample. The imbalance in the methylation levels is caused by a non-random CpG site distribution in the genome, mainly because of the link between the CpG density and DNA methylation.
- **Downstream Analysis:** Differential methylation levels are used to determine the differences between the cases and controls. The methylation levels are estimated based on the probe intensities (methylated and unmethylated probe from the Illumina Infinium assay). The methylated and unmethylated levels of the interrogated CpG sites are commonly identified using two techniques, Beta values and M-values. The Beta value is the ratio of methylated probe intensity to overall intensity, determined by the following equation [41]:

$$Beta_n = \frac{max(y_n,\ methylated,\ 0)}{max(y_n,\ unmethylated,\ 0) + max(y_n,\ methylated,\ 0) + \alpha}$$

Where α is the offset used to regularize the Beta value when both the probe intensities are low.

When the Beta value is close to 0, more sites are unmethylated, and when it is close to 1, more sites are methylated.

M-values are log-ratios used to measure the level of methylation. These are calculated by the following equation:

$$M_n = \log_2 \frac{\max(y_n,\ methylated,\ 0) + \alpha}{\max(y_n,\ unmethylated,\ 0) + \alpha}$$

When the M-value is positive, more sites are methylated, and when the M-value is negative, more sites are unmethylated.

2.4. Feature selection

As mentioned above, a major issue with the analysis of DNA methylation data is the HDLSS challenge. Only a few of these features (genes) are beneficial to the final classification model. Thus, efficient techniques are needed to select only the useful genes from the overall data set. In AI, feature selection is a widely used method to reduce the size of the feature set and keep only the informative features from the high-dimensional data [42]. There are four main approaches used for feature selection: filter, wrapper, embedded, and hybrid.

Filter methods are mainly used as a preprocessing technique and are independent of the classifier used. Filters are mostly feature ranking techniques [43]. Wrapper methods are dependent on the classifier and assist in evaluating the better feature subset by calculating the classifier accuracy [44]. However, wrapper methods are computationally intensive and sensitive to the classifier. Hence, they are not suited for DNA methylation data. Embedded methods are integrated into the classifier during the training process and help to reduce the time needed to reclassify subsets [45]. The hybrid feature selection methods combine two or more feature selection approaches, mainly filters and wrappers. In this study, two tree-based and two regularization-based embedded feature selection approaches were used.

2.4.1. Embedded feature selection

As discussed above, embedded feature selection is embedded in the classifier itself during the training phase. The embedded approaches are grouped under two categories: regularization and tree-based. Both approaches are discussed further below.

2.4.1.1. Regularization-based embedded feature selection. In regularization, a penalty is added to the various parameters of the model to restrict its freedom. This penalization is done to eliminate the noise, to avoid overfitting, and to make the model more generalized. This is done via two processes:

- **LASSO Regression:** The least absolute shrinkage and selection operator (LASSO) is a linear model that helps in eliminating irrelevant features through shrinkage [46]. In LASSO, some of the coefficients are shrunk to zero. Using this approach, some features will be multiplied by zero and removed as they do not contribute to the target feature. LASSO regression is an L1 regularization, and it adds a

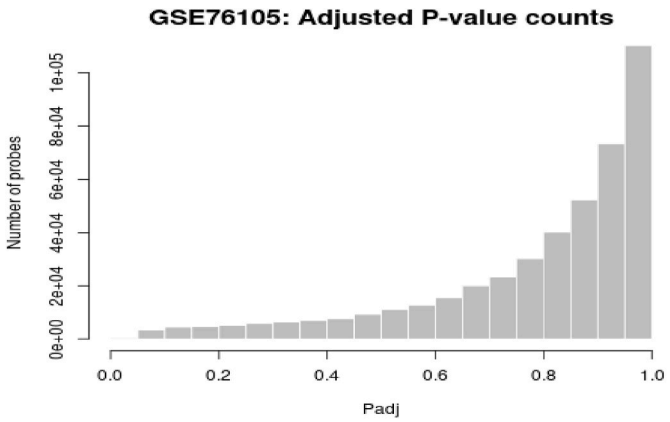

Fig. 2. The adjusted p-value distribution of the analyzed genes.

penalty that is equal to the absolute value of the coefficient's magnitude [47].

- **SVM:** SVM is a traditional and widely used linear classifier. This technique carries out classification based on the supervised approach. SVM can also be used in feature selection with its embedded ability to determine feature importance [48]. After fitting the linear SVM, the classifier coefficients are accessible. The importance of a particular feature is determined by comparing the coefficients with each other. In SVM, the parameter C (penalty factor) controls the sparsity. The smaller the value of C, the smaller the number of features selected. The unimportant features with a value below the threshold are removed from the final feature set based on the feature importance assigned [49].

2.4.1.2. Tree-based embedded feature selection.

- **Ada Boost -** Ada Boost is a set of weak learners that are easy to implement and fast to converge. The aim is to find the weak hypothesis through the weak learners. Ada Boost does not need any prior knowledge about the learners, and it can easily be implemented with other methods [50]. The weak learners are built by adjusting weights, where the weights are increased for misclassified samples and decreased for correctly classified samples [51]. The goal of the weak learner is to find the weak hypothesis.

Ada Boost selects features through the following steps:

- Initialize weights to the samples.

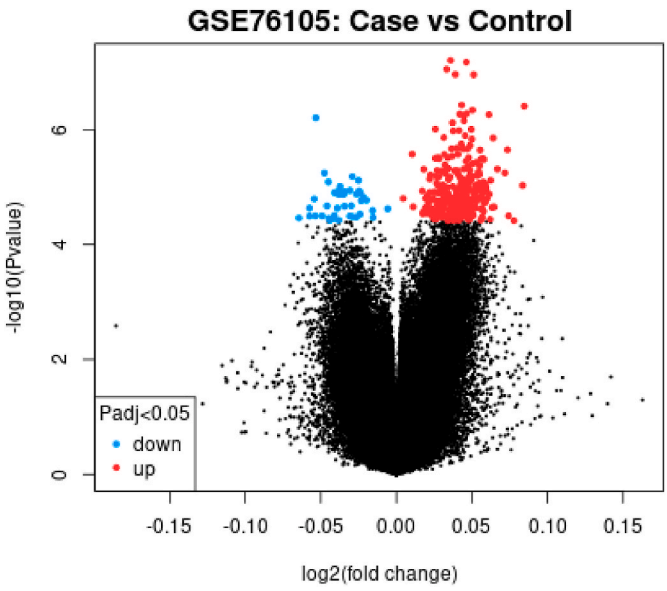

Fig. 4. A volcano plot of the data.

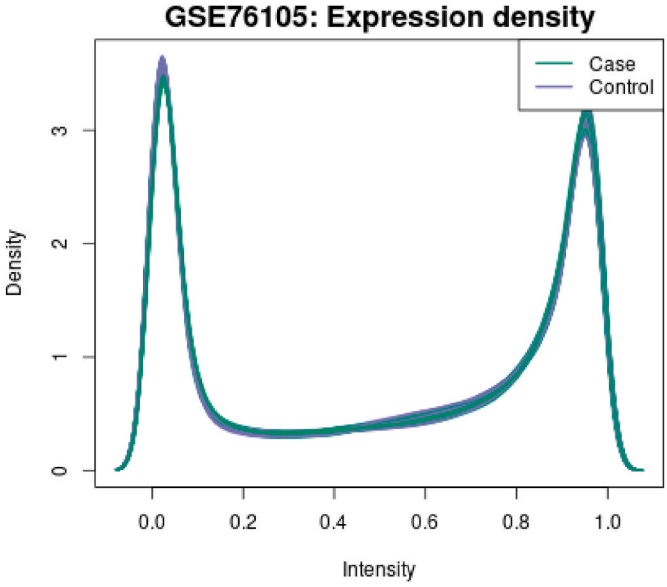

Fig. 5. A density plot of the value distribution of the selected samples.

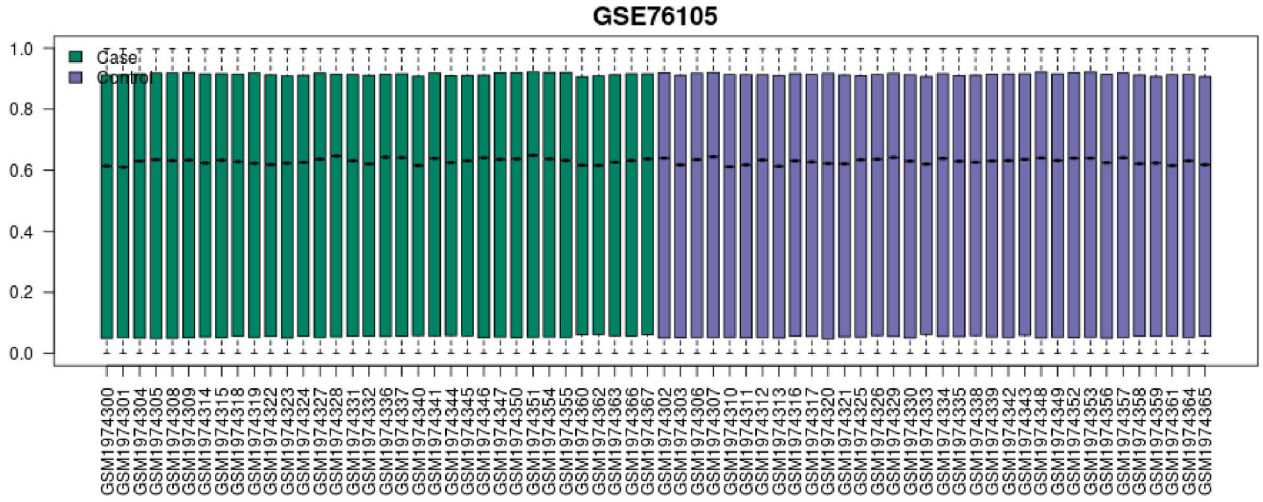

Fig. 3. A box plot of the data after normalization.

Table 1
The accuracy of the feature selection approaches as assessed by logistic regression (LR).

Embedded Feature Selection Approach	5-Fold Cross-Validation	No. of CpGs Selected	Accuracy with LR
Ada Boost	K = 1	195	0.795
	K = 2	186	0.761
	K = 3	12	0.871
	K = 4	66	0.839
	K = 5	112	0.804
Random Forest	K = 1	95	0.752
	K = 2	170	0.784
	K = 3	38	0.821
	K = 4	74	0.814
	K = 5	66	0.792
LASSO	K = 1	18	0.857
	K = 2	46	0.831
	K = 3	58	0.798
	K = 4	20	0.839
	K = 5	105	0.786
SVM	K = 1	210	0.748
	K = 2	56	0.792
	K = 3	28	0.832
	K = 4	117	0.775
	K = 5	21	0.848

- After generating a feature set, a single feature is trained to select the best feature.
- The strong classifier and the weights of the sample are updated.
- Update the feature subset and repeat until stopping criteria are met.
- **Random Forest:** Random forest is a tree-based machine learning approach with simple interpretability, accurate predictive performance, and low overfitting. This straightforward technique estimates the importance of a variable that contributes to a target variable [52]. The random forest is constructed with features extracted randomly. The dataset is divided into two groups, and each group has features that are similar and different to each other. Thus, feature importance is decided based on the purity of each group [53]. For classification, the purity can be determined using the Information Gain or Gini Index [54]. In random forest, a feature with a minor impurity is considered the more important feature and vice versa. The final importance of the feature is estimated by calculating the average decrease in impurity from each feature.

2.5. Classification

An Enhanced Deep Recurrent Neural Network (EDRNN) with stopping criteria was used to classify the cases and controls from the AD dataset.

2.5.1. Deep Recurrent Neural Networks

In a traditional neural network, the inputs and output neurons are independent of each other, whereas, in a recurrent neural network (RNN), there is an internal state or hidden state in addition to the input space. The hidden state acts as a 'memory,' which carries information about previous states. In a RNN, the dependent activations are converted to independent activations. All the layers are allocated with the same biases and weights to reduce the complexity of the model. As all the weights and biases are the same, the hidden layers can be combined into one recurrent layer. The training protocol for this type of network is outlined below:

2.5.1.1. Training.

- Provide input to the network.
- Calculate the current state using the previous state and set of current input states using the following equation:

$$CS_t = f(CS_{f-1},\ IS_t)$$

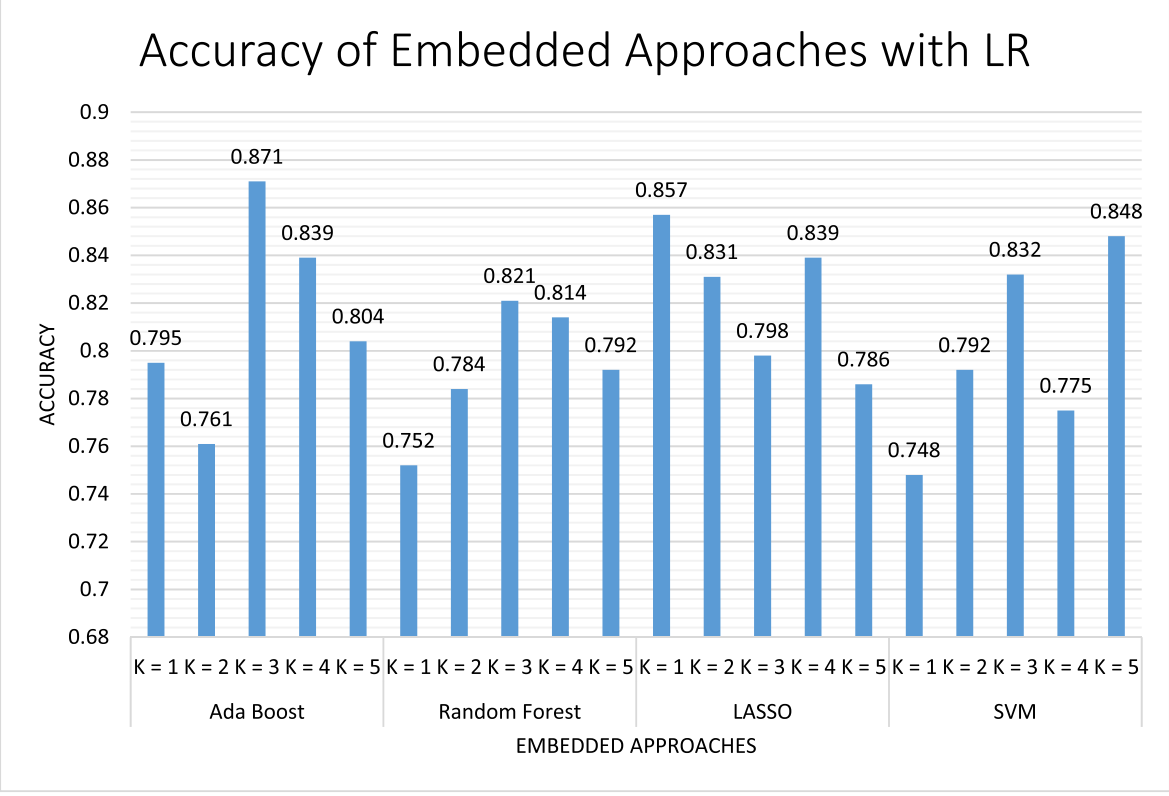

Fig. 6. Performance comparison of the embedded approaches.

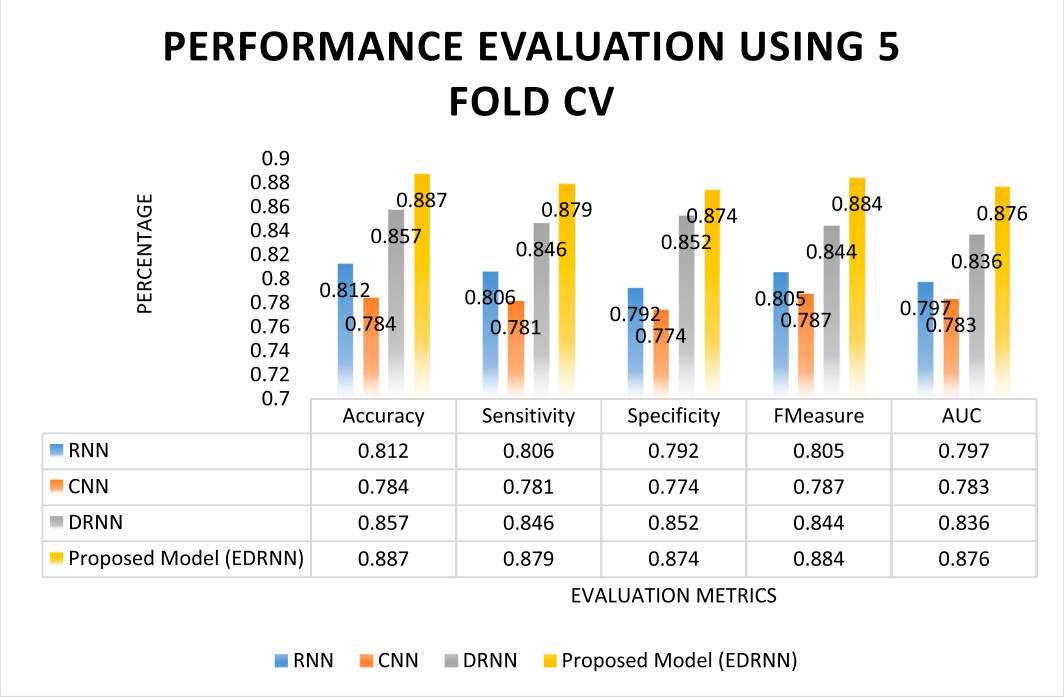

Fig. 7. Performance comparison of the classification models.

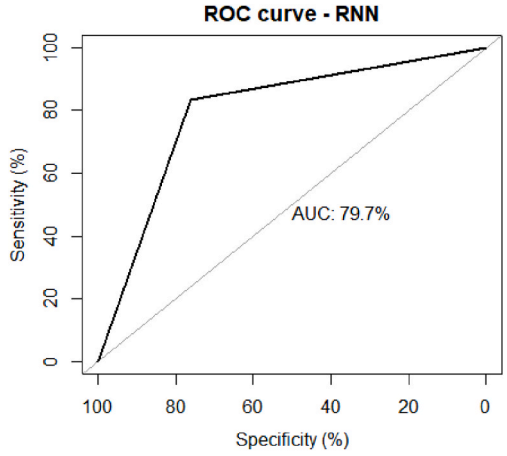

Fig. 8. AUC-ROC of RNN.

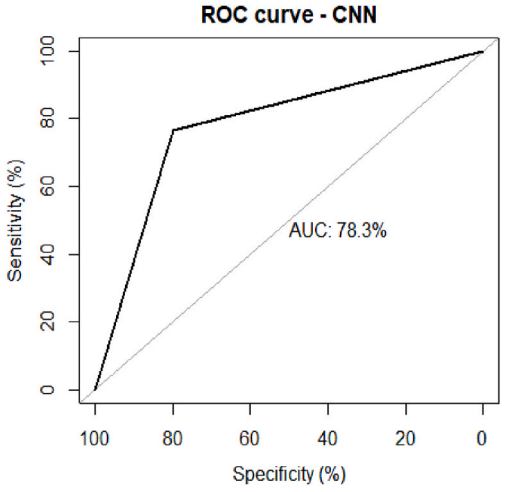

Fig. 9. AUC-ROC of CNN.

Where CS_t = Current State, CS_{f-1} = Previous State, and IS_t = Input State.

- The current state will become the previous state for the next time.
- Repeat for all the steps and the final current state is used to estimate the output.
- The output is compared with the target output and the error is calculated.
- The weights are updated based on the error using the following equation:

$$OL_t = W_{OL} CS_t$$

Where OL_t = Output Layer and W_{OL} = Weight of output layer.

3. Dataset

The dataset used in the current study was downloaded from the GEO Omnibus dataset (GSE76105). The AD DNA methylation profiling was done using an Illumina Infinium HumanMethylation450 array platform with tissue samples from the superior temporal gyrus (STG). The dataset consists of 68 records for 34 cases and 34 controls with 461,272 features (CpG sites) All of the implementations were done using R Studio. For pre-processing, the Bioconductor package from R Studio was used.

4. Performance evaluation

The proposed approach was evaluated using two cross validation techniques, 5-fold cross validation and leave-one-out cross validation (LOOCV; see below for descriptions of these techniques). Accuracy, sensitivity, specificity and FMeasure statistical measures were used as evaluation scores to compare the results of the implemented approach with the existing approaches. Also, we used the ROC curve, which is a standard method to evaluate the prediction and classification models including gene expression and DNA methylation-based analysis [55].

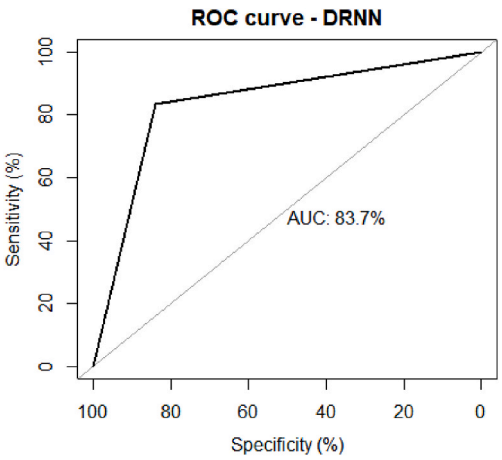

Fig. 10. AUC-ROC of DRNN.

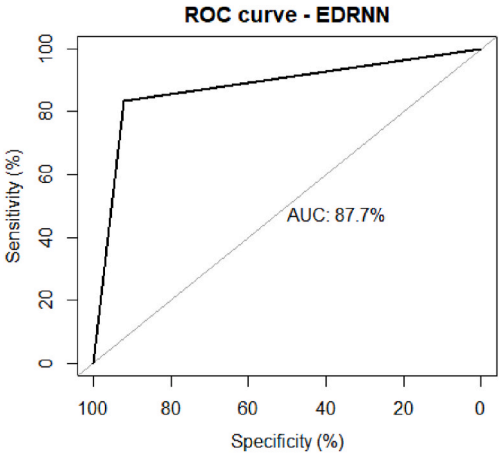

Fig. 11. AUC-ROC of EDRNN.

The AUC-ROC shows the trade-off between specificity and sensitivity.

- ***K-fold Cross Validation:*** In k-fold cross validation, the dataset is divided into an equal 'k' number of groups or folds. In the folds, the first one is considered as the validation set and the remaining k-1 folds are used to fit the model.
- ***LOOCV:*** In LOOCV, each sample is treated as a validation set, whereas the remaining n-1 samples are treated as the training set. LOOCV is used for datasets with lower sample sizes as it is exhaustive and has a high computational cost.

5. Results

Pre-processing techniques were applied to the retrieved data set and poorly performing samples (p-values > 0.01) were eliminated during the quality control. The adjusted p-value distribution of the analyzed genes is shown in Fig. 2. After applying the quality control, the data was highly skewed. Thus, a log2 transformation was applied and Z-scores were used to normalize the data for further processing and to make it comparable across all platforms. The data after normalization are shown as a box plot in Fig. 3. Once the data were normalized, downstream analysis was done to determine the differentially methylated positions (DMPs). The threshold used for this analysis was a fold change (FC) > 2 and p-value > 0.01. A volcano plot which is useful in visualizing differentially expressed genes by plotting the statistical significance versus the magnitude of change is shown Fig. 4. Fig. 5 shows the density of the value distribution of the selected samples. If the density curves differ from each other, it indicates that the data needs more normalization.

After pre-processing, the data was further processed to select only the relevant genes for the classification process. Ada Boost, Random forest, LASSO, and SVM embedded feature selection approaches were implemented to choose the relevant CpG sites for further processing. These approaches were validated using logistic regression (LR). The data was split into a training and testing set with a 5-fold-cross validation. All four embedded feature selection methods were applied to the data, and the number of CpG sites selected with their corresponding accuracies are shown in Table 1. The parameters for the approaches were tuned using grid search. The plot shown in Fig. 6 shows that Ada Boost worked well with the AD DNA methylation dataset and selected the CpG sites with a better accuracy. The highest accuracy (87%) of the Ada Boost was seen during the 3rd fold by selecting 12 CpG sites. We identified 6 hypo- and

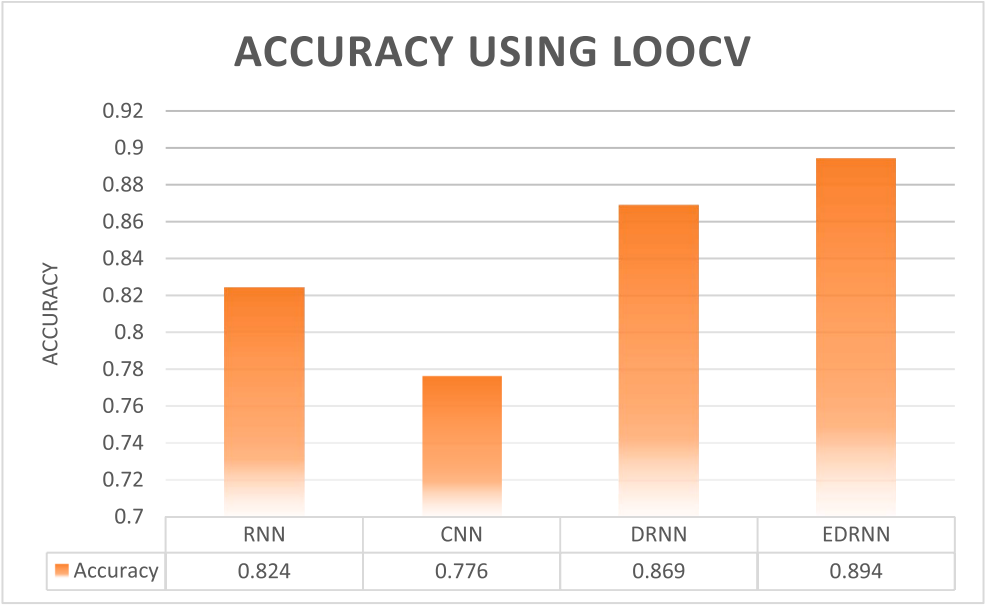

Fig. 12. Accuracy of the of the implemented models using LOOCV.

6 hyper-methylated regions with 9 genes associated with them. The genes associated with the CpG sites were used for classifying the AD patients in the classification stage.

After applying Ada Boost, the selected CpG sites were used for classifying the AD patients. For classification, we used an EDRNN with stopping criteria. We introduced the depth in the RNN in the hidden layer to the output area. The hyper parameters of the EDRNN were tuned using Bayesian optimization with five-fold cross-validation. After the tuning, the parameters chosen were three hidden layers and a dropout rate of 0.7 for the final EDRNN.

The primary issue with any deep net analysis is overfitting. To avoid overfitting and to speed up the model convergence, stopping criteria were introduced. After every 20 epochs, the test accuracy of the last ten epochs was compared with the average test accuracy. In this way, the converging or decreasing accuracy is estimated. Also, along with the test accuracy, the training accuracy was compared. If the model is converging and meets both of the conditions mentioned above, the learning is stopped.

The proposed model was compared with a traditional RNN, a Convolutional Neural Network (CNN), and a Deep RNN (DRNN) without the stopping criteria. The models were evaluated using 5-fold cross validation and LOOCV statistical approaches with the help of evaluation scores, including accuracy, sensitivity, specificity, and FMeasure. In both the cross validations, the performance evaluation showed that the DRNN and EDRNN performances were better than the traditional RNN and the CNN. The results of the 5-fold cross validation of the implemented models are shown in Fig. 7. The average accuracy of LOOCV of RNN, CNN, DRNN and EDRNN were 82.4%, 77.6%, 86.9% and 89.5% respectively. Also, the AUC of EDRNN (87.7%) was better than the other implemented models (Figs. 8–11). The performance evaluation of the models using LOOCV is shown in Fig. 12. Overall, the proposed method with the stopping criteria eliminates the overfitting issues and improves the accuracy.

In addition, a Wilcoxon signed-rank test (WSRT) was used to test the statistical significance of the proposed approach as compared the other existing approaches. To perform the WSRT, a null hypothesis (H_0 – both approaches perform similarly) and an alternate hypothesis (H_1 – there is a significant difference between the two approaches) were proposed. The WSRT was performed on the 5-fold split dataset and initially calculated the difference in performance of the two algorithms. WSRT involves the size of the observation (n), level of significance (α), the critical value (taken from the standard WSRT critical values table), and the test value (min positive ranks, negative ranks). With the estimated ranks from the WSRT, it was found that for RNN-EDRNN the critical value was 0 and the test value was 6, for CNN-EDRNN the critical value was 0 and the test value was 4, and for DRNN-EDRNN the critical value was 0 and the test value was 3. The level of significance used here was $p < 0.05$ and the test was one-tailed. To reject the null hypotheses, the test value must be greater than the critical value. In this case, the test value was larger than the critical value for all three comparisons. Thus, the null hypotheses were rejected and the alternate hypotheses, that the EDRNN is significantly different from the other implemented approaches, was accepted.

6. Discussion and conclusions

AD is commonly found in the elderly. However, there are rare cases of early-onset. This disease progresses slowly and is ultimately fatal. Thus, early detection is needed to slow down the progression of symptoms, and diagnosing AD at the MCI stage is critical. However, it has been difficult to diagnose AD with the existing methods. Brain imaging is widely used to identify AD, but there are challenges associated with using this high-end equipment. Thus, more research is needed on molecular-based AD detection. Here, we implemented a deep learning model to classify AD using DNA methylation profiles. The AD dataset had 68 samples with 34 cases and 34 controls. The associated number of CpG sites in this data set was huge and, to have a high accuracy classification, the irrelevant features must be eliminated. Hence, the data was preprocessed to eradicate the poor signal probes and missing values, and were also normalized to allow for comparisons across all of the platforms. In addition, downstream analysis was done to identify the DMPs.

In total, 12 differentially methylated positions were identified, of which 6 were hypo-methylated and 6 were hyper-methylated. After preprocessing, four embedded-based feature selection methods, Ada Boost, Random forest, LASSO, and SVM, were implemented with 5-fold cross-validation and LOOCV. The hyper parameters of the approaches were tuned using grid search, and all the feature selection approaches were evaluated using LR. The results (tabulated in Table 1) show that the tree-based embedded approach Ada Boost gave a better accuracy than the other three approaches. Thus, Ada Boost was chosen to select the features for classifying AD patients using the proposed deep learning-based classification model.

The selected features using Ada Boost in 5-folds of cross-validation were processed as inputs in the proposed deep learning model. An EDRNN with stopping criteria was implemented to avoid the overfitting problem. Depth in the RNN in the hidden layer to the output area was introduced for better accuracy. The proposed model was compared with three state- of-the-art approaches, including traditional RNN, CNN, and DRNN. The hyperparameters of the models were tuned using the Bayesian optimization technique with 5-fold cross-validation. In the 3rd fold, the Ada Boost selected 12 CpG sites which gave higher accuracy than the other folds. The results of LOOCV also showed there was improved accuracy in the proposed approach compared to the other existing approaches.

The genes associated with the 12 CpG sites selected by the Ada Boost were MS4A4A, MYNN, TXNIP, CORO2B, NOG, BEX2, PIGA, FAM82A1, and CDKN1C. Among these, MYNN and BEX2 are reported in the AlzGene database. MYNN is important for controlling gene expression, and BEX2 is considered critical in inhibiting neuronal differentiation. The selected CpG sites were used for the classification of the AD patients. We implemented the EDRNN and compared the results with the other three approaches, and tabulated the results. The results show that the EDRNN performed better than the other techniques (see Fig. 7). Future work should include analyzing various forms of molecular data sets, such as gene expression, Single Nucleotide Polymorphism (SNP), and Copy Number Variation (CNV), to identify other genes associated with AD. Also, computational methods and bioinformatics tools should be used for a further understanding of the results.

Declaration of competing interest

None declared.

References

[1] T. Kt, et al., "Neurological disorders," Mosaic autoimmun. Nov. Factors autoimmune Dis, Accessed: Sep. 09, 2021, [Online]. Available, http://europepmc.org/books/NBK361950, May 2016, 541-548.
[2] F.J. Charlson, A.J. Baxter, T. Dua, L. Degenhardt, H.A. Whiteford, T. Vos, Excess mortality from mental, neurological and substance use disorders in the Global Burden of Disease Study 2010, Epidemiol. Psychiatr. Sci. 24 (2) (Apr. 2015) 121–140.
[3] A.A. Farooqui, Effect of lifestyle, aging, and phytochemicals on the onset of neurological disorders, Phytochem. Signal Transduction, Neurol. Disord. (2013) 1–29, https://doi.org/10.1007/978-1-4614-3804-5_1.
[4] N.Z. Baquer, et al., A metabolic and functional overview of brain aging linked to neurological disorders, Biogerontology 2009 104 10 (4) (Apr. 2009) 377–413, https://doi.org/10.1007/S10522-009-9226-2.
[5] V.D. Ks Anand, Hippocampus in health and disease: an overview, Ann. Indian Acad. Neurol. 15 (4) (Oct. 2012) 239, https://doi.org/10.4103/0972-2327.104323.
[6] A. Ashraf, S. Naz, S.H. Shirazi, I. Razzak, M. Parsad, Deep transfer learning for alzheimer neurological disorder detection, Multimed. Tool. Appl. (2021) 1–26, https://doi.org/10.1007/S11042-020-10331-8. Jan. 2021.
[7] P.V.W. Henderson, Alzheimer's disease and other neurological disorders 10 (SUPPL. 2) (Oct. 2009) 92, https://doi.org/10.1080/13697130701534097. https://doi.org/10.1080/13697130701534097, 96.

[8] K.L. Lanctôt, R.D. Rajaram, N. Herrmann, Review: therapy for Alzheimer's disease: how effective are current treatments? 2 (3) (Apr. 2009) 163–180, https://doi.org/10.1177/1756285609102724. https://doi.org/10.1177/1756285609102724.

[9] M.I.G.X. Mi Razzak, Big data analytics for preventive medicine, Neural Comput. Appl. 32 (9) (May 2020) 4417–4451, https://doi.org/10.1007/s00521-019-04095-y.

[10] Z. S. Khachaturian, "Diagnosis of Alzheimer's disease," Arch. Neurol., vol. 42, no. 11, pp. 1097–1105, Nov. 1985, doi: 10.1001/ARCHNEUR.1985.04060100083029.

[11] K. Yang, E.A. Mohammed, A review of artificial intelligence technologies for early prediction of Alzheimer's disease, Accessed: Sep. 09, 2021. [Online]. Available, https://arxiv.org/abs/2101.01781v1, Dec. 2020.

[12] K.G. Yiannopoulou, S.G. Papageorgiou, Current and future treatments for Alzheimer's disease, Ther. Adv. Neurol. Disord. 6 (1) (2013) 19–33, https://doi.org/10.1177/1756285612461679.

[13] S. Wang, H. Wang, Y. Shen, X. Wang, Automatic recognition of mild cognitive impairment and alzheimers disease using ensemble based 3D Densely connected convolutional networks, in: Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018, Jan. 2019, pp. 517–523, https://doi.org/10.1109/ICMLA.2018.00083.

[14] V. García, J.S. Sánchez, L. Cleofas-Sánchez, H.J. Ochoa-Domínguez, F. López-Orozco, An insight on the 'large <Emphasis Type="Italic">G</Emphasis>, small <Emphasis Type="Italic">n</emphasis>' problem in gene-expression microarray classification, Lect. Notes Comput. Sci. 10255 (2017) 483–490, https://doi.org/10.1007/978-3-319-58838-4_53. LNCS.

[15] S. de la Fuente Garcia, C. W. Ritchie, and S. Luz, "Artificial intelligence, speech, and language processing approaches to Monitoring Alzheimer's disease: a systematic review," J. Alzheim. Dis., vol. 78, no. 4, pp. 1547–1574, Jan. 2020, doi: 10.3233/JAD-200888.

[16] C. Salvatore, A. Cerasa, I. Castiglioni, MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia 24 Months before probable diagnosis, Front. Aging Neurosci. (May 2018) 135, https://doi.org/10.3389/FNAGI.2018.00135, vol. 0, no. MAY.

[17] C.K. Fisher, A.M. Smith, J.R. Walsh, Machine learning for comprehensive forecasting of Alzheimer's Disease progression, Sci. Reports 2019 91 9 (1) (Sep. 2019) 1–14, https://doi.org/10.1038/s41598-019-49656-2.

[18] A. Ortiz, J. Munilla, J.M. Górriz, J. Ramírez, "Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease," 26 (7) (Aug. 2016), https://doi.org/10.1142/S0129065716500258. https://doi.org/10.1142/S0129065716500258.

[19] S. Thapa, P. Singh, D.K. Jain, N. Bharill, A. Gupta, M. Prasad, Data-driven approach based on feature selection technique for early diagnosis of Alzheimer's disease, Proc. Int. Jt. Conf. Neural Networks (Jul. 2020), https://doi.org/10.1109/IJCNN48605.2020.9207359.

[20] S. Murugan, et al., "DEMNET: a deep learning model for early diagnosis of alzheimer diseases and dementia from MR images, IEEE Access 9 (2021) 90319–90329, https://doi.org/10.1109/ACCESS.2021.3090474.

[21] F. Segovia, J.M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, Early diagnosis of Alzheimer's disease based on partial least squares and support vector machine, Expert Syst. Appl. 40 (2) (Feb. 2013) 677–683, https://doi.org/10.1016/J.ESWA.2012.07.071.

[22] M.H. Modarres, et al., Early diagnosis of Alzheimer's dementia with the artificial intelligence-based Integrated Cognitive Assessment, Alzheimer's Dementia 16 (S6) (Dec. 2020) e042863, https://doi.org/10.1002/ALZ.042863.

[23] R. Pandya, S. Nadiadwala, R. Shah, M. Shah, Buildout of methodology for Meticulous diagnosis of K-complex in EEG for aiding the detection of Alzheimer's by artificial intelligence, Augment. Hum. Res 5 (1) (Oct. 2019) 1–8, https://doi.org/10.1007/S41133-019-0021-6, 2019 51.

[24] S. Esmaeilzadeh, D.I. Belivanis, K.M. Pohl, E. Adeli, End-to-end Alzheimer's disease diagnosis and biomarker identification, Lect. Notes Comput. Sci. 11046 (Sep. 2018) 337–345, https://doi.org/10.1007/978-3-030-00919-9_39.

[25] A. Yilmaz, et al., A community-based study identifying metabolic biomarkers of mild cognitive impairment and Alzheimer's disease using artificial intelligence and machine learning, J. Alzheim. Dis. 78 (4) (Jan. 2020) 1381–1392, https://doi.org/10.3233/JAD-200305.

[26] M. Song, H. Jung, S. Lee, D. Kim, M. Ahn, Diagnostic classification and biomarker identification of Alzheimer's disease with random forest algorithm, Brain Sci. 11 (4) (Apr. 2021) 453, https://doi.org/10.3390/BRAINSCI11040453, 2021, Vol. 11, Page 453.

[27] L. Scheubert, M. Luštrek, R. Schmidt, D. Repsilber, G. Fuellen, Tissue-based Alzheimer gene expression markers-comparison of multiple machine learning approaches and investigation of redundancy in small biomarker sets, BMC Bioinf. 13 (1) (2012), https://doi.org/10.1186/1471-2105-13-266.

[28] T. Lee, H. Lee, Prediction of Alzheimer's disease using blood gene expression data, Sci. Reports 2020 101 10 (1) (Feb. 2020) 1–13, https://doi.org/10.1038/s41598-020-60595-1.

[29] L. Wang, Z.-P. Liu, Detecting Diagnostic biomarkers of Alzheimer's disease by integrating gene expression data in six brain regions, Front. Genet. 157 (2019), https://doi.org/10.3389/FGENE.2019.00157 vol. 0, no. MAR,.

[30] C. Park, J.R. Kim, J. Kim, S. Park, Machine learning-based identification of genetic interactions from heterogeneous gene expression profiles, PLoS One 13 (7) (2018) 1–15, https://doi.org/10.1371/journal.pone.0201056.

[31] S. Perera, K. Hewage, C. Gunarathne, R. Navarathna, D. Herath, R.G. Ragel, Detection of Novel biomarker genes of Alzheimer's disease using gene expression data, MERCon 2020 - 6th Int. Multidiscip. Moratuwa Eng. Res. Conf. Proc. (Jul. 2020) 1–6, https://doi.org/10.1109/MERCON50084.2020.9185336.

[32] R. Ramaswamy, P. Kandhasamy, S. Palaniswamy, Feature selection for Alzheimer's gene expression data using modified binary particle swarm optimization, 2021, https://doi.org/10.1080/03772063.2021.1962747. https://doi.org/10.1080/03772063.2021.1962747.

[33] W.S. Liang, et al., Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain 28 (3) (Feb. 2007) 311–322, https://doi.org/10.1152/PHYSIOLGENOMICS.00208.2006. https://doi.org/10.1152/physiolgenomics.00208.2006.

[34] A. Ni, A. Sethi, For the A. D. N. Initiative, "Functional Genetic Biomarkers of Alzheimer's Disease and Gene Expression from Peripheral Blood, " *bioRxiv*, Jan. 2021, p. 2021, https://doi.org/10.1101/2021.01.15.426891, 01.15.426891.

[35] J. Ren, B. Zhang, D. Wei, Z. Zhang, Identification of methylated gene biomarkers in patients with Alzheimer's disease based on machine learning, BioMed Res. Int. 2020 (2020), https://doi.org/10.1155/2020/8348147.

[36] A. Unnikrishnan, W.M. Freeman, J. Jackson, J.D. Wren, H. Porter, A. Richardson, The role of DNA methylation in epigenetics of aging, Pharmacol. Ther. 195 (Mar. 2019) 172–185, https://doi.org/10.1016/J.PHARMTHERA.2018.11.001.

[37] A. Bird, DNA methylation patterns and epigenetic memory, Genes Dev. 16 (1) (Jan. 2002) 6–21, https://doi.org/10.1101/GAD.947102.

[38] C.S. Wilhelm-Benartzi, et al., Review of processing and analysis methods for DNA methylation array data, Br. J. Cancer 2013 1096 109 (6) (Aug. 2013) 1394–1402, https://doi.org/10.1038/bjc.2013.496.

[39] W. Z, W. X, and W. Y, "A framework for analyzing DNA methylation data from Illumina Infinium HumanMethylation450 BeadChip, BMC Bioinf. 19 (5) (Apr. 2018), https://doi.org/10.1186/S12859-018-2096-3.

[40] T. N and T. J, Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation, Epigenomics 4 (3) (Jun. 2012) 325–341, https://doi.org/10.2217/EPI.12.21.

[41] D. P, et al., Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis, BMC Bioinf. 11 (Nov. 2010), https://doi.org/10.1186/1471-2105-11-587.

[42] J.R. Vergara, P.A. Estévez, A review of feature selection methods based on mutual information, Neural Comput. Appl. 24 (1) (2014) 175–186, https://doi.org/10.1007/s00521-013-1368-0.

[43] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, M. Lang, Benchmark for filter methods for feature selection in high-dimensional classification data, Comput. Stat. Data Anal. 143 (Mar. 2020) 106839, https://doi.org/10.1016/J.CSDA.2019.106839.

[44] S. Jadhav, H. He, K. Jenkins, Information gain directed genetic algorithm wrapper feature selection for credit rating, Appl. Soft Comput. 69 (Aug. 2018) 541–553, https://doi.org/10.1016/J.ASOC.2018.04.033.

[45] H. Liu, M. Zhou, Q. Liu, An embedded feature selection method for imbalanced data classification, IEEE/CAA J. Autom. Sin. 6 (3) (May 2019) 703–715, https://doi.org/10.1109/JAS.2019.1911447.

[46] V. Roth, The generalized LASSO, IEEE Trans. Neural Network. 15 (1) (Jan. 2004) 16–28, https://doi.org/10.1109/TNN.2003.809398.

[47] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B 58 (1) (Jan. 1996) 267–288, https://doi.org/10.1111/J.2517-6161.1996.TB02080.X.

[48] Y.-W. Chen, C.-J. Lin, Combining SVMs with various feature selection strategies, Stud. Fuzziness Soft Comput. 207 (2006) 315–324, https://doi.org/10.1007/978-3-540-35488-8_13.

[49] H.C. Kim, S. Pang, H.M. Je, D. Kim, S.Y. Bang, Constructing support vector machine ensemble, Pattern Recogn. 36 (12) (Dec. 2003) 2757–2767, https://doi.org/10.1016/S0031-3203(03)00175-4.

[50] R. Wang, AdaBoost for feature selection, classification and its relation with SVM, A review, Phys. Procedia 25 (Jan. 2012) 800–807, https://doi.org/10.1016/J.PHPRO.2012.03.160.

[51] D.B. Redpath, K. Lebart, Boosting feature selection, Lect. Notes Comput. Sci. 3686 (2005) 305–314, https://doi.org/10.1007/11551188_33. PART I.

[52] M. Saraswat, K.V. Arya, Feature selection and classification of leukocytes using random forest, Med. Biol. Eng. Comput. 52 (12) (Oct. 2014) 1041–1052, https://doi.org/10.1007/S11517-014-1200-8, 2014 5212.

[53] L. Breiman, Random forests, Mach. Learn. 45 (1) (Oct. 2001) 5–32, https://doi.org/10.1023/a:1010933404324.

[54] C. Nguyen, Y. Wang, H.N. Nguyen, in: Random Forest Classifier Combined with Feature Selection for Breast Cancer Diagnosis and Prognostic, vol. 2013, May 2013, pp. 551–560, https://doi.org/10.4236/JBISE.2013.65070.

[55] Supervised machine learning models and protein-protein interaction network analysis of gene expression profiles induced by omega-3 polyunsaturated fatty acids.", accessed, https://assets.researchsquare.com/files/rs-49619/v2/1b638212-e354-4314-88b9-fd58f513e636.pdf?c=1631853372. (Accessed 8 November 2021).