

Comparative Analysis of Performance Metrics for Machine Learning Classifiers with a Focus on Alzheimer's Disease Data

Sivakani Rajayyan ¹ , Syed Masood Mohamed Mustafa ² 

¹ Department of Computer Science and Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, India

² Department of Computer Applications, B.S. Abdur Rahman Crescent Institute of Science and Technology, India

Corresponding author: Syed Masood Mohamed Mustafa (ms.masood@crescent.education)

Abstract

Alzheimer's disease is a brain memory loss disease. Usually, it will affect persons over 60 years of age. The literature has revealed that it is quite difficult to diagnose the disease, so researchers are trying to predict the disease in the early stage. This paper proposes a framework to classify Alzheimer's patients and to predict the best classification algorithm. The Bestfirst and CfsubsetEval methods are used for feature selection. A multi-class classification is done using machine learning algorithms, namely the naïve Bayes algorithm, the logistic algorithm, the SMO/SMV algorithm and the random forest algorithm. The classification accuracy of the algorithms is 67.68%, 84.58%, 87.42%, and 88.90% respectively. The validation applied is 10-fold cross-validation. Then, a confusion matrix is generated and class-wise performance is analysed to find the best algorithm. The ADNI database is used for the implementation process. To compare the performance of the proposed model, the OASIS dataset is applied to the model with the same algorithms and the accuracy of the algorithms is 98%, 99%, 99% and 100% respectively. Also, the time for the model construction is compared for both datasets. The proposed work is compared with existing studies to check the efficiency of the proposed model.

Keywords

Cognitive normal; Early mild cognitive impairment; Late mild cognitive impairment; Naïve Bayes algorithm; Logistic regression algorithm; SVM algorithm; Random forest algorithm.

Citation: Rajayyan, S., & Mustafa, S. M. M. (2023). Comparative Analysis of the Performance Metrics for Machine Learning Classifiers with a Focus on Alzheimer's Disease Data. *Acta Informatica Pragensia*, 12(1), 54–70. <https://doi.org/10.18267/j.aip.198>

Special Issue Editors: Mazin Abed Mohammed, University of Anbar, Iraq
Seifedine Kadry, Noroff University College, Norway
Oana Geman, Ștefan cel Mare University of Suceava, Romania

Academic Editor: Zdenek Smutny, Prague University of Economics and Business, Czech Republic

Copyright: © 2023 by the author(s). Licensee Prague University of Economics and Business, Czech Republic.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY 4.0).

1 Introduction

Alzheimer's disease (AD) is a type of brain disorder. It affects memory little by little and in the end, it will not allow the person to think. It happens due to death of memory cells in the brain. A protein is deposited in the form of an amyloid and tangles in the brain. Due to this, memory cells die and in the end, the size of the brain shrinks, so memory loss occurs. The initial stage of brain damage affects the hippocampus. The next stage of memory deficiency is mild cognitive impairment; widespread death of memory cells leads to the severe stage and inevitably death occurs. AD is the sixth leading disease that causes death in the United States (National Institute on Aging, 2021).

1.1 Stages of Alzheimer's disease

Alzheimer's disease has been categorized into four stages (MCI Screen, 2021); they are shown in Figure 1:

- **Cognitive normal (CN)**
Cognitive normal is the normal cognitive ageing process (MCI Screen, 2021; Han et al., 2020). People in this category experience healthy ageing. They do not have any AD symptoms.
- **Early mild cognitive impairment (EMCI)**
Early mild cognitive impairment is the early stage of AD (MCI Screen, 2021; Guo et al., 2020). In this stage, the small changes in the cognitive normal are considered EMCI. Not all EMCI stages progress to AD: some of the EMCI does return to the cognitive normal stage. Thus, this stage is considered harmless.
- **Late mild cognitive impairment (LMCI)**
Late mild cognitive impairment is the next stage of EMCI (MCI Screen, 2021; Guo et al., 2020). Most patients in this stage will progress to AD. Few patients return to the EMCI stage.
- **Alzheimer's disease (AD)**
This is the final stage of the memory loss disease (MCI Screen, 2021; Han et al., 2020; Guo et al., 2020). This is an incurable stage.

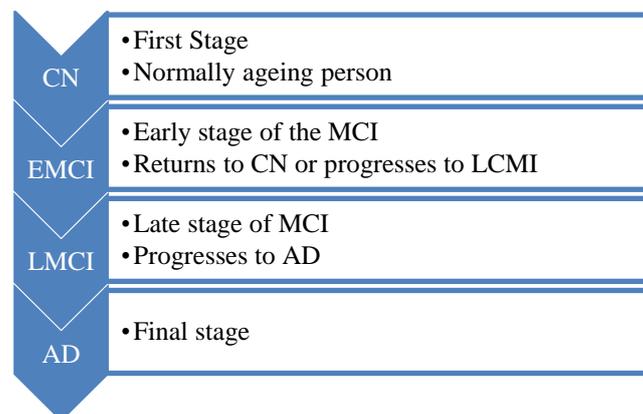


Figure1. Stages of Alzheimer's disease.

The first stage is the CN; in this stage, people can think and react. This happens due to the normal ageing process. The next stage is EMCI; it affects the medial temporal lobe in the hippocampus and exhibits symptoms of short-term memory loss. It progresses to the next level called LMCI, which affects the lateral and parietal lobes of the brain. Symptoms of this stage include reading problems, poor object recognition and poor sense of direction. The next stage is AD, which affects the front temporal lobe and the occipital lobe of the brain. Symptoms of this stage include poor judgment, impulsivity, short attention and visual problems.

The first level of AD starts from the entorhinal cortex and the hippocampus, and then it slowly influences other parts of the brain. Researchers have identified that 80% of people with MCI progress to AD within 6 years. The Mayo clinic found that 15-20% of MCI patients progress to AD each year and the progression rate is 1-2% (MCI Screen, 2021).

1.2 Literature review

A pre-training model was introduced by Han et al. (2020) for extraction and transfer of features with a focus on age-related attributes. The result was compared with 8 classification models and proved that this is a competitive model for predicting the MCI to AD progression. The future plan was to explore the model performance on multiple neuroimaging attributes.

Guo et al. (2020) introduced a cerebral similarity network using the sparse regression algorithm. In the first level, a new dynamic morphological feature was defined, and based on that the network was constructed. SVM classification and the leave-one-out cross-validation were made. The accuracy of this method was 92.31% and was considered a more sensitive biomarker in the prediction of MCI to AD progression. In future work, the validation would be done on a larger dataset.

To find the different stages in AD, the BLS diagnosing model and convolutional variants were used by Gao et al. (2020) and Sivakani and Ansari (2020). The validity of the model was evaluated using MRI data images from the ADNI database. The accuracy of the proposed model was compared with the PCA-SVM and the InceptionNet techniques and proved to be the best model. Feature extraction was also a focus of the paper.

Syed et al. (2020) proposed an ensemble classification model using the linear SVM and LR algorithms. Recursive feature elimination (RFE) and L1 regularization method were applied for feature selection. An MCI dataset was used for the implementation and the biomarkers focused on were cystatin, matrix metalloproteinases and the tau protein. The proposed model was deployed as a web-based application for the detection of early AD.

You et al. (2020) proposed a cascade neural network for faster and more accurate classification. It was a two-step process: feature extraction and classification. Using a sensor, EEG data were collected from patients. Three-way classification was done and the proposed model gave the best results.

A convolutional auto-encoder was proposed by Oh et al. (2019) for binary classification. Unsupervised learning technique was applied for the AD, and NC classification and the supervised transfer learning technique was applied for the pMCI and sMCI classification. Gradient-based visualization was done. The accuracy for AD classification and pMCI classification was 86.60% and 73.95% respectively.

A novel framework was proposed by Gupta et al. (2019), based on machine learning techniques. The biomarkers used for the classification were a combination of FDG-PET, sMRI, CSF and APOE. The ADNI database was used for the study. A multi-classification was done using the kernel SVM classifier and the grid-search method. The AU-ROC of this model gave better results than other state-of-the-art methods.

Eitel and Ritter (2019) trained a CNN model for checking the robustness of the methods of gradient input, guided back-propagation, layer-wise relevance propagation and occlusion for AD classification. A visual comparison of the methods was made. Li et al. (2018) introduced the MKSCDDL algorithm for AD classification, and the results were better than other state-of-the-art methods.

Martinez-Murcia et al. (2020) proposed a deep convolutional autoencoder (CAE) model for AD diagnosis. Feature extraction and classification were done using the regression method. More than 80% of classification accuracy was achieved. Maqsood et al. (2019) developed an efficient classification method with the utilization of a pre-trained network, AlexNet, and CNN, and the performance was evaluated using the OASIS database. The accuracy for multi-class classification was 92.85%.

Nozadi et al. (2018) designed a pipeline for the group classification with learned features of PET images, and the evaluation was made on the ADNI database with an accuracy of 91.2%. A novel grading biomarker was developed by Tong et al. (2017) using sparse representation techniques with selected features, age and cognitive measurements to provide a more accurate prediction of progression from MCI to AD. The AUC range of this biomarker was 84-92%.

Zhang et al. (2019) developed a multi-state Markov model to predict progression from MCI to AD. They focused on a time axis instead on the age attribute. They also focused on the internal censoring problem in longitudinal data. Further effort focused on other data types such as genetic data, PET scans, DTI, fMRI and other related research questions.

Li et al. (2019) came up with a biomarker using auxiliary data of AD and normal control subjects. In the first step, a projection vector was obtained and then the projection vector was integrated with a self-weight grading to develop the novel biomarker. Finally, the biomarker was developed from multiple morphological features to predict progression from MCI to AD.

A novel EEG-based method was introduced by Mammone et al. (2018) for evaluation of MCI subjects. A dissimilarity matrix was developed using the coupling strength of each pair of the EEG signals; then, hierarchical clustering was applied to the related electrodes. Wavelet coherence (WC) and permutation Jaccard distance (PJD) coupling were introduced. 25 MCI patients were involved in the test; after three months, 4 subjects were observed to have progressed to AD and showed connectivity density reduction. The remaining patients did not manifest such behaviour.

Leandrou et al. (2018) reviewed the methods employed in studying progression from MCI to AD. They identified that the entorhinal cortex provides a better classification and predictor for progression from MCI to AD. Minhas et al. (2018) and Sivakani and Ansari (2020) developed an autoregressive model with 3 arrangements of longitudinal data and performed a test. Then they estimated for the future biomarker and made an SVM classification for predicting progression from MCI to AD. Five-fold cross-validation was done; the AUC values generated were 88.93% and 88.13% for 2 years and 3 years of progression from MCI to AD. The study was done only for 3 years and feature selection and boosting algorithms were planned to apply.

Minhas et al. (2017) introduced a supervised non-parametric method the classification of a longitudinal dataset for progression from MCI to AD. The similarity between the clusters was estimated using the Euclidean space and the feature values were selected using the linear regression method. Leave-one-out cross-validation was done; the accuracy and the precision values predicted were 93.33% and 89.66%. In the future, the missing data computation and validation with other biomarkers were to be focused on.

Li et al. (2015) developed diagnosing models to find the different stages of AD using BLS and convolutional variants. MRI images were collected from the ADNI database and validated using the proposed algorithm. The outcome of this new algorithm was compared with a state-of-the-art algorithm for accuracy and training time; the authors concluded that the new proposed model was better than the compared algorithm.

A comparative analysis has been made to find defects in prediction. Logistic regression, naïve Bayes and decision tree classifiers have been used for prediction of cost-sensitive classification (Moser et al., 2008; Tuan et al., 2022; Weakley et al., 2015). Bari Antor et al. (2021) carried out a comparative result analysis with various machine learning techniques for prediction of dementia disease, and concluded that the support vector machine model is the best for dementia prediction for the OASIS dataset. The machine learning models considered were support vector machine, logistic regression, decision tree and random forest.

Mahyoub et al. (2018) deployed various machine learning models on the ADNI dataset to rank the risk factors of Alzheimer's disease. Bansal et al. (2018) carried out a comparative study for the ADNI dataset using J48, multi-layer perceptron, Naïve Bayes and random forest algorithms. The conclusion was that the J48 algorithm is best for the detection of dementia. Segovia et al. (2012) compared two machine learning approaches to predicting Alzheimer's disease.

It can be observed from the literature that classifications of Alzheimer's disease have been done using various machine learning algorithms and the performance of the models has been analysed using one or two algorithms, but in our model, the model performance is analysed using four algorithms. The proposed model is compared with an existing model (Bari Antor et al., 2021) and shows better performance. This paper aims to develop a framework for classifying the various stages of Alzheimer's disease and to analyse the performance of the model using various evaluation metrics.

This paper focuses on the following tasks. We perform a classification of Alzheimer's disease using the naïve Bayes algorithm, linear regression algorithm, SVM algorithm and random forest algorithm. We use the ADNI database for the classification and carry out a 10-fold cross-validation. The class-wise performance is analysed. We compare the performance of the algorithm and the model building time for the ADNI dataset with the OASIS dataset. The classification accuracy is compared and the best algorithm is analysed using confusion matrix components. Finally, we compare the model performance with an existing model.

2 Proposed Model

The proposed model is shown in Figure 2. It has seven steps. In the first step, the ADNI dataset is considered for processing, then we carry out pre-processing and implement feature selection. Validation is made for the classification and the results are compared to find the best classifier.

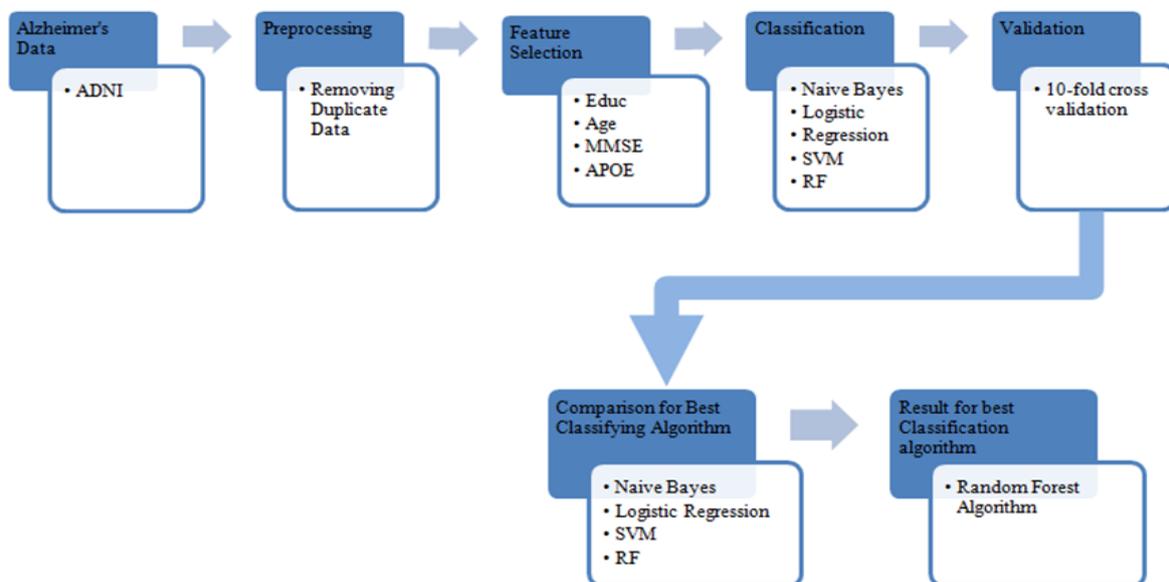


Figure 2. Framework of proposed model.

In this framework, the ADNI data pre-processing and feature selection are made; then the cross-validation is performed on the classified ADNI data. In the pre-processing stage, duplicate data are removed and then subjected to feature selection. After applying feature selection, the attributes selected are MMSE, Education and APOE, and Age. Naïve Bayes, logistic regression, SVM, and random forest (RF) algorithms are used for the classification. Ten-fold cross-validation is applied. The results obtained are compared to find the best algorithm.

2.1 Methodology

The methodology used in this proposed model is discussed below with a detailed description.

Machine learning algorithm

Machine learning algorithms will create the model for the dataset (Bari Antor et al., 2021) based on the problem. In this paper, the following algorithms are used for the classification: naive Bayes algorithm, logistic regression, SVM/SMO and random forest.

Naïve Bayes

The naïve Bayes algorithm classifies the dataset using the Bayes rule. Based on the probability observed in the training data, the classification is made using all the features. It is a supervised learning algorithm (Bari Antor et al., 2021). The classification is made based on the probability Equation (1).

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (1)$$

where $P(A | B)$ is the conditional probability of A given B, $P(B | A)$ is the conditional probability of B given A, $P(A)$ is the probability of event A, and $P(B)$ is the probability of event B.

Logistic regression

The logistic regression algorithm is used for the classification task. This algorithm works based on probability. It is a supervised learning algorithm (Bari Antor et al., 2021). The classification is made based on the hypothesis function Equation (2).

$$0 \leq h_0(x) \leq 1 \quad (2)$$

where $h_0(x)$ is the hypothesis function.

Support vector machines (SVM)

The SVM algorithm is a classification algorithm. It works based on the segregation process. It is a supervised learning algorithm (Bari Antor et al., 2021). For the implementation of the SMV algorithm, sequential minimal optimization (SMO) is used. The classification is made based on Equation (3).

$$\left[\frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i (w^T x_i - b)) \right] + \Lambda \|w\|^2 \quad (3)$$

Random forest

The random forest algorithm is a classification algorithm. Using the ensemble learning concept will classify the dataset. The training process takes place based on the bagging process. It is a supervised learning algorithm (Bari Antor et al., 2021). The classification is made using Equation (4).

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2 \quad (4)$$

where N is the number of instances, f_i is the result returned by the model, and y_i is the actual value for the instances.

3 Results and Discussion

The classification is made on the ADNI preprocessed dataset using the machine learning models. The performance is analysed and a comparison is made with the OASIS dataset.

3.1 Data description

The data are acquired from the ADNI dataset (ADNI, 2017). The attributes given in the dataset are PTID, Age, PTgender, Pteduca, APOE4, MMSE, APOE Genotype and DXBL.

Table 1. Data description for ADNI dataset.

No.	Data attribute	Attribute name	Description
1	PTID	Patient ID	Identification number of the patient
2	Age	Age	Age of the patient
3	PTgender	Patient gender	Gender of the patient
4	Pteduca	Patient education	Education details of the patient
5	APOE4	Apolipoprotein gene type 4	Type 4 gene for the Apolipoprotein, it is high risk in AD
6	MMSE	Mini mental state examination	Mini mental state examination taken for the patient
7	APOEGenotype	Apolipoprotein gene type pairs	Apolipoprotein gene type pair of gene present in the patient
8	DXBI	Diagnosis	Type of AD

The OASIS longitudinal dataset (OASIS, 2021) is taken for the evaluation. The attributes in the dataset are Subject ID, Age, M/F, Hand, EDUC, SES, MMSE, CDR, eTIV, nWBV, ASF and Group.

Table 2. Data description for OASIS dataset.

No.	Data attribute	Attribute name	Description
1	Subject ID	Patient ID	Identification number of the patient
2	Age	Age	Age of the patient
3	M/F	Patient gender	Gender of the patient
4	Hand	Dominant hand	Dominant hand of the patient
5	EDUC	Education level	Education level of the patient
6	SES	Socio-economic status	Socio-economic status of the patient
7	MMSE	Mini mental state examination	Mini mental state examination taken for the patient
8	CDR	Clinical dementia rating	Clinical dementia rating of the patient
9	eTIV	Estimated total intracranial volume	Estimated total intracranial volume of the brain
10	nWBV	Normalized whole brain volume	Normalized whole brain volume of the patient
11	ASF	Atlas scaling factor	Scaling factor of the brain
12	Group	Converted, demented, nondemented	Specifies the stage of dementia

3.2 Data pre-processing

Pre-processing is an essential step for processing a dataset (Sivakani & Ansari, 2020). The pre-processing is done on the ADNI dataset. In this dataset, there are 1534 instances. It is identified that some of the instances are repeated, so the duplicated data are removed using the “remove duplicates” function in the Weka tool; then the instances obtained for classification are 1343. The classification is implemented on

these 1343 instances. In the OASIS dataset, there are 374 instances and all the instances are considered for evaluation.

3.3 Feature selection

Feature selection is a process of selecting the required input instances for processing (Sivakani & Ansari, 2020). The Bestfirst and CfssubsetEval methods are applied and it is found that the PTGENDER and the APOE Genotype attributes are of minimal importance for the classification, so those attributes are removed and the processing is done with the rest of the attributes.

3.4 Classification using naïve Bayes algorithm

The naïve Bayes algorithm is applied to the dataset and the classification results are displayed in Table 3. It shows that the instances classified correctly and incorrectly are 67.68% and 32.31%. The root mean squared error, the root relative squared error and the kappa statistic values are 0.33%, 78.11%, and 0.54%. The model for this algorithm was constructed with a timing of 0.1 seconds. The validation applied for the classification was 10-fold cross-validation.

Table 3. Classification using naïve Bayes algorithm.

No.	Task	Value (%)
1	Correctly classified instances	67.68
2	Incorrectly classified instances	32.31
3	Root mean squared error	0.33
4	Root relative squared error	78.11
5	Kappa statistic	0.54

3.5 Classification using logistic regression algorithm

The logistic regression algorithm is applied to the dataset and the classification results are displayed in Table 4. It shows that the instances classified correctly and incorrectly are 84.58% and 15.41%. The root mean squared error, the root relative squared error and the kappa statistic values are 0.27%, 0.78%, and 64.94%. The model for this algorithm was constructed with a timing of 1.66 seconds.

Table 4. Classification using logistic regression algorithm.

No.	Task	Value (%)
1	Correctly classified instances	84.58
2	Incorrectly classified instances	15.41
3	Root mean squared error	0.27
4	Root relative squared error	64.94
5	Kappa statistic	0.78

3.6 Classification using SVM/SMO algorithm

The SVM algorithm is applied to the dataset and the classification results are displayed in Table 5. It shows that the instances classified correctly and incorrectly are 87.42% and 12.58%. The root mean squared error, the root relative squared error and the kappa statistic values are 0.33%, 78.77%, and 0.54%. The model for this algorithm was constructed with a timing of 3.75 seconds.

Table 5. Classification using SVM algorithm.

No.	Task	Value (%)
1	Correctly classified instances	87.42
2	Incorrectly classified instances	12.58
3	Root mean squared error	0.33
4	Root relative squared error	78.55
5	Kappa statistic	0.54

3.7 Classification using random forest algorithm

The random forest algorithm is applied to the dataset and the classification results are displayed in Table 6. It shows that the instances classified correctly and incorrectly are 88.90% and 11.09%. The root mean squared error, the root relative squared error and the kappa statistic values are 0.25%, 59.34%, and 0.84%. The model for this algorithm was constructed with a timing of 0.92 seconds.

Table 6. Classification using random forest algorithm.

No.	Task	Value (%)
1	Correctly classified instances	88.90
2	Incorrectly classified instances	11.09
3	Root mean squared error	0.25
4	Root relative squared error	59.34
5	Kappa statistic	0.84

Table 7. Comparison of algorithms.

No.	Algorithm	Classification
1	Naïve Bayes	67.68%
2	Logistic regression	84.58%
3	Support vector machine	87.42%
4	Random forest	88.90%

Classification accuracy of algorithms

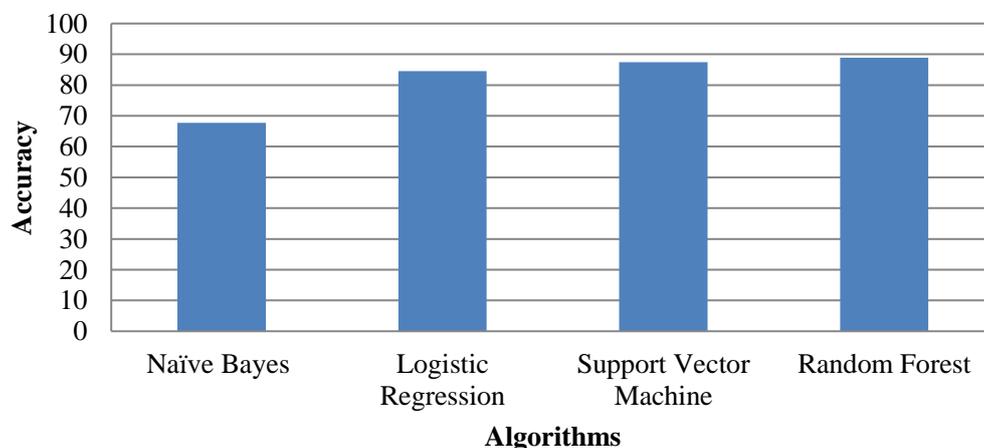


Figure 3. Comparison of algorithms.

Table 7 shows a comparison of the algorithms. The random forest algorithm gives the best results. Figure 3 shows a comparison of the classification accuracy. The graphical representation shows that the random forest algorithm has the highest accuracy.

3.8 Class-wise performance measurement using confusion matrix

The confusion matrix is one of the evaluation metrics to measure the performance of an algorithm (Chicco et al., 2020). The performance can be evaluated for each class. The components of the confusion matrix are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Correctly predicted positive instances are referred to as TP, correctly predicted negative instances are labelled TN; wrongly predicted positive instances are FP, and wrongly predicted negative instances are FN (Chicco et al., 2020; Chicco et al., 2021; Brown, 2018). The best classification will give the TP and TN values as greater than FP and FN. In this ADNI dataset, there are four types of classes: CN, LMCI, EMCI and AD. For each class, the TP, TN, FP and FN are calculated.

Table 8. Confusion matrix generated by naïve Bayes algorithm.

No.	CN	LMCI	AD	EMCI	Predicted best value
1	333	40	0	10	Positive
2	110	334	31	26	Positive
3	3	80	151	0	Positive
4	83	51	0	91	Positive

Table 7 shows the confusion matrix generated for the naïve Bayes algorithm. The positive and the negative instances predicted are 909 and 434. The positive prediction has a greater value than the negative prediction. The coloured part represents the positively predicted values and the rest are the negatively predicted values.

When evaluating the confusion matrix for each class, the CN class has TP, FP, TN and FN values of 333, 50, 764, and 196. For the LMCI class, the TP, FP, TN and FN are 334, 167, 671, and 171. For the AD class, the TP, FP, TN and FN are 151, 83, 1078, and 31. For the EMCI class, the TP, FP, TN and FN are 91, 134, 1082, and 36. The TN value shows the highest score for all the classes.

Also, the highest score for each component is analysed; the TP value is greater for the LMCI class and lesser for EMCI; the FP value is greater for the LCMI class and lesser CN; the TN value is greater for the EMCI class and lesser for LMCI; and the FN value is greater for the CN class and lesser for the AD class.

Table 9. Confusion matrix for logistic regression algorithm.

No.	CN	LMCI	AD	EMCI	Predicted best value
1	264	29	4	86	Positive
2	0	457	8	36	Positive
3	0	25	207	2	Positive
4	0	15	2	208	Positive

Table 9 shows the confusion matrix generated for logistic regression. The positive and negative instances predicted are 1136 and 207. The positive prediction has a greater value than the negative prediction. The coloured part represents the positively predicted values and the rest are the negatively predicted values.

For the CN class, the TP, FP, TN and FN values are 264, 264, 964, and 0. For the LMCI class, the TP, FP, TN and FN are 457, 44, 773, and 69. For the AD class, the TP, FP, TN and FN values are 207, 27, 1095, and

14. For the EMCI class, the TP, FP, TN and FN are 208, 17, 994, and 124. The TN value shows the highest score for all the classes.

Also, the highest score for the metrics is analysed; the TP value is greater for the LMCI class and lesser for AD; the FP value is greater for the CN class and lesser for EMCI; the TN value is greater for the AD class and lesser for LMCI; and the FN value is greater for the EMCI class and lesser for the CN class.

Table 10. Confusion matrix for SVM algorithm.

No.	CN	LMCI	AD	EMCI	Predicted best value
1	347	31	0	5	Positive
2	32	461	5	3	Positive
3	1	39	192	2	Positive
4	34	17	0	174	Positive

Table 10 shows the confusion matrix generated for the SVM algorithm. The positive and negative instances predicted are 1174 and 169. The positive prediction has a greater value than the negative prediction. The coloured part represents the positively predicted values and the rest are the negatively predicted values.

For the CN class, the TP, FP, TN and FN values are 347, 36, 893, and 67. For the LMCI class, the TP, FP, TN and FN are 461, 40, 755, and 87. For the AD class, the TP, FP, TN and FN values are 192, 42, 1104, and 5. For the EMCI class, the TP, FP, TN and FN are 174, 51, 1108, and 10. The TN value shows the highest score for all the classes.

Also, the highest score for the metrics is analysed; the TP value is greater for the LMCI class and lesser for AD; the FP value is greater for the class EMCI and lesser for CN; the TN value is greater for the AD class and lesser for LMCI; and the FN value is greater for the LMCI class and lesser for the AD class.

Table 11. Confusion matrix for random forest algorithm.

No.	CN	LMCI	AD	EMCI	Predicted best value
1	354	29	0	0	Positive
2	24	477	0	0	Positive
3	0	39	195	0	Positive
4	29	28	0	168	Positive

Table 11 shows the confusion matrix generated for the random forest algorithm. The positive and negative instances predicted are 1194 and 149. The positive prediction has a greater value than the negative prediction. The coloured part represents the positively predicted values and the rest are the negatively predicted values.

For the CN class, the TP, FP, TN and FN values are 354, 29, 907, and 53. For the LMCI class, the TP, FP, TN and FN are 477, 24, 747, and 96. For the AD class, the TP, FP, TN and FN values are 195, 39, 1109, and 0. For the EMCI class, the TP, FP, TN and FN are 168, 57, 1118, and 0. The TN value shows the highest score for all the classes.

Also, the highest score for the metrics is analysed; the TP value is greater for the LMCI class and lesser for EMCI; the FP value is greater for the EMCI class and lesser for LMCI; the TN value is greater for the EMCI class and lesser for LMCI; and the FN value is greater for the LMCI class and lesser for AD and EMCI.

Table 12. Comparison of metrics for algorithms – ADNI dataset.

No.	Algorithm	Pos	Neg	ROC	TPR	FPR	Prec	Rec	F1
1	Naïve Bayes	909	434	0.87	0.67	0.14	0.76	0.69	0.67
2	Logistic regression	1136	207	0.95	0.84	0.05	0.89	0.87	0.84
3	SVM	1174	169	0.93	0.87	0.06	0.82	0.88	0.87
4	Random forest	1194	149	0.98	0.88	0.05	0.97	0.88	0.88

Table 12 shows a comparison of the metrics to evaluate the best classification algorithm for the ADNI dataset. The evaluation metrics considered for the evaluation are the positive(Pos), negative(Neg), receiver operative characteristics(ROC), true positive rate(TPR), false positive rate(FPR), precision(Prec), recall(Rec) and F1score(F1). When comparing all the metrics for these algorithms, the random forest algorithm gives the best results.

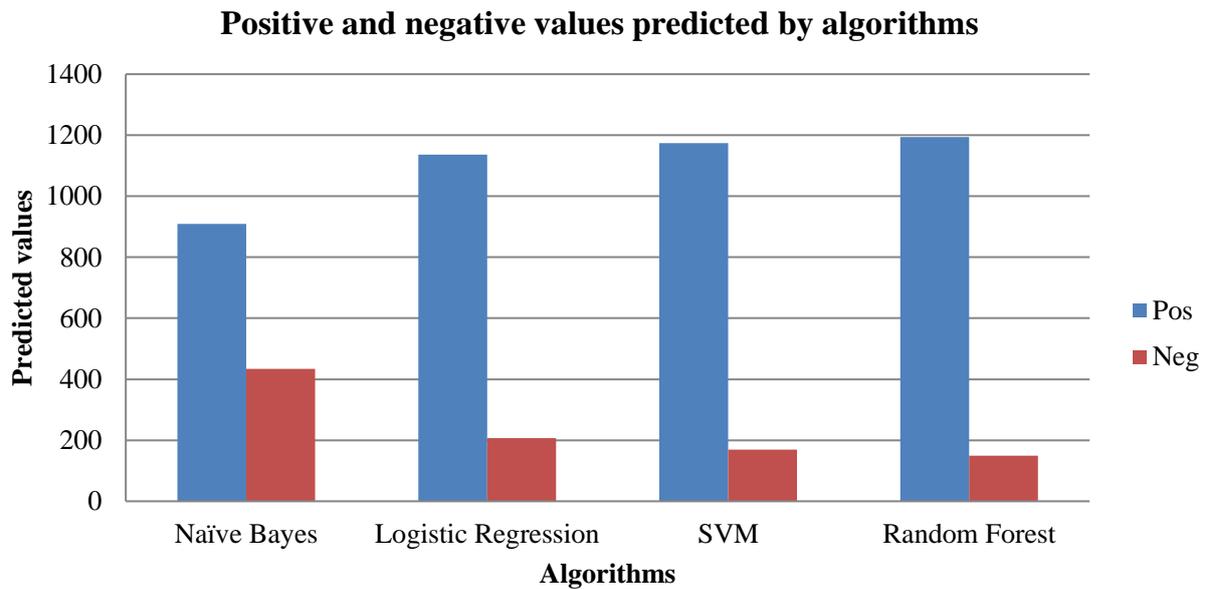


Figure 4. Positive and negative classification values for algorithms.

Figure 4 shows the positive and negative classification values; the positive value should be high for the best algorithm (Hicks et al., 2022). To show the efficiency of the proposed model, the OASIS dataset (OASIS, 2021) is applied to the model and the evaluation metrics are tabulated in Table 12, which shows a comparison of the metrics to evaluate the best classification algorithm for the OASIS dataset. When comparing all the metrics for these algorithms, the random forest algorithm gives the best results.

Table 13. Comparison of metrics for algorithms – OASIS dataset.

No.	Algorithm	ROC	TPR	FPR	Prec	Rec	F1
1	Naïve Bayes	0.98	0.92	0.46	0.92	0.92	0.92
2	Logistic regression	0.99	0.98	0.15	0.98	0.98	0.98
3	SVM	0.99	0.98	0.12	0.98	0.98	0.97
4	Random forest	1.00	0.96	0.34	0.96	0.95	0.93

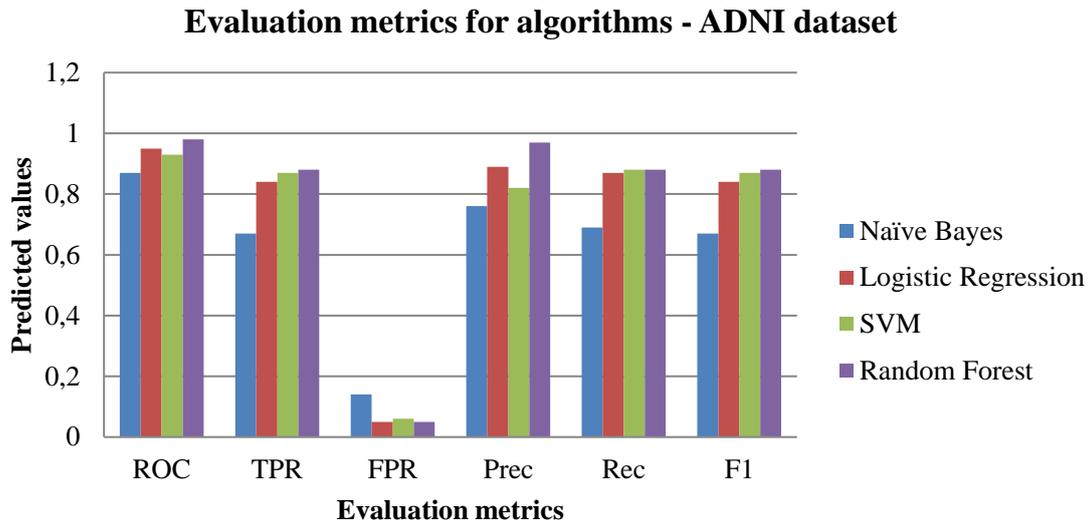


Figure 5. Evaluation metrics for ADNI dataset.

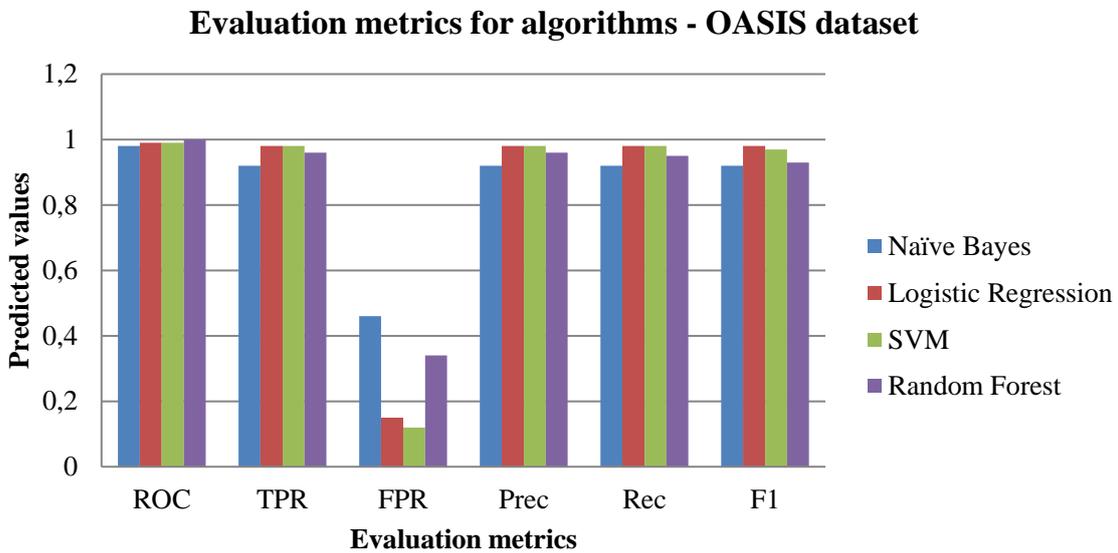


Figure 6. Evaluation metrics for OASIS dataset.

Figure 5 shows the evaluation metrics of the algorithms for the ADNI dataset. All the evaluation metrics have the highest values for the random forest algorithm. Figure 6 shows the evaluation metrics of the algorithms for the OASIS dataset. Again, all the evaluation metrics have the highest values for the random forest algorithm. For both datasets, the random forest algorithm gives the best results. The accuracy of both the datasets for each algorithm are tabulated in Table 13; the random forest algorithm shows the best results for both datasets.

Table 14. Accuracy comparison for models – both datasets.

No.	Algorithm	ADNI	OASIS
1	Naïve Bayes	0.87	0.98
2	Logistic regression	0.95	0.99
3	SVM	0.93	0.99
4	Random forest	0.98	1.00

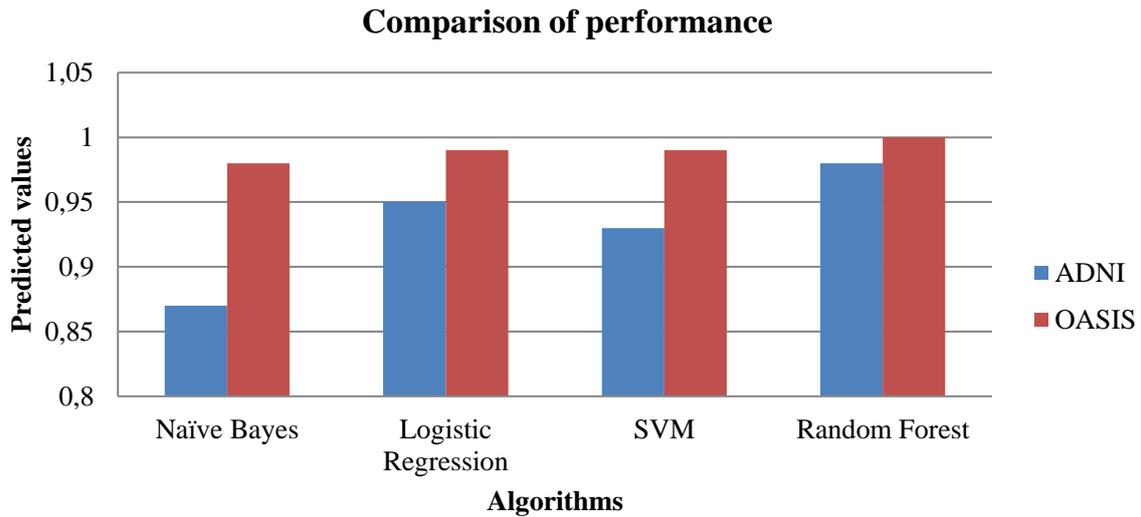


Figure 7. Comparison of accuracy for ADNI and OASIS dataset.

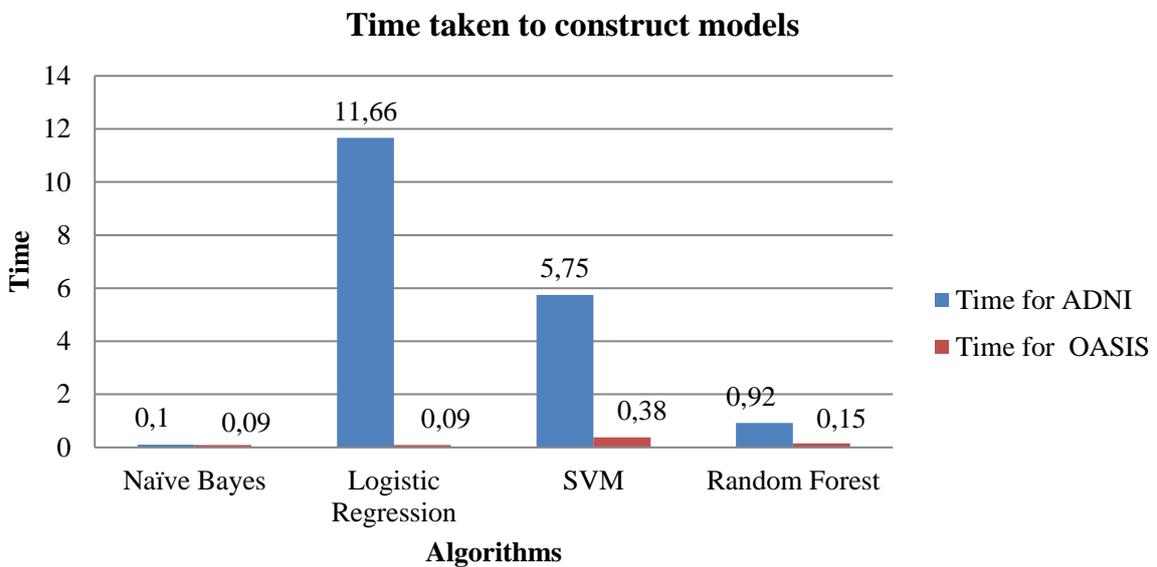


Figure 8. Comparison of time for constructing models for ADNI and OASIS dataset.

Figure 7 shows the accuracy of the algorithms for the ADNI and OASIS datasets. The random forest algorithm shows the best result among all the algorithms. Figure 8 shows the model construction time of the algorithms for both the ADNI and OASIS datasets. For both datasets, the naïve Bayes algorithm gave the best result. The time (in seconds) taken to construct the model for both datasets for each algorithm is tabulated in Table 15; the naïve Bayes algorithm shows the best results for both datasets.

Table 15. Comparison of model construction time.

No.	Algorithm	Time for ADNI	Time for OASIS
1	Naïve Bayes	0.1	0.09
2	Logistic regression	11.66	0.09
3	SVM	5.75	0.38
4	Random forest	0.92	0.15

In the classification, the random forest algorithm has the best performance and when comparing all the evaluation metrics, the random forest algorithm again gives the best result for both datasets. While comparing the time for constructing the model, the naïve Bayes algorithm gives the best results. Considering all the evaluation metrics, the conclusion is that among these four algorithms, the random forest algorithm gives the pre-eminent results.

4 Conclusions

In this paper, we made a classification on the ADNI dataset using four machine learning algorithms. Naïve Bayes, logistic regression, SVM and random forest. The dataset had a multi-class label, so multi-class classification was performed and the classes were CN, EMCI, LMCI and AD. The validation made on the dataset was 10-fold cross-validation. Then the evaluation metrics were compared and it was found that the random forest algorithm gives better results. The evaluation metrics considered were positive (Pos), negative (Neg), receiver operative characteristics (ROC), true positive rate (TPR), false positive rate (FPR), precision (Prec), recall (Rec) and F1score (F1). The correctly classified instances for naïve Bayes, logistic regression, SVM and random forest were 67.68%, 84.58%, 87.42%, and 88.90%. From the correctly classified instances, it was found that the random forest algorithm gives the best results.

A confusion matrix was generated and the positive and negative classified values were calculated for the algorithms and also for each class. The evaluation metrics were also compared and it was found that the random forest algorithm gives the best results. The OASIS dataset was applied to the proposed model with the same algorithms and the accuracy rates were 98%, 99%, 99% and 100%. The best results were again achieved by the random forest algorithm. The model construction time was compared for both datasets and the naïve Bayes algorithm gave the best results for this. In the future, we plan to perform classification using other algorithms.

Additional Information and Declarations

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: S.R.: Conceptualization, Data curation, Software, Methodology, Validation, Writing – review & editing. M.M.S.M.: Supervision, Writing – original draft.

Institutional Review Board Statement: Not applicable, because the authors use external datasets that already exist.

Data Availability: The data supporting this study's findings are openly available by the Alzheimer's Disease Neuroimaging Initiative (ADNI) at <https://adni.loni.usc.edu/data-samples/access-data/> and by the Brains project OASIS at <http://www.oasis-brains.org/#data>.

References

- ADNI. (2017). *Alzheimer's Disease Neuroimaging Initiative*. <https://adni.loni.usc.edu/data-samples/access-data/>
- Bansal, D., Chhikara, R., Khanna, K., & Gupta, P. (2018). Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia. *Procedia Computer Science*, 132, 1497–1502. <https://doi.org/10.1016/j.procs.2018.05.102>
- Bari Antor, M., Jamil, A. H. M. S., Mamtaz, M., Monirujjaman Khan, M., Aljahdali, S., Kaur, M., Singh, P., & Masud, M. (2021). A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. *Journal of Healthcare Engineering*, 2021, e9917919. <https://doi.org/10.1155/2021/9917919>
- Brown, J. B. (2018). Classifiers and their Metrics Quantified. *Molecular Informatics*, 37(1-2), 1700127. <https://doi.org/10.1002/minf.201700127>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1). <https://doi.org/10.1186/s12864-019-6413-7>
- Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, 14(1). <https://doi.org/10.1186/s13040-021-00244-z>

- Eitel, F., & Ritter, K. (2019). Testing the robustness of attribution methods for convolutional neural networks in MRI-based Alzheimer's disease classification. *arXiv:1909.08856*. <https://doi.org/10.48550/arXiv.1909.08856>
- Gao, F., Yoon, H., Xu, Y., Goradia, D., Luo, J., Wu, T., & Su, Y. (2020). AD-NET: Age-adjust neural network for improved MCI to AD conversion prediction. *NeuroImage: Clinical*, 27, 102290. <https://doi.org/10.1016/j.nicl.2020.102290>
- Guo, M., Li, Y., Zheng, W., Huang, K., Zhou, L., Hu, X., Yao, Z., & Hu, B. (2020). A novel conversion prediction method of MCI to AD based on longitudinal dynamic morphological features using ADNI structural MRIs. *Journal of Neurology*, 267(10), 2983–2997. <https://doi.org/10.1007/s00415-020-09890-5>
- Gupta, Y., Lama, R. K., & Kwon, G.-R. (2019). Prediction and Classification of Alzheimer's Disease Based on Combined Features From Apolipoprotein-E Genotype, Cerebrospinal Fluid, MR, and FDG-PET Imaging Biomarkers. *Frontiers in Computational Neuroscience*, 13. <https://doi.org/10.3389/fncom.2019.00072>
- Han, R., Chen, C. L. P., & Liu, Z. (2020). A Novel Convolutional Variation of Broad Learning System for Alzheimer's Disease Diagnosis by Using MRI Images. *IEEE Access*, 8, 214646–214657. <https://doi.org/10.1109/access.2020.3040340>
- Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, 12(1), 5979. <https://doi.org/10.1038/s41598-022-09954-8>
- Leandrou, S., Petroudi, S., Kyriacou, P. A., Reyes-Aldasoro, C. C., & Pattichis, C. S. (2018). Quantitative MRI Brain Studies in Mild Cognitive Impairment and Alzheimer's Disease: A Methodological Review. *IEEE Reviews in Biomedical Engineering*, 11, 97–111. <https://doi.org/10.1109/rbme.2018.2796598>
- Li, F., Tran, L., Thung, K.-H., Ji, S., Shen, D., & Li, J. (2015). A Robust Deep Model for Improved Classification of AD/MCI Patients. *IEEE Journal of Biomedical and Health Informatics*, 19(5), 1610–1616. <https://doi.org/10.1109/jbhi.2015.2429556>
- Li, Q., Wu, X., Xu, L., Chen, K., & Yao, L. (2018). Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Cognitively Unimpaired Individuals Using Multi-feature Kernel Discriminant Dictionary Learning. *Frontiers in Computational Neuroscience*, 11. <https://doi.org/10.3389/fncom.2017.00117>
- Li, Y., Fang, Y., Zhang, H., & Hu, B. (2019). Self-Weighting Grading Biomarker Based on Graph-Guided Information Propagation for the Prediction of Mild Cognitive Impairment Conversion. *IEEE Access*, 7, 116632–116642. <https://doi.org/10.1109/access.2019.2936415>
- Mahyoub, M., Randles, M., Baker, T., & Yang, P. (2018). Comparison Analysis of Machine Learning Algorithms to Rank Alzheimer's Disease Risk Factors by Importance. In *2018 11th International Conference on Developments in eSystems Engineering (DeSE)*, (pp. 1–11). IEEE. <https://doi.org/10.1109/DeSE.2018.00008>
- Maqsood, M., Nazir, F., Khan, U., Aadil, F., Jamal, H., Mehmood, I., & Song, O. (2019). Transfer Learning Assisted Classification and Detection of Alzheimer's Disease Stages Using 3D MRI Scans. *Sensors*, 19(11), 2645. <https://doi.org/10.3390/s19112645>
- Mammone, N., Ieracitano, C., Adeli, H., Bramanti, A., & Morabito, F. C. (2018). Permutation Jaccard Distance-Based Hierarchical Clustering to Estimate EEG Network Density Modifications in MCI Subjects. *IEEE Transactions on Neural Networks and Learning Systems*, 29(10), 5122–5135. <https://doi.org/10.1109/tnnls.2018.2791644>
- Martinez-Murcia, F. J., Ortiz, A., Gorriz, J.-M., Ramirez, J., & Castillo-Barnes, D. (2020). Studying the Manifold Structure of Alzheimer's Disease: A Deep Learning Approach Using Convolutional Autoencoders. *IEEE Journal of Biomedical and Health Informatics*, 24(1), 17–26. <https://doi.org/10.1109/jbhi.2019.2914970>
- MCI Screen. (2021). *Medical Care Corporation*. <https://www.mccare.com/education/alzprogression.html>.
- Minhas, S., Khanum, A., Riaz, F., Khan, S. A., & Alvi, A. (2018). Predicting Progression From Mild Cognitive Impairment to Alzheimer's Disease Using Autoregressive Modelling of Longitudinal and Multimodal Biomarkers. *IEEE Journal of Biomedical and Health Informatics*, 22(3), 818–825. <https://doi.org/10.1109/jbhi.2017.2703918>
- Minhas, S., Khanum, A., Riaz, F., Alvi, A., & Khan, S. A. (2017). A Nonparametric Approach for Mild Cognitive Impairment to AD Conversion Prediction: Results on Longitudinal Data. *IEEE Journal of Biomedical and Health Informatics*, 21(5), 1403–1410. <https://doi.org/10.1109/jbhi.2016.2608998>
- Moser, R., Pedrycz, W., & Succi, G. (2008). A comparative analysis of the efficiency of change metrics and static code attributes for defect prediction. In *Proceedings of the 30th international conference on Software engineering*, (pp. 181–190). ACM. <https://doi.org/10.1145/1368088.1368114>
- National Institute on Aging. (2021, July 08). *NIH National Institute on Aging*. <https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet#stages>
- Nozadi, S. H., & Kadoury, S. (2018). Classification of Alzheimer's and MCI Patients from Semantically Parcelled PET Images: A Comparison between AV45 and FDG-PET. *International Journal of Biomedical Imaging*, 2018, Article ID 1247430. <https://doi.org/10.1155/2018/1247430>
- OASIS. (2021). *The Alzheimer's Association: Brains project OASIS*. <http://www.oasis-brains.org/#data>
- Oh, K., Chung, Y.-C., Kim, K. W., Kim, W.-S., & Oh, I.-S. (2019). Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-54548-6>
- Segovia, F., Górriz, J. M., Ramírez, J., Salas-Gonzalez, D., Álvarez, I., López, M., & Chaves, R. (2012). A comparative study of feature extraction methods for the diagnosis of Alzheimer's disease using the ADNI database. *Neurocomputing*, 75(1), 64–71. <https://doi.org/10.1016/j.neucom.2011.03.050>

- Sivakani, R., & Ansari, G.A.** (2020). Machine Learning Framework for Implementing Alzheimer's Disease. In *2020 International Conference on Communication and Signal Processing (ICCSP)*, (pp. 588–592). IEEE.
<https://doi.org/10.1109/ICCSP48568.2020.9182220>
- Syed, A. H., Khan, T., Hassan, A., Alromema, N. A., Binsawad, M., & Alsayed, A. O.** (2020). An Ensemble-Learning Based Application to Predict the Earlier Stages of Alzheimer's Disease (AD). *IEEE Access*, 8, 222126–222143.
<https://doi.org/10.1109/access.2020.3043715>
- Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., & Initiative, A. D. N.** (2017). A Novel Grading Biomarker for the Prediction of Conversion From Mild Cognitive Impairment to Alzheimer's Disease. *IEEE Transactions on Biomedical Engineering*, 64(1), 155–165. <https://doi.org/10.1109/tbme.2016.2549363>
- Tuan, T. A., Pham, T. B., Kim, J. Y., & Tavares, J. M. R. S.** (2022). Alzheimer's diagnosis using deep learning in segmenting and classifying 3D brain MR images. *International Journal of Neuroscience*, 132(7), 689–698.
<https://doi.org/10.1080/00207454.2020.1835900>
- Weakley, A., Williams, J. A., Schmitter-Edgecombe, M., & Cook, D. J.** (2015). Neuropsychological test selection for cognitive impairment classification: A machine learning approach. *Journal of Clinical and Experimental Neuropsychology*, 37(9), 899–916. <https://doi.org/10.1080/13803395.2015.1067290>
- You, Z., Zeng, R., Lan, X., Ren, H., You, Z., Shi, X., Zhao, S., Guo, Y., Jiang, X., & Hu, X.** (2020). Alzheimer's Disease Classification With a Cascade Neural Network. *Frontiers in Public Health*, 8. <https://doi.org/10.3389/fpubh.2020.584387>
- Zhang, L., Lim, C. Y., Maiti, T., Li, Y., Choi, J., Bozoki, A., & Zhu, D. C.** (2018). Analysis of conversion of Alzheimer's disease using a multi-state Markov model. *Statistical Methods in Medical Research*, 28(9), 2801–2819.
<https://doi.org/10.1177/0962280218786525>

Editorial record: The article has been peer-reviewed. First submission received on 14 June 2022. Revisions received on 1 August 2022, 5 October 2022 and 30 October 2022. Accepted for publication on 2 November 2022. The editors coordinating the peer-review of this manuscript were Mazin Abed Mohammed , Seifedine Kadry , and Oana Geman . The editor in charge of approving this manuscript for publication was Zdenek Smutny .

Special Issue: Deep Learning Blockchain-enabled Technology for Improved Healthcare Industrial Systems.

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951
