Dementia Prediction Using Machine Learning

Dhakal, Sara; Azam, Sami; Hasib, Khan Md; Karim, Asif; Jonkman, Mirjam; Farhan Al Haque, A. S.M.

Published in:
Procedia Computer Science

DOI:
10.1016/j.procs.2023.01.414

Published: 01/01/2023

Document Version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 219 (2023) 1297–1308

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2022

Dementia Prediction Using Machine Learning

Sara Dhakal[a], Sami Azam[a], Khan Md. Hasib[b], Asif Karim[a]*, Mirjam Jonkman[a], A S M Farhan Al Haque[c]

[a]*College of Engineering, IT and Environment, Charles Darwin University, NT, Australia*

[b]*Department of Computer Science and Engineering, Ahsanullah University of Science Technology, Dhaka, Bangladesh*

[c]*Department of Electronic Systems, Aalborg University, Copenhagen, Denmark*

Abstract

Dementia is a chronic and degenerative condition, which has become a major health concern among the elderly. With ever-continuing cases of dementia, it has become a very challenging task in the 21st century to provide care for patients with dementia. This paper proposes a framework for the prediction of dementia using the data collected from the OASIS (Open Access Series of Imaging Studies) project which was made available by the Washington University Alzheimer's Disease Research Centre. Different techniques have been implemented for data imputation, pre-processing and data transformation to create suitable data for training the model. Machine learning approaches like Adaboost (AB), Decision Tree (DT), Extra Tree (ET), Gradient Boost (GB), K-Nearest Neighbour (KNN), Logistic Regression (LR), Naïve Bayes (NB), Random Forest (RF), and SVM (Support Vector Machine) has been used for a combination of features. These techniques have been applied to the full set of features and features selected from Least Absolute Shrinkage and Selection Operator (LASSO) techniques. A comparison between the accuracy, precision, and other metrics based on the results of the classification algorithms has been provided. The experimental results show that the highest accuracy of 96.77% was obtained by Support Vector Machine (SVM) with full features. The proposed methodology is promising and if developed and deployed can be helpful for the rapid assessment of Alzheimer's Disease (AD).

Keywords: *Alzheimer's Dementia, Health Informatics, LASSO, Machine Learning;*

* Corresponding author. *E-mail address:* asif.karim@cdu.edu.au

1. Introduction

Dementia is a progressive and irreversible cognitive deterioration. There are about 40 million people

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: asif.karim@cdu.edu.au

10.1016/j.procs.2023.01.414

suffering from dementia in this world and thus it is recognized as one of the major health challenges of our century [1]. The two major causes of dementia are Alzheimer's Disease and dementia caused by vascular factors. Alzheimer's disease (AD) is one of the most common causes of dementia. In today's world it is responsible for 70-80% of cases affecting 50 million people [2]. The symptoms of dementia are associated with aging; however, some symptoms form at an early age. Alzheimer's disease causes changes in the brain affecting the structural and functional aspects. Machine learning algorithms seem to be effectively used in various industries such as healthcare, banking, transport, social media, education, etc. In this research, we aim to develop algorithms that can learn and progress over time and can be used for predictions. Introduction to various machine learning algorithms and the study revolving around dementia prediction has been introduced by many researchers in their study. However, no tests such as a proper standardized dementia test exist so far [3]. Though, it is very challenging to determine whether the diagnosis will lead towards dementia or people will return to healthy cognitive form. With application of ML, it is possible to identify people with cognitive complaints who are on verge of getting dementia or suffer from Mild Cognitive Impairment. The diagnosis at mild cognitive impairment (MCI) stage needs to be handled carefully as it has been discovered that 10-15% of MCI patients are converted to AD [4]. The use of machine learning techniques in health informatics can easily solve the problem and is very cost-effective. With an aim to create an algorithm, implement the model in the existing hospital systems globally, this proposed model has been accurately developed and made efficient.

2. Background

Alzheimer's Dementia prediction using machine-learning systems has been described in different research studies. Extensive research has been done centered around supervised learning approaches for resolving issues in diverse areas [5, 6] and supervised learning methods [7, 8-11] for detection of many diseases by the support of many computer-aided systems. Annette and her team [12] for their paper have compared the machine learning methods by identifying the biomarkers for early detection. Their paper used 10 machine learning algorithms on the ADNI & Sydney Memory and Age Study (MAS) dataset. MAS showed the accuracy of 0.82 for the performance values, whereas ADNI showed 0.93. A variation of dementia detection methods using clinic collectable predictors has been highlighted by Anastasia et al. [13]. As a part of an ongoing study, they have taken 78 Parkinson's disease dementia (PDD) and 62 Dementia with Lewy bodies (DLB) subjects. Anastasia's team has explored logistic regression, K-Nearest Neighbors (K-NNs), Support Vector Machine (SVM), Naive Bayes Classifier, and ensemble model, for their ability to predict PDD or DLB. The paper by the team of Ahmad et al. [14] was intended to diagnose and classify Alzheimer's Disease. Convolutional Neural Network has been implemented using MRI images from the ADNI dataset. A model was built using a dataset of 1512 mild, 2633 normal, and 2480 AD. Achieving accuracy of 99%, the model has been compared with pre-existing models designed using the OASIS dataset. Gloria et al. [24] have optimized the diagnosis between Alzheimer disease (AD) and vascular dementia (VD) differentiation by combining machine learning (ML) with magnetic resonance imaging (MRI). Their study was categorized into two levels, firstly, identifying if machine learning algorithms combined with MRI features could be supportive towards VD and AD classification. Secondly, dominant features prevalent with "VD-AD dementia" patients were fed as input for the Machine Learning algorithms.

This study was very successful to predict dementia symptoms in an early stage helping in an early prognosis to support physicians' diagnostic evaluations and patients altogether. Compared to earlier pieces of work, this paper entails a thorough description of supervised machine learning algorithms. Based on the literature review, a simple observation has been noted that very few researchers have used more than four or five machine learning algorithms in their research. For this research, we aim to implement nine of the supervised machine learning techniques namely: AdaBoost, Decision Tree, Extra Tree, Gradient Boost, KNN, Logistic Regression, Naïve Bayes, Random Forest, and

Support Vector Machine (SVM). The detection performance of machine learning on different diseases has been verified using diverse approaches and datasets.

3. Aim of Research

The main aim of the research is to develop a prediction model of dementia based on the longitudinal data on OASIS dataset [16]. A model has been accurately developed and made efficient taking the following approaches for this study:

1. Analysis of dataset, and the features within the dataset for prediction of dementia.
2. Explanatory data analysis on the OASIS dataset.
3. Perform various Feature Selection techniques.
4. Improve the efficiency of the proposed model using feature selection and cross-validation techniques.
5. Propose the best model for implementation which gives the best accuracy.

This paper intends to develop a novel algorithm for best prediction results progressively making the model computationally efficient. Several machine learning algorithms have been successfully applied to differentiate AD patients from elderly (healthy) subjects. To differentiate between elderly/ healthy or dementia state person, the model requires a training set of the population having proper categorisation. To make that happen, it is a must the classifier predicts the correct classification of an unseen data.

4. Approach

4.1 Overview of the Proposed Model

Dataset is chosen from the OASIS project. A framework devised is shown in Fig. 1. The model involves the following steps:

- Firstly, the data was collected. Then preparation and pre-processing of data were done.
- Exploratory analysis of data was done to determine the hidden relationships and patterns in the dataset. Data pre-processing also involves the imputation of null, label encoding, data transformation, feature selection, and feature scaling.
- Different variations of feature selection techniques have been implemented for the purpose of learning and building model. This will be further described in the methodology in Section III.
- After feature selection, the splitting of data into train and test sets was done on the ratio of 3:1. We assign 70% of the dataset for training purposes and 30% for testing purposes.
- Data splitting was followed by the process of selecting inputs to predict outcomes. These were fed into the classifiers to learn.
- A qualified model was formed, which was loaded with testing dataset to categorise it into demented and non-demented. Supervised classifiers have been used in our model.

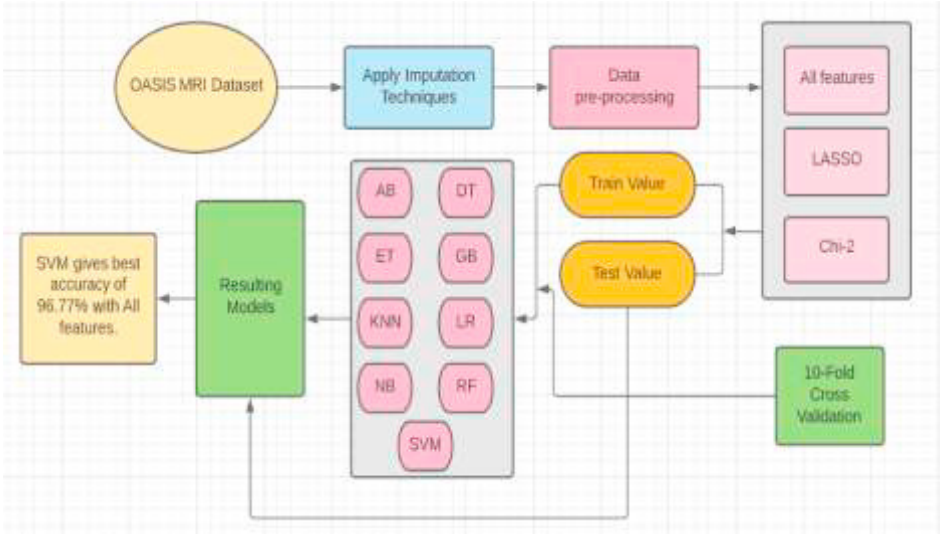

Figure **1**: Graphical representation of our proposed model.

4.2 Supervised Machine Learning

Machine learning works on the basis of past examples and past experiences [17]. There are three categories of machine learning namely, supervised learning, unsupervised learning, and reinforcement learning. This research paper revolves around the use of supervised techniques of machine learning. The use of various algorithms in different feature selection scenarios has been implemented. Different kinds of supervised machine learning algorithms were employed in this study.

5. Methods and Methodologies

5.1 Dataset

The dataset that has been used in this paper was provided by the Open Access Series of Imaging Studies (OASIS) project which aims at making MRI data sets of the brain freely available for future discoveries. Among the two types of datasets provided: Cross-sectional MRI data in young, middle aged, nondemented and demented older adults and Longitudinal MRI data in nondemented and demented older adults, we have chosen the latter one for our prediction [16]. Longitudinal MRI data has the following attributes:

- This dataset consists of a longitudinal collection of 150 subjects aged 60 to 96.
- Each subject was scanned on two or more visits.
- There are both men and women and everyone is right-handed.

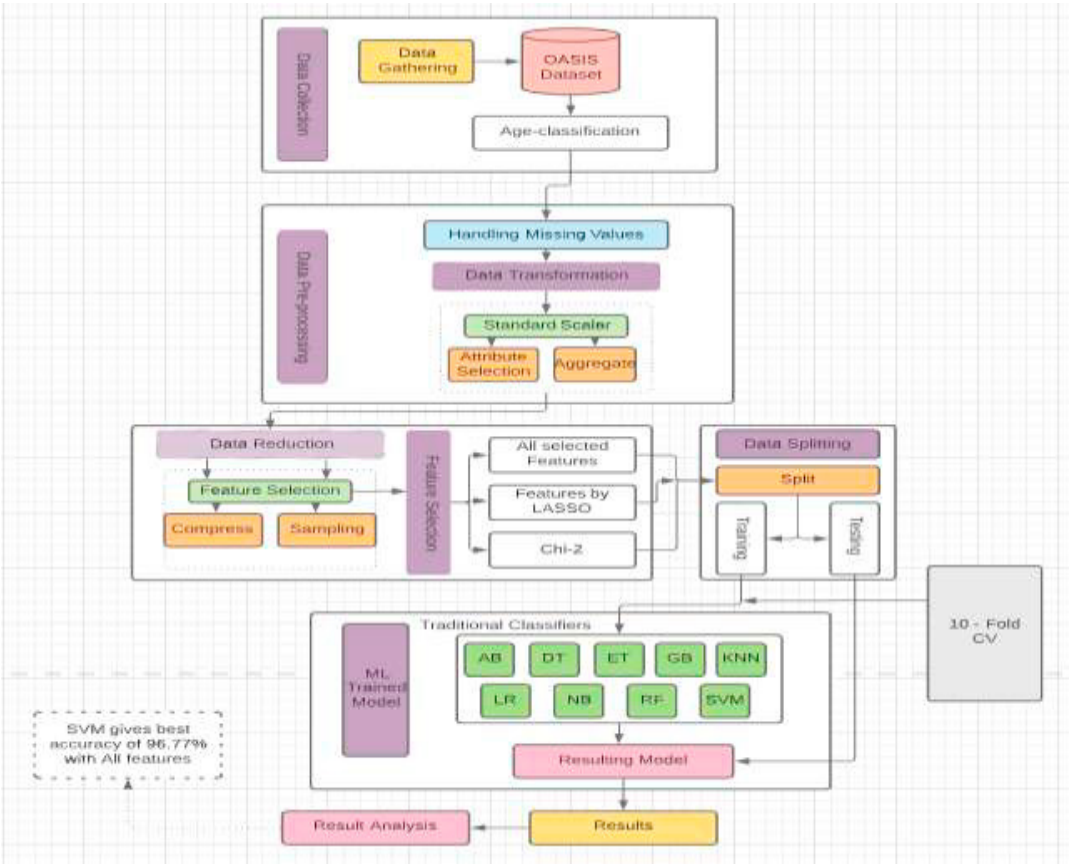

Figure **2**: Architecture of the Proposed Model.

- 72 subjects were characterised as 'nondemented'.
- 64 subjects were characterised as 'demented'.
- 14 subjects were grouped as 'Nondemented' at the time of their visit, who later were characterised as 'Demented' on their following visits. These fall under the 'Converted' category.

5.2 Discussion of the Architecture

This research focused mainly on implementing the steps as proposed by the architecture and creating a working model which can be applied in a real-time setting. After careful research and literature review, an architecture of the proposed model has been formed. Fig, 2 illustrated the architecture of the proposed model. Comprising of major four stages, it is very necessary to implement and understand the working of all the sub-stages.

The first stage comprised of data preparation, which involved gathering the MRI data, and preparing it as an input for the following stage. This stage involved the gathering of data, which is the most important stage. The second stage comprised data-pre-processing. Data pre-processing consists of data visualization, data imputation, and data transformation. This approach handles missing data, grouping the "age" feature, removing existing outliers, normalization, standardization, and feature selection. Data visualization helps us see the raw data based on different charts, histograms, distribution, etc. This stage has a major role in affecting the accuracy at the last stage because the output from the second stage forms as the input to the third stage. The third stage focuses on data segregation, which

is basically splitting the dataset into train data and test data. I also have performed data segregation. The split data was then used in the next stage of the model which is model building. This stage of model building has a few more substages namely: model training, model evaluation, and cross-validation. This stage was dedicated towards the working of the machine learning algorithms, where various ML classifiers were trained. The model was evaluated based on the accuracy generated, also improving the accuracy by cross-validation. The model evaluation was performed by employing various ML algorithms for the learning and classification of data for model generation. The final step is the result generation, which includes the step of model prediction, and evaluation of the model generated in the previous step. Finally, the produced result is evaluated based on the performance measure indices as shown graphically.

5.3 Data Collection

For this study, the dataset was chosen from the OASIS project. (https://www.oasis-brains.org/). The initial OASIS dataset consists of cross-sectional MRI Data in all age groups from young, middle-aged to nondemented and demented older adults. There were 416 subjects which included both males and females ranging from age 18 to 96. They were all right-handed subjects. 100 among the subjects over 60 years showed symptoms of AD who had their MRI taken multiple times on a single occurrence. The data obtained from the data preparation stage is further pre-processed. This stage has following sub-stages: data visualisation, data imputation, and transformation. The dataset consists of 373 rows and 10 columns.

5.4 Handling Missing Data using Mean Imputation

A large amount of data can be collected through the internet, surveys, and experiments. These raw datasets usually contain missing values, noise, and distortions.

Imputation is the process of substituting appropriate values for the missing values/ null. Data imputation needs to be performed carefully by preserving the relations in the data [12]. For the dataset, this problem was resolved by using the mean imputation method. Firstly, we calculated the mean of all the values for that attribute. Then the missing value for a given attribute in a record was filled in with the mean. SES and MMSE have 19 and 2 nulls, they both were replaced using mean imputation technique.

5.5 Feature Selection

Feature selection is the method of choosing the features to be used and removing input features that are not important for classification [18-19]. For this dataset, certain features were irrelevant. However, no redundant feature set existed. There are so many algorithms for feature selection, however, we have selected two of the feature selection algorithms which are the Least Absolute Shrinkage and Selection Operator (LASSO) and Chi-square test.

A. Least Absolute Shrinkage and Selection Operator (LASSO):

LASSO is a shrinkage method that performs both variable selection and regularization [20].

B. Chi-square:

The Chi-square test is used in statistics to measure the lack of correlation between the features [19]. For this research, Chi-square measured how the expected count and observed count deviated from each other.

The formula for Chi-square is,

$$\chi^2 = \frac{\Sigma((Observed - Expected)^2}{Expected}$$

The more the difference, the higher the chi-square value[21].

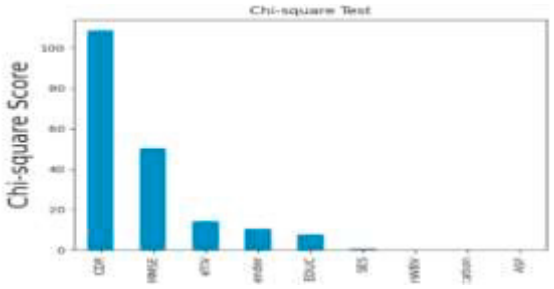

Figure **3**: Feature ranking after applying Chi-square method.

After applying the chi-square technique, the feature was be ranked as shown in figure 3. A feature having a significant value of 0.05 was chosen, while the other were eliminated. For training the model, we have taken six features 'CDR', 'MMSE', 'eTIV', 'Gender', 'EDUC' and 'SES'. Three features have been eliminated which have a lower chi-square than 0.05.

5.6 Data Segregation

The next step is data segregation. The clean data obtained from the data pre-processing was further segregated. Data segregation helps to avoid overfitting, underfitting issues. It is seen that, when the model is trained for familiar data, it tends to fail to generalise with unseen data. Data splitting is used to resolve this issue of generalization. The dataset was split into two portions: training and testing. The train and test set were divided in the ratio of 70:30. The model for train-test sometimes is not reliable, K-fold cross-validation provides a solution to this problem. For our research, we also have used the 10-fold cross-validation technique. The data was divided into N = 10 samples for better validation.

5.7 Model Building

The data prepared in the model segregation step was used for model building. Building a model is training a ML algorithm that can predict the label (target variable) from the features (independent variables). The training data was supplied into the ML algorithm, where it learnt, and a model was generated. In the training phase, a ML algorithm was trained by feeding the dataset. A machine learning classifier was passed with train data. The output of the model training process was a ML model which was used to make predictions using the sample output (test data). This is called model fitting. The input (train data) has an influence on the output. Learning takes place in the model training step. In our dataset, the target variable has two values: demented and non-demented.

6 Analysis and Discussion of Results

6.1 Model Validation

This section compares the outcomes of the different classification models with different input features. Firstly, all the machine learning algorithms are applied to all features of the OASIS dataset. Secondly, Least Absolute Shrinkage and Selection Operator Feature Selection Algorithm was implemented to extract some relevant features and the nine machine learning algorithms were applied again. Finally, the features selected by Chi-2 were used as input to the classifiers. Different performance metrics as explained in the previous section are also evaluated to evaluate the predicted outcomes.

Table 1: Scores for all Features without Validation

	Accuracy	Precision	F1-Score	Recall	FPR
AdaBoost	88.7	87.93	87.93	87.93	10.60
Decision Tree	89.51	86.6	88.1	89.6	10.60
Extra Tree	83.87	79.68	83.6	87.93	**19.69**
Gradient Boost	88.7	87.93	87.93	87.93	9.09
K-NN	95.16	98.14	**93.59**	**91.37**	1.51
Logistic Regression	94.35	100	93.57	87.93	0
Naïve Bayes	94.35	100	93.57	87.93	0
Random Forest	94.35	100	93.57	87.93	0
Support Vector Machine	**96.77**	**100**	93.57	87.93	0

Table 2: Scores for LASSO selected features without Validation

	Accuracy	Precision	F1-Score	Recall	FPR
AdaBoost	88.70	87.5	87.93	87.93	10.60
Decision Tree	89.51	87.5	87.93	89.65	10.60
Extra Tree	76.61	73.77	75.63	73.77	**24.24**
Gradient Boost	89.51	87.5	87.93	87.93	9.09
K-NN	**95.16**	**100**	**94.54**	**100**	1.51
Logistic Regression	94.35	**100**	93.57	87.93	0
Naïve Bayes	94.35	**100**	93.57	87.93	0
Random Forest	94.35	**100**	93.57	87.93	0
Support Vector Machine	94.35	**100**	93.57	87.93	0

Table 3: Scores for Chi-2 selected Features without Validation

	Accuracy	Precision	F1-Score	Recall	FPR
AdaBoost	88.7	**100**	**93.57**	87.93	0
Decision Tree	91.93	92.85	91.22	**89.65**	0
Extra Tree	91.12	91.22	90.43	**89.65**	**7.575**
Gradient Boost	93.54	98.07	92.72	87.93	1.515
K-NN	**94.35**	**100**	**93.57**	87.93	0
Logistic Regression	**94.35**	**100**	**93.57**	87.93	0
Naïve Bayes	**94.35**	**100**	**93.57**	87.93	0
Random Forest	**94.35**	**100**	**93.57**	87.93	0
Support Vector Machine	**94.35**	**100**	**93.57**	87.93	0

Table 1 gives the Scores for all Features without Validation. Considering full features, the highest level of accuracy of 96.77% was achieved by SVM. The least accurate prediction of 83.87% was obtained from the ET classifier. Considering full features, ET had the lowest precision score of 79.68% in precision. The highest precision of 100% was shared between four classifiers, LR, NB, FR and SVM. KNN achieved the highest recall score (91.37%) and DT received the second-highest score (89.6%) when applied to all sets of features. For full features, the highest F1-score is achieved with the KNN classifier which outperformed other algorithms. For full features, the FPR is significantly high for AB, DT, ET, and GB. The highest is 19.69% by ET.

Table 2 gives the Scores for LASSO selected features without Validation. When evaluating LASSO features, KNN generates the highest accuracy (95.16%). We got the lowest accuracy for ET of 76.61%. The best precision result was recorded from LR, NB, RF, SVM and KNN with an accuracy of 100%. The lowest precision of 73.77% was for ET. KNN showed the highest recall score of 100%. A very poor recall score (73.77%) has been generated by ET algorithm. the F1-scores have changed only slightly as compared to the full features. The highest F1-score is 94.54% for KNN. FPR for ET increase dramatically to a value of 24.24%. However, a decrease in the FPR has been observed for other classifiers.

Table 3 gives the Scores for Chi-2 selected Features without Validation. Analysing the accuracy achieved with Chi-2 features, the highest accuracy of 94.35 % was demonstrated by six of the classifiers: AB, KNN, LR, NB, RF, and SVM. All the score was above 90% with ET giving the lowest precision score of 91.22%. 100% precision was attained by AB, LR, NB, RF, KNN, and SVM. The highest recall score has been shown by DT and ET algorithms of

89.65%. For 6 features, the highest F1- score of 93.57% was given by AB, LR, NB, RF, SVM and KNN algorithms and ET had the lowest F1- score of 90.43% for 6 features. The outcomes of FPRs have decreased significantly after the application of Chi-2 feature selection. AB, DT, LR, NB, RF, SNM, KNN have 0% FPR. ET has the highest FPR of 7.575% and GB has 1.515%

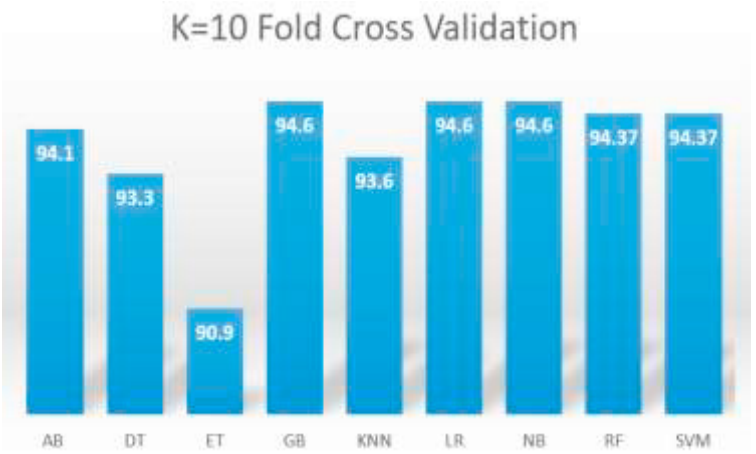

Figure **4**: K-fold Cross Validation

Based on the results and values of performance indices, it can be realized that the proposed model is effective. It can be implemented in real-time scenarios to detect and perform diagnoses. However, the model was trained only to a particular dataset causing issues of overfitting or bias when introducing to an unknown dataset. To overcome any such issues and evaluate the robustness, cross-validation approach was applied to the full set of features. The accuracy scores given are shown in Fig. 4.

Table **4**: Validation score for all features

	Accuracy	Precision	F1-Score	Recall	FPR
AdaBoost	88.7	87.93	87.93	87.93	10.60
Decision Tree	88.71	88.66	88.13	**89.65**	12.12
Extra Tree	83.87	79.68	83.6	87.93	**19.69**
Gradient Boost	88.7	87.93	87.93	87.93	10.606
K-NN	**95.166**	98.18	92.59	86.2	1.51
Logistic Regression	94.35	**100**	**93.57**	87.93	0
Naïve Bayes	94.35	**100**	**93.57**	87.93	0
Random Forest	94.35	**100**	**93.57**	87.93	0
Support Vector Machine	94.35	**100**	**93.57**	87.93	0

Table **5**: Validation score for LASSO selected features

	Accuracy	Precision	F1-Score	Recall	FPR
AdaBoost	88.7	87.93	87.93	87.93	10.60
Decision Tree	89.516	88.13	88.8	**89.65**	10.606
Extra Tree	76.16	76.16	75.63	77.58	**24.24**
Gradient Boost	89.51	89.47	88.69	87.93	9.09
K-NN	**95.166**	**100**	**94.54**	89.65	0
Logistic Regression	94.35	**100**	93.57	87.93	0
Naïve Bayes	94.35	**100**	93.57	87.93	0
Random Forest	94.35	**100**	93.57	87.93	0
Support Vector Machine	94.35	**100**	93.57	87.93	0

Table **6**: Validation score for Chi-2 selected features

	Accuracy	Precision	F1-Score	Recall	FPR
AdaBoost	**94.35**	100	**93.57**	87.93	0
Decision Tree	91.93	92.85	91.22	89.65	0.6
Extra Tree	91.92	91.22	90.43	89.65	**7.575**
Gradient Boost	93.54	98.07	92.72	87.93	1.515
K-NN	**94.35**	98.18	93	**93.1**	1.515
Logistic Regression	**94.35**	**100**	**93.57**	87.93	0
Naïve Bayes	**94.35**	**100**	**93.57**	87.93	0
Random Forest	**94.35**	**100**	**93.57**	87.93	0
Support Vector Machine	**94.35**	**100**	**93.57**	87.93	0

Table 4, 5 and 6 gives the performance comparison between various algorithms with cross validation. Again, Table 7 provides comparison of the proposed model with other approaches for various research in Alzheimer's Dementia. It can be noticed that compared to several previous studies using ADNI and OASIS database, our model has shown promising accuracy. It has been seen that, the study by Ahmad et al. [22] reported a highest achieved performance of 99% using deep learning techniques in ADNI dataset.

Table **7**: Performance comparison between different algorithms

Author	Targeted Work	Dataset	M/L Algorithms	Performance
Ahmad et al. [22]	Early Diagnosis and classification of Alzheimer's disease.	ADNI	Deep Learning	Accuracy -99%
Annette et al. [12]	Survival analysis of Clinical Data	Sydney Memory and Age Study (MAS) & ADNI	10 M/L Algorithms	Performance Values: 0.82 for MAS 0.93 for ADNI
Anastasia et al. [13]	Prediction between Parkinson's and Lewy body Disorder	138 patients with PDD and DLB.	Logistic Regression	87.5%
J Neelavani & M.S Geetha [23]	Alzheimer Prediction	Psychological parameters like age, number of visits, MMSE and education.	SVM, DT	SVM: 85%, DT: 83%
Bo et al. [37]	Diagnosis of MCI, CN and Dementia	Parkinson's Progression Markers Initiative dataset	SVM	Accuracy from Multi-level ROI-based features: 92.35%, 3.91%, 80.84%.

Rammurti et al. [15]	Dementia Detection	OASIS dataset	Stacking GB and ANN	GB: 87%, ANN: 89%, GB + ANN: 89%
Gloria et al. [24]	Differential Diagnosis of Alzheimer and Vascular Dementia.	77 subjects from Neurological Institute IRCCS Mondino Foundation	ANN, SVM, A Neuro-fizzy Inference System	ANFIS – 84%
Proposed Model	Prediction of dementia	OASIS	AB, DT, ET, GB, KNN, LR, NB, RF, SVM	**Highest Accuracy - SVM 96.77%**

7 Conclusions

This research is based on the literature review of previous works and aims to answer the research gap as observed. With the primary motive to design and develop a model for dementia prediction OASIS dataset has been utilized. The motive was to predict the results of the classification of Alzheimer's Dementia from the available data by employing machine learning methods in a way to increase and come up with higher accuracy and improved performance. Nine supervised machine learning algorithms have been applied to different feature sets to identify the best combination for the detection of dementia. The best result was achieved with the Support Vector Machine (SVM) with the full set of features. Analyzing the process involved, it was significant that the selection of features and the pre-processing strategy were key for determining the best model for detection of dementia. Based on the findings, the use of classification algorithms has shown an accuracy of above 90% with SVM giving the highest accuracy. Considering full features, the highest level of accuracy of 96.77% was achieved by SVM. In future research, the study will be more dedicated to improving the efficiency and accuracy of the proposed model. Implementation of ensemble techniques, other feature selection techniques, and feature reduction techniques can be implemented to make the model more accurate.

References

[1] Iadecola, C. Vascular and metabolic factors in Alzheimer's disease and related dementias: introduction. Springer, City, 2016.

[2] Tanveer, M., Richhariya, B., Khan, R., Rashid, A., Khanna, P., Prasad, M. and Lin, C. Machine Learning Techniques for the Diagnosis of Alzheimer's Disease: A Review. ACM transactions on multimedia computing communications and applications, 16, 1s (2020), 1-35.

[3] Nithya, B. and Ilango, V. Predictive analytics in health care using machine learning tools and techniques. City, 2017.

[4] Basheer, S., Bhatia, S. and Sakri, S. B. Computational Modeling of Dementia Prediction Using Deep Neural Network: Analysis on OASIS Dataset. IEEE access, 9 (2021), 42449-42462.

[5] Spooner, A., Chen, E., Sowmya, A., Sachdev, P., Kochan, N. A., Trollor, J. and Brodaty, H. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. Scientific reports, 10, 1 (2020), 20410-20410.

[6] Bougea, A., Efthymiopoulou, E., Spanou, I. and Zikos, P. A Novel Machine Learning Algorithm Predicts Dementia With Lewy Bodies Versus Parkinson's Disease Dementia Based on Clinical and Neuropsychological Scores. Journal of Geriatric Psychiatry and Neurology (2021), 0891988721993556.

[7] Hasib, K.M., Iqbal, M., Shah, F.M., Mahmud, J.A., Popel, M.H., Showrov, M., Hossain, I., Ahmed, S. and Rahman, O., 2020. A survey of methods for managing the classification and solution of data imbalance problem. arXiv preprint arXiv:2012.11870.

[8] Hasib, K.M., Towhid, N.A. and Alam, M.G.R., 2021, November. Online Review based Sentiment Classification on Bangladesh Airline Service using Supervised Learning. In 2021 5th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) (pp. 1-6). IEEE.

[9] Salehi, A. W., Baglat, P. and Gupta, G. Alzheimer's disease diagnosis using deep learning techniques. Int. J. Eng. Adv. Technol, 9, 3 (2020).

[10] Stamate, D., Alghamdi, W., Ogg, J., Hoile, R. and Murtagh, F. A Machine Learning Framework for Predicting Dementia and Mild Cognitive Impairment. City, 2018.

[11] Lee, G. G., Huang, P., Xie, Y. and Pai, M. Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Cognitively Normal Based on Neuropsychological Data via Supervised Learning. City, 2019.

[12] Spooner, A., Chen, E., Sowmya, A., Sachdev, P., Kochan, N. A., Trollor, J. and Brodaty, H. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. Scientific reports, 10, 1 (2020), 20410-20410.

[13] Bougea, A., Efthymiopoulou, E., Spanou, I. and Zikos, P. A Novel Machine Learning Algorithm Predicts Dementia With Lewy Bodies Versus Parkinson's Disease Dementia Based on Clinical and Neuropsychological Scores. Journal of Geriatric Psychiatry and Neurology (2021), 0891988721993556.

[14] Salehi, A. W., Baglat, P., Sharma, B. B., Gupta, G. and Upadhya, A. A CNN Model: Earlier Diagnosis and Classification of Alzheimer Disease using MRI. City, 2020.

[15] Sivakani, R. and Ansari, G. A. Machine Learning Framework for Implementing Alzheimer's Disease. City, 2020.

[16] Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C. and Buckner, R. L. Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults. Journal of Cognitive Neuroscience, 22, 12 (2010), 2677-2684.

[17] Manlangit, S., Azam, S., Shanmugam, B., Karim, A. Novel machine learning approach for analyzing anonymous credit card fraud patterns, International Journal of Electronic Commerce Studies. 10 (2019). doi:10.7903/ijecs.1732.

[18] Karim, A., Azam, S., Shanmugam, B., Kannoorpatti, K., Alazab, M. A comprehensive survey for intelligent spam email detection, IEEE Access. 7 (2019) 168261–168295. doi:10.1109/access.2019.2954791.

[19] Antony, L., Azam, S., Ignatious, E., Quadir, R., Beeravolu, A. R., Jonkman, M. and Boer, F. D. A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction. IEEE Access, 9 (2021), 126481-126501.

[20] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F.J.M., Ignatious, E., Shultana, S., Beeravolu, A.R. and De Boer, F., 2021. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access, 9, pp.19304-19326.

[21] Sumaiya Thaseen, I. and Aswani Kumar, C. Intrusion detection model using fusion of chi-square feature selection and multi class SVM. Journal of King Saud University - Computer and Information Sciences, 29, 4 (2017/10/01/ 2017), 462-472.

[22] Arif, M., Ahmad, S., Ali, F., Fang, G., Li, M. and Yu, D.-J. TargetCPP: accurate prediction of cell-penetrating peptides from optimized multi-scale features using gradient boost decision tree. Journal of computer-aided molecular design, 34, 8 (2020).

[23] Neelaveni, J. and Devasana, M. G. Alzheimer disease prediction using machine learning algorithms. IEEE, City, 2020.

[24] Castellazzi, G., Cuzzoni, M. G., Cotta Ramusino, et al. A Machine Learning Approach for the Differential Diagnosis of Alzheimer and Vascular Dementia Fed by MRI Selected Features. Frontiers in Neuroinformatics, 14, 25 (2020-June-11 2020).