A systematic analysis of assorted machine learning classifiers to assess their potential in accurate prediction of dementia

Afreen Khan and Swaleha Zubair
Department of Computer Science, Faculty of Sciences, Aligarh Muslim University, Aligarh, India, and

Samreen Khan
*Department of Community Medicine,
Integral Institute of Medical Sciences and Research, Integral University, Lucknow, India*

Abstract

Purpose – This study aimed to assess the potential of the Clinical Dementia Rating (CDR) Scale in the prognosis of dementia in elderly subjects.

Design/methodology/approach – Dementia staging severity is clinically an essential task, so the authors used machine learning (ML) on the magnetic resonance imaging (MRI) features to locate and study the impact of various MR readings onto the classification of demented and nondemented patients. The authors used cross-sectional MRI data in this study. The designed ML approach established the role of CDR in the prognosis of inflicted and normal patients. Moreover, the pattern analysis indicated CDR as a strong cohort amongst the various attributes, with CDR to have a significant value of $p < 0.01$. The authors employed 20 ML classifiers.

Findings – The mean prediction accuracy varied with the various ML classifier used, with the bagging classifier (random forest as a base estimator) achieving the highest (93.67%). A series of ML analyses demonstrated that the model including the CDR score had better prediction accuracy and other related performance metrics.

Originality/value – The results suggest that the CDR score, a simple clinical measure, can be used in real community settings. It can be used to predict dementia progression with ML modeling.

Keywords Alzheimer's disease, Clinical dementia rating scale, Cross-sectional, Dementia, Machine learning, MRI, Prediction

Paper type Research paper

1. Introduction

Dementia has been considered a brain malfunctioning condition that adversely affects the cognitive attributes of an individual (Barrett & Burns, 2014; Emanuele, Alessandro, Victòria, Federica, & Lavinia, 2021). Alzheimer's disease (AD) is a type of dementia characterized by a

Rights: This study does not contain any data directly procured from human or animal participants.
Informed consent: Not applicable.

decline in cognitive functions, particularly memory, as well as language and problem-solving (Barrett & Burns, 2014; Linz, Troger, Alexandersson, Konig, Robert, & Wolters, 2017; Porsteinsson, Isaacson, Knox, Sabbagh, & Rubino, 2021). Aside from AD, other conditions such as stroke, traumatic brain injuries and brain tumors can also cause this condition (Linz *et al.*, 2017). Considering the world's growing aging population, global estimates of AD prevalence are anticipated to rise from 25 to 63 million by 2030 and 114 million by the year 2050 (Cacchione, Powlishta, Grant, Buckles, & Morris, 2003). There have been subtle but meaningful advances in the pharmacological therapy of this deadly disease over the last two decades. Even so, early and accurate diagnosis of dementia in its earliest stages is required for successful management and treatment to limit disease development (Wimo, Winbald, Torres, & Strauss, 2003; Ribeiro, Lopes, & Lourenço, 2013).

The early prognosis of dementia could be made plausible by the identification of disease-associated reliable markers (Ranson *et al.*, 2021). In this regard, various neuropsychological, biochemical and genetic-based markers may help in monitoring dementia progression. Sheehan highlighted various assessment scales to assess the severity of dementia. Numerous small dementia screening tests have been reported in this paper, which are suitable for primary and secondary healthcare (Sheehan, 2012). They are Mini Mental State Examination (MMSE), Abbreviated Mental Test Score, Clock-drawing test, Six-item Cognitive Impairment Test, General Practitioner Assessment of Cognition, Mini-Cog, Test Your Memory, Montreal Cognitive Assessment (MoCA), Addenbrookes Cognitive Assessment and Memory Impairment Screen (Sheehan, 2012). For overall dementia severity, Clinical Dementia Rating (CDR), Global Deterioration Scale and Clinicians Global Impression of Change were showed in his findings (Sheehan, 2012). Keeping this fact into consideration, it would be appropriate to look for a cost-effective and simple diagnostic marker that can be easily accessed in routine clinical settings. A variety of studies have recommended CDR and MMSE as a strong predictor for the progression of AD-related dementia (Nakata *et al.*, 2009; Daly *et al.*, 2000). In a recent study, Wessels, Dowsett & Sims performed a comparative analysis of ADAS-Cog (Alzheimer's Disease Assessment Scale – Cognitive Subscale) and CDR-SB (Clinical Dementia Rating – Sum of Boxes [SOB]) for detecting the treatment group differences in AD (Wessels, Dowsett, & Sims, 2018). They found that the ADAS cognitive subscale was more frequent than the CDR-SB scale in the detection of AD treatment group differences (Wessels *et al.*, 2018).

CDR scale, proposed by the Washington University, is a global assessment tool that generates two outputs, namely (1) global and (2) SOB scores. The global score is frequently employed to grade various stages of dementia severity in inflicted patients (O'Bryant *et al.*, 2010). This can be widely exploited in both research as well as clinical settings (Hughes, Berg, Danziger, Coben, & Martin, 1982). The CDR Scale is significantly used to evaluate the functional and cognitive impairment status in AD patients (Morris, 1993; Morris *et al.*, 1997). It is represented by a five-point scale and is used to assess six major fields of functional and cognitive performances (Hughes *et al.*, 1982; Morris, 1993). These domains include memory, judgment and problem-solving, orientation, home and hobbies, community affairs, personal care, *etc*. This scale has been introduced to build diagnostic differences among the nondemented healthy versus demented patients. It covers demented patients very mild score (i.e. CDR 0.5) (Lim, Chong, & Sahadevan, 2007; Balsis, Miller, Benge, & Doody, 2011). In line with this, Coley *et al.* in their study reported the suitability of the CDR-SB scale as a primary measure in AD tests (Coley *et al.*, 2011). They further claimed that CDR-SB alone can be employed as a sole cognitive assessment scale for AD trials (Coley *et al.*, 2011).

Apart from the CDR testing variable, MoCA, MMSE, etc. are other standardized tests that have been formulated for dementia. But CDR score can alone aid in the detection of dementia and related brain deterioration. The diagnosis of dementia and other cognitive related diseases has remained an extremely challenging task (Barrett & Burns, 2014). For this reason,

the development of easy to operate and efficient techniques are crucial for timely detection leading to proactive interventions (Prince, Herrera, Knapp, Guerchet, & Karagiannidou, 2016; Prince, Albanese, Guerchet, & Prina, 2014).

In the healthcare sector, the implementation of machine learning (ML) can provide an effective method of using dense information needed for an accurate diagnosis. ML, a science of pattern learning, has the unique ability to deal with bulky datasets leading to the development of précised predictive models (Khan & Zubair, 2018). ML allows the automatic selection of high-value predictors from a pool of possible inputs (Amoroso et al., 2017). The application of magnetic resonance imaging (MRI) along with the complex ML algorithms is being used to distinguish the healthy brain from that of the mildly demented brain (Battineni, Sagaro, Chinatalapudi, Amenta, 2020a; Battineni, Chintalapudi, Amenta, & Traini, 2020b). Battineni et al. reviewed 435 articles published between 2015 and 2019, based on applications of ML in the diagnosis of chronic diseases (Battineni et al., 2020a, b). They finally selected 22 studies to present a comparative analysis (Battineni et al., 2020a, b). Among these 22 reviewed papers, they found dementia as one of the chronic diseases with case-control as a study type, MRI as input features and support vector machine (SVM) classifier was employed for ML modeling (Battineni et al., 2020a, b). Furthermore, in another referenced paper, Battineni et al. built an ML model for dementia prediction using the SVM classifier (Battineni, Chintalapudi, & Amenta, 2019). They performed ML modeling on the longitudinal pool of 150 MRI patients and reached a prediction accuracy of 65.75% (Battineni et al., 2019). Khan & Zubair proposed an improved multi-modal ML pipeline for the prognostication of AD (Khan & Zubair, 2020a, b, c, d). Their findings showed to have an accuracy of 87.0% on the random forest ML classifier, built on Open Access Series of Imaging Studies (OASIS) longitudinal MRI data (Khan & Zubair, 2020a, b, c, d).

Deep learning is a branch of ML that involves several layers of information processing steps in a hierarchical way for unsupervised feature learning and pattern classification challenges (Deng & Yu, 2014). Deep neural models typically outperform shallow models when confronted with challenging learning problems (Onan & Toçoğlu, 2021; Onan, 2022). It is a current direction of research in ML that tries to develop a classification scheme with greater prediction performance based on numerous layers of nonlinear information processing (Onan, 2020). Also, it is gradually making its way into novel tools with high-value clinical applications in the real world. Innovative patient-facing applications and a few surprisingly established methodologies in image analytics and diagnostics are among the most promising use cases.

As stated above, a number of studies and extensive research have been done recently on the prediction, progression and diagnosis of dementia and AD based on ML methodologies (Nori, Hane, Martin, Kravetz, & Sanghavi, 2019). Among these, several studies reported employing only traditional-based ML classifiers, which employs classifiers with no tweaking of hyperparameters or ensembling approach. This caused to have a model with lower accuracy and reduced performance. Besides, a proper sequential data preprocessing workflow that can deal effectively in handling the missing data, outliers and imbalanced classification problem is the kernel of any productive ML model (Lim et al., 2007; Balsis et al., 2011). Thus, this is a severe issue and hence a challenge to build such a system where dementia and related disorders can be predicted effectively (with higher accuracy and improved performance) based on neuropsychological and demographic data. There is a considerable need for an effective, reliable, cost-effective and scalable screening method for AD and its early stages, which can encompass a proportion of patients with mild cognitive impairment (MCI), or subjective memory complaints (SMCs), to expedite preventative clinical investigations and address under-diagnosis in the population. Multiple classifier systems (often referred as ensemble classifiers) have been extensively used in the realm of pattern

recognition to construct robust classification schemes due to their significant improvement in generalizability and predictive accuracy (Onan, 2018a, b).

Thus, the present study is significant in that we attempted to develop a cross-sectional MRI features-based ensemble ML strategy to diagnose dementia. Following that, we used ML to assess the efficacy of the CDR score. This ML program was further used in the prediction of dementia progression in older adults. On the basis of cognitive and functional measures of demented versus healthy brains, the predictive power of as-developed ML was compared. The primary goal of this study was to use demographic and neuropsychological data to evaluate CDR scores and their application in dementia diagnosis. We generated different in-house developed ML models using OASIS cross-sectional MRI data. Aside from classifying subjects, another objective of this paper was to determine the data from neuropsychological tests and demographic data that can be effectively used in the accurate diagnosis of dementia in the affected patients.

2. Methods
2.1 Study design and participants
The study included 416 individuals who were enrolled for 434 MRI sessions. The cross-sectional MRI data used in this study were acquired from the OASIS database (http://www.oasis-brains.org/). The set was built on the basis of the collection of MRI data of the subjects reported through the Washington University Alzheimer Disease Research Center (ADRC).

Out of 416 participants, there were 160 male and 256 female participants. All the participants were aged between 18 and 96 years. For the present analysis, we included all subjects. The structural MRI scans were T1-weighted magnetization. These were obtained on a 1.5-T (Tesla) Vision Scanner. A high-resolution MPRAGE (Magnetization Prepared Rapid Acquired Gradient Echo) sequence was used for examining the MRI data. For each study participant, 3 to 4 distinct images were obtained. The MRI acquisition parameters are as follows: flip angle (°) = 10, TD (msec) = 200, TE (echo time in msec) = 4.0, TI (msec) = 20, TR (repetition time in msec) = 9.7, orientation = sagittal, resolution (pixels) = 256×256 (1×1 mm), slice number = 128 and thickness and gap (mm) = 1.25 and 0 (Marcus et al., 2007).

The dataset consisted of 12 features, corresponding to both demographic and neuropsychological data, which were recorded during the MR imaging acquisition. These features include identification number, gender, handedness (left-handed or right-handed), age, educational years, socioeconomic status (SES), MMSE, CDR, estimated total intracranial volume (eTIV), normalized whole brain volume (nWBV), atlas scaling factor (ASF) and magnetic resonance delay time. The above-specified features were assigned before the image acquisition.

2.2 Baseline clinical assessment
The status of dementia was established and classified based on the CDR Scale as suggested by earlier studies (Morris, 1993; Morris et al., 2001). We also made sure that the determination of Alzheimer's disease or its control status was centered only on the clinical approaches, with no reference to psychometric performance. We also ensured that any possible alternate causes of dementia were absent during data acquisition. There were three scores, namely, SES, MMSE and CDR, present in the dataset. These scores, along with the other MRI attributes, helped in uncovering the status of dementia. According to the Hollingshead Index, SES refers to the social position of an individual. It was classified into two groups explicitly into the lowest status and highest status. The lowest status corresponds to the value "0", while the highest status refers to "5" (Lynch & Kaplan, 2000).

MMSE, also called as Folstein test, consists of a questionnaire. It is broadly used to assess the cognitive ability of the inflicted patient. The MMSE score ranges from 0 to 30. A score below 10 represents an extreme impairment, a score in the range of 10-19 implies moderate dementia, score varying between 19 and 24 signifies an early stage of Alzheimer's disorder and a score above 25 corresponds to a normal healthy individual (Magni *et al.*, 1996).

CDR is a five-point score that is determined through a semi-structured discussion with the patient. Besides this, three other anatomical measures, namely nWBV, eTIV and ASF, have also been implicated. These are employed to analyze the brain anatomical features in MRI images, specifically relative to aging.

2.3 CDR as an assessment tool

To diagnose the dementia disorder, multiple cognitive impairments along with functional impairments must be present. These impairments result from the cognitive impairments when no sign of delirium or any other nonorganic aetiology (cf. major depression) is associated. CDR is a tool to study various stages of dementia. It provides a five-point score, which has been classified into six domains. Among these, memory is the only primary subscale; while orientation, community affairs, judgment, problem-solving tasks, hobbies, and personal care represent the five secondary subscales.

In the present study, we used the CDR score to determine whether a particular individual is demented. The CDR score can also be used to assess the severity of dementia. Based on collateral source and examination of the subject, an overall CDR score is obtained from each of the six different domains. A global CDR score of 0 denotes no dementia whereas CDRs of 0.5, 1, 2 and 3 represent questionable deficit, mild, moderate and severe dementia, respectively.

2.4 Machine learning modeling

ML methods have been widely used in the diagnosis of individuals suffering from MCI or dementia. On the contrary, differentiating these categories through only one model has remained challenging. For that reason, the objective of this study was to use the demographic and neuropsychological data to predict the status of the dementia disorder on the basis of CDR scores (i.e. normal [CDR = 0], very mild dementia [CDR = 0.5; proxy for MCI], mild dementia [CDR = 1], moderate dementia [CDR = 2] and severe dementia [CDR = 3]) by applying the various well-known ML models. Through this, we were able to establish the role of CDR in predicting the stage of impairment. At first, the clinical diagnosis was achieved by considering scores based on neuropsychological features. Because of this, we trained the ML algorithms in such a way that result in higher classification accuracy when predicting the CDR values instead of the other diagnostic values.

The abundance of available data poses a great challenge for ML-based classification. The volumetric data make it harder to train the learning algorithms in an appropriate duration of time and reduces the classification accuracy of the resulting model. In order to create robust and effective classification models while lowering training time, feature selection becomes a crucial challenge (Onan & Korukoğlu, 2017). The following independent variables were used in the ML analysis: gender, age, years of education, SES, MMSE, eTIV, nWBV and ASF, while identification number, handedness and delay features were dropped. CDR testing variable was the dependent variable, i.e. the response attribute.

Constructing an ML model means training an ML algorithm that can predict the labels, i.e. target variable among the set of independent variables, tune it and validate it. We applied twenty supervised ML models as these resulted in more accurate results. Ensemble learning can be used to provide more reliable classification strategies (Onan, Korukoğlu, & Bulut, 2017). It seeks to improve the predictive accuracy of a classification model by mixing the

results of various learning algorithms (Onan, 2018a, b). The data preprocessing is the initial step in ML modeling. In concordance, we applied ten data pre-processing steps as reported in one of our previous study (Khan, Zubair, & Sabri, 2019). Missing data is a very common and challenging issue, mainly when dealing with real-world datasets. The effect of such data on the ML model has revealed that the result gets degraded by assigning a random value to the missing data (Khan & Zubair, 2019). In this study, we handled the class of missing data by applying the "imputation by median" method. This is used because it operates extremely well for continuous data that comprises of outliers. During the model building stage, we applied repeated stratified K-Fold cross-validation. This approach aids in improving the prediction accuracy of the ML model. It repeats stratified K-Fold n times with distinct randomization in every repeat. This further helps in reducing the noisy performance of the entire ML model. The basic characteristic of cross-validation was twofold. Initially, it divided the data into two segments to compare them statistically. Next, it evaluated the as-developed algorithms.

An imbalanced classification problem is a severe problem in cases where the data distribution is usually biased across the target variable (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). It occurs in ML applications in which one class (referred as the minority class) has a minuscule number of instances while the other class (termed as the majority class) has an enormous number of instances (Onan, 2019). It poses a challenge while building an ML model. If not treated well, the imbalance classification leads to the development of an ML model that ignores and result in a poor performing model with lower accuracy. In the present study, we resolved the imbalanced classification problem using Synthetic Minority Oversampling Technique (SMOTE). This is a data augmentation technique that specifically deals with minority class (Chawla *et al.*, 2002). There were four classes of the target variable, i.e. CDR, consisting of normal, very mild, moderate and severe demented subjects. In particular, there was a disparity found in the last two CDR classes. Because of this, the built ML model resulted in poor performance and hence lower accuracy. Hence, we employed SMOTE analysis, which oversampled the minority CDR class. Oversampling only balances the class distribution; it does not add any extra information to the ML model. Before testing each classifier, it was tuned with the hyperparameters. Hyperparameters primarily help in structuring the ML model (ML tuning is a kind of optimization problem). Post access to a collection of hyperparameters, we tried to discover the correct combination of their values. This helps in examining the performance metrics of the classifiers with maximum accuracy and other related metrics.

We worked on several ML classifiers but present only those models which gave more accurate results. The following classifiers were employed while building an ML model.

2.4.1 Generalized linear models:.

(1) Logistic regressionCV: In the field of ML, it is a well-known mathematical modeling algorithm applied to epidemiology datasets (Jayatilake & Ganegoda, 2021). It is based on a regularized logistic regression ML algorithm and comprises built-in cross-validation. It does optimization by employing the LIBLINEAR library (library for large linear classification). This library is created for large linear-based classification (Fan, Chang, Hsieh, Wang, & Lin, 2008). It supports both L1 and L2 type of regularization (Pedregosa *et al.*, 2011). This poses a great advantage over other classifiers.

(2) Passive aggressive: It is an online learning ML classifier. In such a type of classifier, data are learned in a sequential manner instead of batch learning. While inputting the data sequentially, the ML model is updated at every step. As it is a group of online algorithms, here distinct ML classifiers, which are based on binary and multiclass regression, classification, uniclass prediction and sequence prediction, are analyzed specifically (Crammer, Dekel, Keshet, Shwartz, & Singer, 2006).

(3) Perceptron: A perceptron ML classifier for classification employs linearly separable functions. In other words, it performs classification and builds a predictive model based on a linear-predictor function, which combines a group of weights along with the feature vector.

(4) Ridge classifierCV: This classifier includes an in-built cross-validation characteristic, where a generalized cross-validation technique is applied by default. This is implemented in a leave-one-out manner (Pedregosa *et al.*, 2011). The advantage of this classifier over other classifiers is that it works efficiently when the number of features is larger than the sample size (Pedregosa *et al.*, 2011).

(5) Stochastic Gradient Descent (SGD): It is a type of optimization-based classifier. It applies regularized linear ML algorithms together with SGD. It is based on discriminative learning, where it is used to find the functions' parameter values that reduce the value of the cost function.

(6) eXtreme Gradient Boosting (XGB): This classifier implements gradient boosting along with Decision Trees classifier to enhance the speed and hence the performance.

2.4.2 Naïve Bayes:.

(7) BernoulliNB: It is an implementation of Naïve Bayes ML algorithm, built specifically for multivariate Bernoulli models (Fan *et al.*, 2008). This classifier works smoothly on discrete data.

(8) GaussianNB: Naïve Bayes classifier is a robust classifier built for predictive modeling (Jayatilake & Ganegoda, 2021). It follows the Bayes' theorem, which states that the probability of a hypothesis can be calculated from its preceding probability (Khan & Zubair, 2020a, b, c, d). Naïve Bayes classifier is further extended, which based on Gaussian/normal distribution, called as GaussianNB (Pedregosa *et al.*, 2011). In this, only the mean and standard deviation is estimated from the training set, which is hence the easiest to work with (Pedregosa *et al.*, 2011).

2.4.3 Support vector machine:.

(9) LinearSVC: This is an SVM classifier, built on a linear kernel. Instead of libsvm library, it implements LIBLINEAR library libsvm (Pedregosa *et al.*, 2011). Besides, it fits the training data, and as an output, it gives a hyperplane that segregates and performs classification of the training data effectively. From this hyperplane, classification and prediction of the input data are modeled.

(10) Support Vector Classification (SVC): This is also an SVM classifier, however, it is built on a radial basis function (rbf) as a kernel. It implements libsvm library instead of LIBLINEAR library. The multi-classification is dealt in a one-vs-one approach (Pedregosa *et al.*, 2011). With a sample size greater than tens of thousands, this classifier fails to perform well as the fitting time gauges quadratically at the least with the sample size (Pedregosa *et al.*, 2011). This is a major disadvantage if the sample size is huge. The advantages of utilizing SVC include its potential to accurately predict without sacrificing generalizability and its robustness to outliers (Alanazi, 2022).

2.4.4 Discriminant analysis:.

(11) Linear discriminant analysis (LDA): This is a classifier that has a linear decision boundary. This boundary is built by fitting class conditional densities to the input data, employing Bayes' theorem. The predictive model is generated by fitting the Gaussian density to every class/label (Pedregosa *et al.*, 2011).

(12) Quadratic discriminant analysis (QDA): This classifier has a quadratic decision boundary. It means that this boundary is built by fitting class conditional densities to the input data, employing Bayes' theorem. Besides, the predictive model is produced by fitting the Gaussian density to every class/label; which is similar to the working of LDA (Pedregosa *et al.*, 2011). The only difference between LDA and QDA lies in the decision boundary.

2.4.5 Neighbor-based:.

(13) K-nearest beighbors (KNN): It is a supervised classification algorithm, based on instance learning. It collects the training data instances and considers those observations only that lies in the near vicinity of the collected instances that the KNN classifier predicts (Zhang, 2016). It works well for smaller sample size but eventually fails to perform well as the input data size grows. A prediction for a new observation is achieved by first identifying the most similar occurrences and then aggregating the output variable based on these occurrences (Jayatilake & Ganegoda, 2021).

2.4.6 Tree-based:.

(14) Decision Trees: It is a supervised ML classifier, which splits the data constantly according to the passed parameters. It comprises nodes and leaves. The nodes are the decision nodes from where the data are split. And the leaves are the one that represents the final result. This classifier learns from a pattern, splits accordingly and generates a decision output. The decision tree helps to build in such an ML model that predicts the class/label by learning/training from the input decision rules, which are determined from the input data variables (Amancio *et al.*, 2014).

2.4.7 Ensemble-based:.

(15) Adaptive Boosting (AdaBoost): It is an ensemble-based classifier. Ensemble classifiers are those ML classifiers that execute by the amalgamation of numerous other ML classifiers (Khan & Zubair, 2020a, b, c, d). Through the ensembling technique, an improved and well-performing model is achieved. The AdaBoost classifier works effectively on the weak ML algorithms, thereby generating a robust ML model as a result (Pedregosa *et al.*, 2011).

(16) Bagging: This is also an ensemble-based meta-estimator. It fits the base algorithms (distinct algorithms that are chosen to form an ensemble) on arbitrary subsets of the input data. It aggregates the prediction results attained from each base algorithm, either by averaging or voting approach to generate a final prediction as an output result. It reduces the variance of a base algorithm by including randomization while the model is built, and then an ensemble ML model is created out of it (Pedregosa *et al.*, 2011).

(17) Extremely Randomized Trees (Extra Trees): It is also based on the ensembling technique. The results attained from the numerous de-correlated decision trees are aggregated in a forest-like structure, from where the output result is generated (Geurts, Ernst, & Wehenkel, 2006). It more likely behaves on the patter of the Random Forest classifier. The only difference lies in the way the decision trees are built within the extra trees classifier.

(18) Gradient boosting: It creates an additive ML model in a staged manner and performs optimization of random differentiable loss functions (Pedregosa *et al.*, 2011). Moreover, it produces a predictive model consisting of an ensemble of several weak ML classifiers, in general decision trees (Natekin & Knoll, 2013).

(19) Random Forest: This classifier is also based on the ensembling method. But it includes only one ML classifier as a base estimator, i.e. several decision trees are used to form a forest-like structure. It fits decision trees on distinct subsets, which acts as an ensemble (Denisko & Hoffman, 2018). The algorithm's generalization error is determined by the power of the individual trees and the relationship between trees (Onan, Korukoğlu, & Bulut, 2016).

2.4.8 Neural network:.

(20) Multi-layer perceptron (MLP): It is based on the architecture of a neural network. It comprises of three node layers, i.e. input, hidden and output layer. It performs classification using backpropagation for training the input data. Also, the optimization of the log-loss function is done employing stochastic gradient descent or lbfgs as a solver, which belongs to a class of quasi-Newton techniques (Pedregosa *et al.*, 2011).

2.5 Performance evaluation metrics
A performance evaluation method is an approach for evaluating an ML model. It is the execution of measurement for the predictions made by a trained model on the test set. The first metric is classification accuracy. It computes the proportion of correct predictions. When predicting positive instances, precision measures how often the prediction is correct. It also determines whether or not the model is correct. The true positive rate is defined as recall. It counts the number of times the prediction is correct when the actual value is positive. The F1 score assesses the model's performance by measuring the accuracy of the test set. It is the harmonic mean that maintains a balance between precision and recall (Robinson, Tang, & Taylor, 2015). It ranges from 0 to 1, where 0 signifies the worst F-measure while 1 denotes the best F-measure.

3. Data analysis
ML models are probabilistic and statistical models that lookup for insights and obtain specific trends and patterns in the data by employing computational algorithms (Khan & Zubair, 2020a, b, c, d). The data analysis was achieved in two different stages: the first being, sample characteristics and the second, exploratory data analysis (EDA).

The data analysis is a practice of collecting, discovering and presenting voluminous data to determine the underlying patterns. It is vital for making data-centered decisions. EDA is a method that analyzes the datasets and summarizes their chief properties. It is more of a graphical and visual approach. An EDA is not the same as statistical visualization; even though both of the terms are used interchangeably (NIST/SEMATECH, 2003). The statistical analysis focuses only on one data characterization part. In contrast to this, EDA covers a larger aspect. It follows a more straightforward methodology of letting the data itself disclose its underlying structural model. The results of the performed analysis are discussed in the below sub-sections.

3.1 Sample characteristics
The demographic description of the subjects used in the study is shown in Table 1. The dataset comprises of subjects across the adult lifespan, i.e. from 18 to 96 years. The entire age and their respective diagnostic properties are presented in Table 2.

From Tables 1 and 2, it can be deduced that of the elder subjects, 98 subjects turned up to have a CDR score of 0, implying no dementia. While 100 subjects exhibited a CDR score greater than 0 indicating to be inflicted by very mild to moderate dementia to moderate AD.

Factors	No of patients	Female	Male	Age range (in years)	Age M ± SD	Education Range (in years)	Education M ± SD	MMSE score Range (in years)	MMSE M ± SD
Normal (CDR = 0)	316	197	119	60-94	75.9 ± 9.0	8–23	14.5 ± 2.9	25–30	29.0 ± 1.2
Very mild dementia (CDR = 0.5)	70	39	31	63-92	76.4 ± 7.0	6–20	13.8 ± 3.2	14–30	25.6 ± 3.5
Mild dementia (CDR = 1)	28	19	9	62-96	77.2 ± 7.5	7–20	12.9 ± 3.2	15–29	21.7 ± 3.8
Moderate (CDR = 2)	2	1	1	78-86	82 ± 5.7	8–14	11 ± 4.2	15	15.0 ± 0.0

Table 1.
subjects' demographic
status (M = mean,
SD = STANDARD
deviation)

Age class	Frequency	No dementia Male	No dementia Female	No dementia Mean	Count of CDR	With dementia Male	With dementia Female	With dementia Mean	Count of CDR 0.5/1.0/2.0
0-20	19	10	9	18.53	0	0	0	0.0	0/0/0
20-30	119	51	68	22.82	0	0	0	0.0	0/0/0
30-40	16	11	5	33.38	0	0	0	0.0	0/0/0
40-50	31	10	21	45.58	0	0	0	0.0	0/0/0
50-60	33	11	22	54.36	0	0	0	0.0	0/0/0
60-70	40	7	18	64.88	0	6	9	66.13	12/3/0
70-80	83	10	25	73.37	0	20	28	74.42	32/15/1
80-90	62	8	22	84.07	0	13	19	82.88	22/9/1
>=90	13	1	7	91.00	0	2	3	92.00	4/1/0
Total	416	119	197			41	59		

Table 2.
Diagnostic
characteristics of the
subjects according to
their age group

Furthermore, the statistical description of the values included in the MRI dataset is given in Table 3 (excluding the NaN values). It gives the central tendency and dispersion of the dataset features.

3.2 Exploratory data analysis
It is a data analysis method that employs numerous graphical techniques to perform a variety of tasks such as, to uncover the dataset's underlying structure, extracting essential features,

	Age	Educ	SES	Attributes MMSE	CDR	eTIV	nWBV	ASF
Count	434.00	235.00	216.00	235.00	235.00	434.00	434.00	434.00
Mean	51.36	3.18	2.49	27.06	0.29	1481.92	0.79	1.20
Standard deviation	25.27	1.31	1.12	3.70	0.38	158.74	0.06	0.13
Minimum	18.00	1.00	1.00	14.00	0.00	1123.00	0.64	0.88
25%	23.00	2.00	2.00	26.00	0.00	1367.75	0.74	1.11
50%	54.00	3.00	2.00	29.00	0.00	1475.50	0.81	1.19
75%	74.00	4.00	3.00	30.00	0.50	1579.25	0.84	1.28
Maximum	96.00	5.00	5.00	30.00	2.00	1992.00	0.89	1.56

Table 3.
statistical description

identifying outliers and uncover anomalies and to determine optimum settings (Khan & Zubair, 2020a, b, c, d). Before ML modeling, EDA was performed on the MRI data. It was executed to gain insights and the strong and weak correlation among the various MRI features.

Outliers are observation points that are completely distant from other observations. Generally, a box-whisker plot is used as a visualization tool to locate the outliers in the data. It shows the spread of quantitative data, which helps in making comparisons among attributes. Thus, we can infer from Figure 1 that Educ, SES, MMSE, CDR, eTIV and nWBV feature columns consist of outliers.

The linearity of the feature variables is determined by a density plot, which displays the distribution of features. In this study, the density plot was used to study the skewness of the various dependent variables. From the density plot, as given in Figure 2, we can note that eTIV, nWBV and ASF have an almost normal distribution. While the features, Educ, SES and MMSE have multimodal distribution.

To construct an effective ML model, an indispensable condition is to remove the correlated features. The correlation matrix using the heat map has been displayed in Figure 3. This is a multivariate plot, which indicates whether or not any kind of dependency and correlation exists amongst various features in the dataset. The heat map shows that eTIV and ASF have
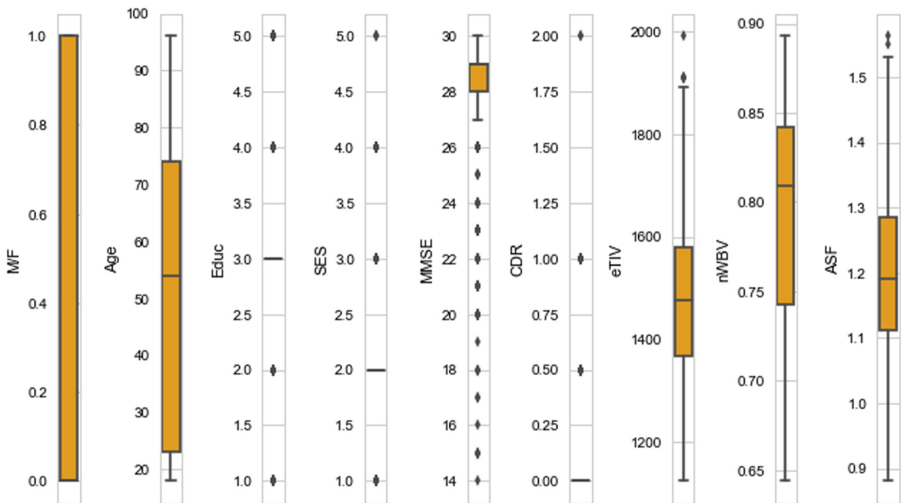

Figure 1.
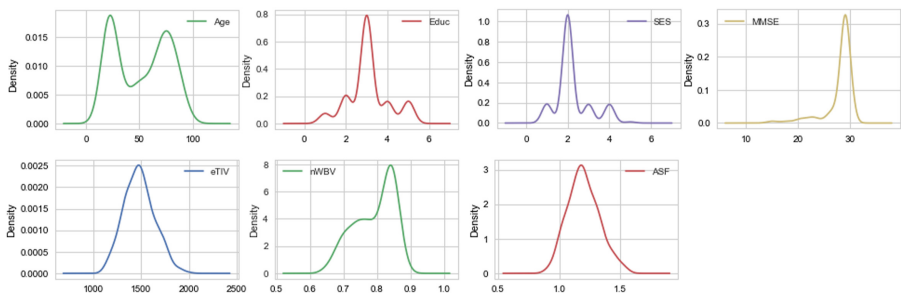Box-Whisker plot showing the outliers

Figure 2.
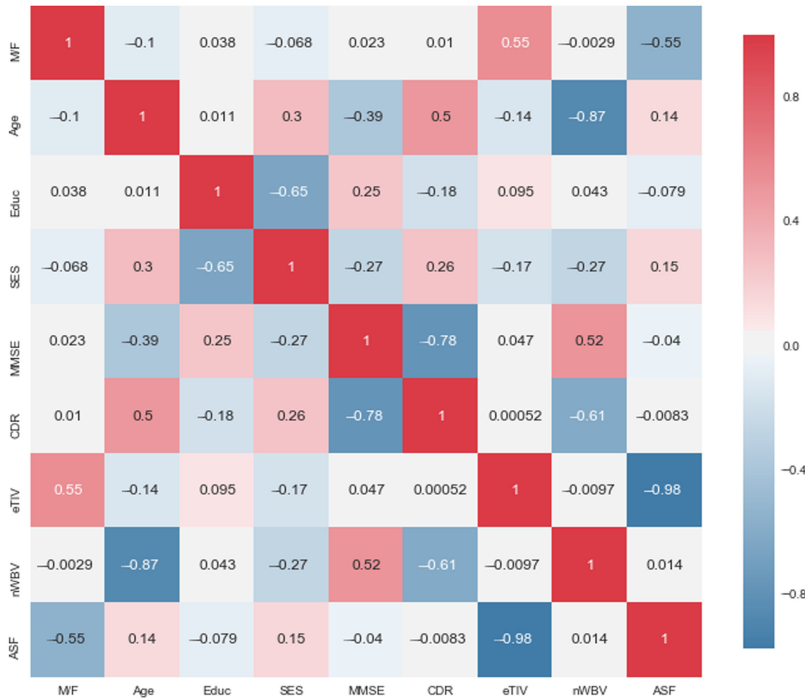Density plot showing the skewness of dependent variables

a strong negative correlation. However, a strong positive correlation can be seen between gender and eTIV.

4. Results
4.1 Comparison of CDR with other variable features
In the present study, we contrasted various independent MRI features with CDRs (target variable). Two types of plots, namely violin and KDE, were employed in the comparative analysis. Table 4 presents the comparison result for the plots depicted in Figure 4a–h.

Among various attributes, the features such as gender, age, education, SES, MMSE, eTIV, nWBV and ASF were found to be considerably related with that of CDR and helpful in the prognosis of dementia. Our analysis shows that subjects aged between 70 and 80 years display a high clustering of dementia than nondemented subjects.

4.2 Comparison of functioning of various as-proposed ML models in the study of dementia progression
We employed ML algorithms to approximate the aggregate CDR score from a pool of dataset features. In a given pool, we have presented the results of 20 selected ML classifiers. The classifiers with mean test accuracy above 50% were included in the study. Table 5 presents the accuracies of training, cross-validation and testing, their precision, recall and F1 values. In this study, the mean prediction accuracies ranged from 52 to 94% (Table 5).

The comparative analysis of the tested predicted models is illustrated in Figure 5. The classification accuracies calculated in Table 5 and its corresponding comparison chart shown

| | | CDR | | | |
		Normal (0.0)	Very mild (0.5)	Moderate (1.0)	Severe (2.0)
a	Gender and CDR	A higher concentration of female than male	A higher concentration of female than male	A higher concentration of female than male	An equal number of female and male
b	Age by CDR	Range from 18 to 96, with a higher concentration of 25-65 years old	Range from 60 to 100, with a higher concentration of 70-80 years old	Range from 60 to 100, with a higher concentration of 75-80 years old	Range from 70 to 85. (It contains only 2 values out of 436)
c	Education by CDR	Middle educated	High educated	In between middle and high educated	Low educated
d	SES by CDR	High status	Vary between high-low status (concentrating towards high status)	Vary between high-low status (concentrating towards low status)	Low status
e	MMSE by CDR	Higher concentration between 27 and 29	Higher concentration between 19 and 24	Higher concentration between 10 and 19	A score of 15 for both of two values
f	eTIV by CDR	Higher concentration between 1300 and 1600	Higher concentration between 1300 and 1500	Higher concentration between 1400 and 1500	Higher concentration between 1300-1400 and 1500-1600
g	nWBV by CDR	Higher concentration between 0.8 and 0.9	Higher concentration between 0.7 and 0.8	Higher concentration between 0.65 and 0.75	Higher concentration between 0.65 and 0.69 and 0.7-0.075
h	ASF by CDR	Higher concentration between 10 and 13	Higher concentration between 11.5 and 12.5	Higher concentration between 11 and 13	Higher concentration between 11-12 and 12-14

Table 4.
Effect of various factors on CDR in demented/nondemented subjects

in Figure 5 emerge to be reliable. For the AD diagnosis, bagging with random forest produced a high accuracy of 93.67%, while BernoulliNB produced the lowest accuracy of 52.29%. In terms of precision, bagging had a higher value of 90.00%, indicating that it is a better diagnostic predictor. Table 5 shows that bagging and random forest have the highest recall, both with 88.00%. The F1 score of 0.89, which is close to one for both bagging and random forest, indicates that these ML models were a better diagnostic predictor for AD.

It can be attributed to the precise classification of the mild and moderate demented subjects (CDR = 0.5 and CDR = 1.0). After dealing with the imbalanced classification using SMOTE analysis and repeated stratified cross-validation, we obtained an improved performance and hence an overall increase in the prediction results ML modeling for all the 20 classifiers. We can see from the results that the cross-validation value accurately displays the actual performance that we can envisage from our model in the real operation.

4.3 Bagging ensemble with random forest as a prediction model for dementia
The bagging classifier shows the highest prediction accuracy amongst varied employed classifiers (Table 5 and Figure 5). This is because the Bagging approach improves a single estimate by aggregating multiple estimates. It builds *n* classification trees from train data using bootstrap sampling and then integrates their predictions to give a final metaprediction.

When compared to other nineteen ML models, this approach demonstrated improvements in terms of accuracy, precision, recall and F1 score. We applied five ML classifiers, namely Gradient Boosting, Random Forest, Decision Tree, LDA and Extra Trees, as a base estimator
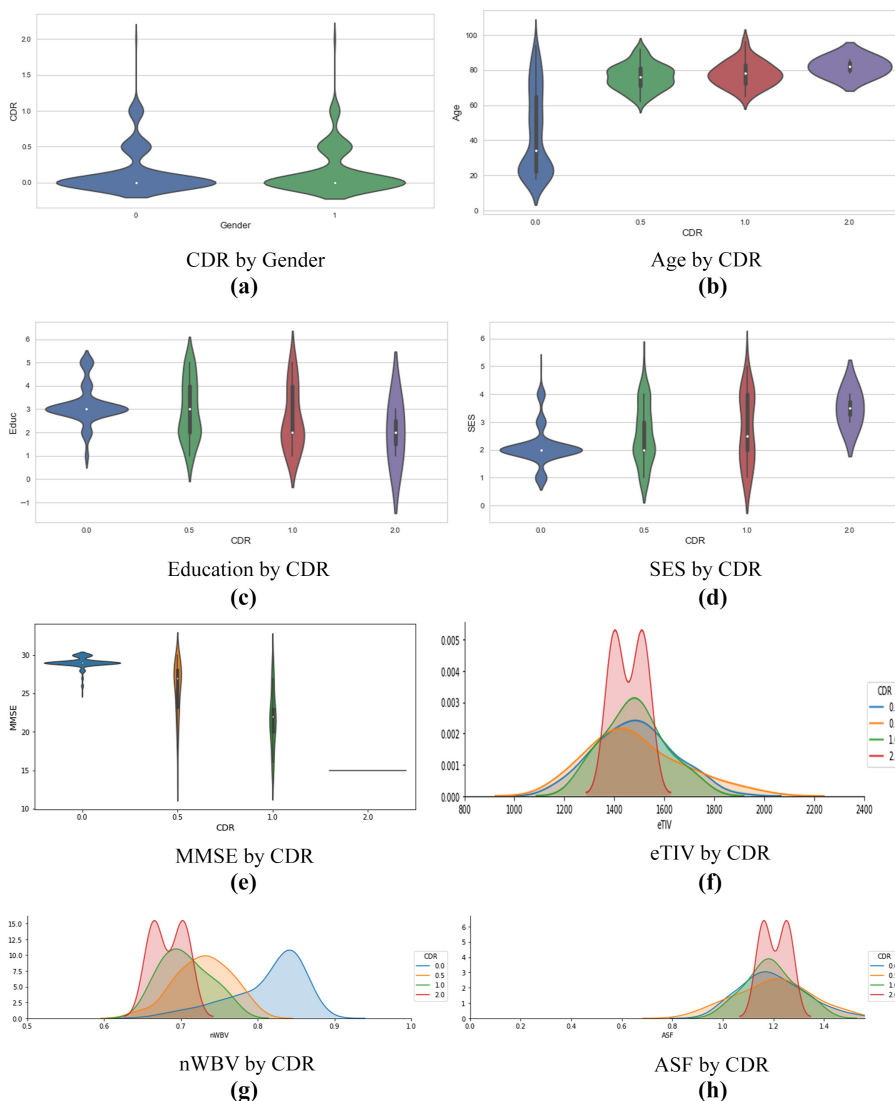
CDR by Gender
(a)

Age by CDR
(b)

Education by CDR
(c)

SES by CDR
(d)

MMSE by CDR
(e)

eTIV by CDR
(f)

nWBV by CDR
(g)

ASF by CDR
(h)

for training a Bagging classifier. Amongst these, the improved performance was given by the
Random Forest after hyperparameter tuning and varying the number of estimators. The
mean cross-validation accuracy was 91.45%, mean prediction accuracy was around 93.67%,
with 90.0% mean precision, 88.0% mean recall and 0.89 F1 score. In general, the precision
correlates with the accuracy of the given ML model. It computes actual positive out of the
positive predicted cases employed in the study. While recall computes how many of the actual
positives that are captured by the ML model are true positive. Furthermore, the F1 measure is
the weighted average (harmonic mean) of both the precision and recall. Thus, in this study,
Bagging approach used with Random Forest considerably enhanced model stability by

ML classifier	Train accuracy (%)	Cross-validation accuracy (%)	Prediction accuracy (%)	Precision (%)	Recall (%)	F1 measure
Generalized linear models						
1 Logistic RegressionCV	84.71	84.26	80.73	84.0	81.0	0.82
2 Passive Aggressive	77.37	84.26	78.90	86.0	79.0	0.81
3 Perceptron	82.87	84.83	79.82	70.0	80.0	0.74
4 Ridge ClassifierCV	77.06	84.83	79.82	86.0	80.0	0.82
5 Stochastic Gradient Descent (SGD)	74.01	83.29	75.23	69.0	75.0	0.72
6 XGB	100.0	84.83	86.24	88.0	86.0	0.87
Naïve Bayes models						
7 BernoulliNB	55.96	83.29	52.29	60.0	52.0	0.55
8 GaussianNB	77.98	83.29	79.82	88.0	80.0	0.82
Support vector machine models						
9 LinearSVC	80.73	83.29	81.65	79.0	82.0	0.77
10 SVC	85.32	84.83	78.90	87.0	79.0	0.81
Discriminant analysis models						
11 LDA	77.37	84.83	78.90	88.0	79.0	0.81
12 QDA	84.40	83.29	79.82	86.0	80.0	0.81
Neighbor-based model						
13 KNN	86.85	84.78	76.15	85.0	76.0	0.79
Tree-based model						
14 Decision Tree	100.0	84.83	84.40	86.0	84.0	0.85
Ensemble-based models						
15 AdaBoost	100.0	84.83	85.32	86.0	85.0	0.86
16 Bagging	*94.20*	*91.45*	*93.67*	*90.0*	*88.0*	*0.89*
17 Extra Trees	100.0	84.83	85.32	88.0	85.0	0.86
18 Gradient Boosting	99.70	83.65	86.24	88.0	86.0	0.87
19 Random Forest	97.0	86.76	88.07	90.0	88.0	0.89
Neural network						
20 MLP	85.02	84.83	82.57	86.0	83.0	0.84

Table 5.
Machine learning analysis for dementia prognosis

improving accuracy and, as a result, reducing variance, thereby eliminating the issue of overfitting.

Figure 6 shows the detailed distribution of precision and recall measures of Bagging classifier for multinomial classification of CDR dementia into normal, very mild and moderate classes. These measures set a base for the computation of overall class metrics.

It can be seen from Figure 6 that all three of the CDR classes possess clinically acceptable accuracies. To make the classification results more meaningful (clinically), the two subjects in the dataset with severe dementia (CDR = 2.0) were merged with moderate dementia subjects. This leads to a total of only three different CDR classes.

The precision value for the normal class was estimated to be 0.9867%. Therefore, about 98% of the subjects predicted as normal were normal in reality. On the other hand, the recall
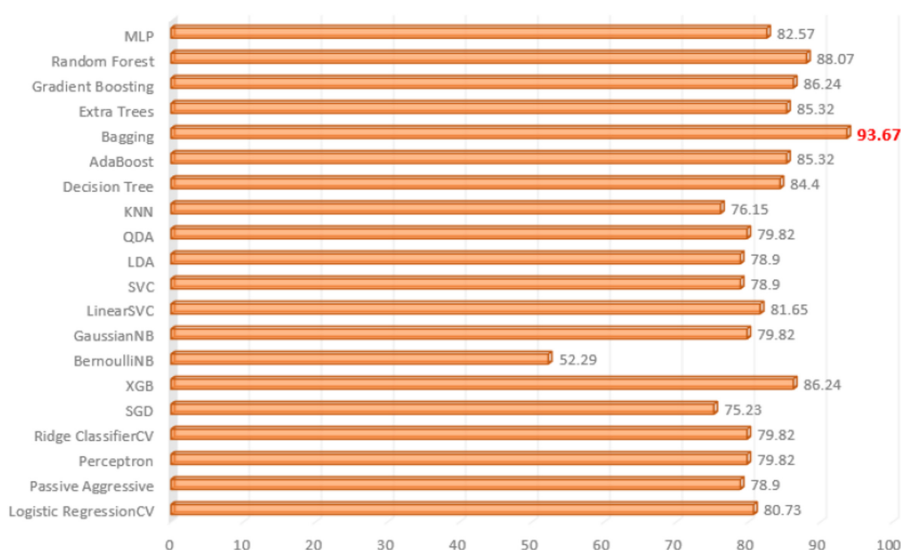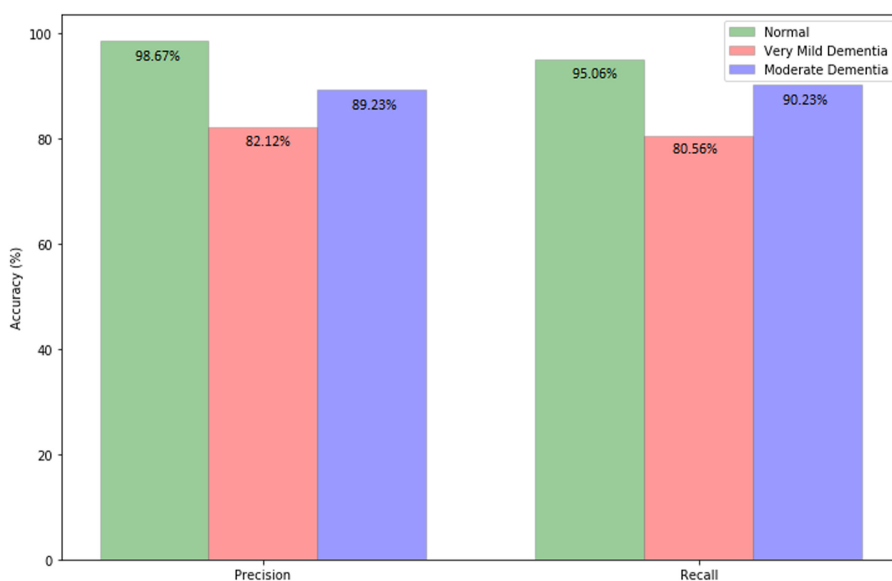
value for the normal class came out to be 0.9506. This means that almost 95% of the normal subjects in the dataset were correctly predicted as normal. Similarly, the precision and recall value was calculated for the other two classes: very mild dementia and moderate dementia. For very mild dementia class, the numbers are 82.12% and 80.56%, respectively. While for moderate dementia class, the numbers for precision and recall are 89.23 and 90.23%. Thus, the mean precision and mean recall of the ML model comes out to be 90.0% 88.62%. In general, a high precision value signifies that a classifier gave more significant results than the insignificant ones. While high recall value means that a classifier returned most of the

significant results. This classification structure was however useful since the essential clinical distinctions are amongst normal aging, very mild and moderate dementia. Hence, the classification accuracies are relatively acceptable. The clinical adjustments which cannot employ any computer-based algorithm fail to apply the results achieved from ML classifiers other than bagging classifier or any tree or decision-rule based classifier. This ML model with a bagging classifier can classify with a subset of the features characterized in the rule set for each patient. Thus, making this characteristic clinically strong and that the generated ML model as practically stable.

5. Discussion

The CDR Scale is a coherent and valid assessment feature that has been effectively used in several dementia studies across the globe. The CDR-based scale exhibits notable interrater reliability. It remained extremely correlated with other performance cognitive measures, like the Abbreviated Mental Test, MMSE and comprehensive psychometric tests (Lim, Chin, Lam, Lim, & Sahadevan, 2005; Otoyama *et al.*, 2000).

In the present study, we applied the ML approach on a cross-sectional MRI-based data to examine the usefulness of the CDR Scale in the prediction of dementia progression. We compared its predictive power with other cognitive and functional measures. In addition to this, we used the demographic and neuropsychological data to predict and classify the group of demented and nondemented patients based on the CDR Scale. We observed that 100 (~23%) out of 436 individuals progressed to dementia within the follow-up interval. When compared with the two-year follow-up amnestic MCI subjects' data from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, our results were in line to ADNI analysis (i.e. 35%) (Gomar, Bascaran, Goldberg, Davies, & Goldberg, 2011).

The computational analysis indicated that CDR alone as a testing variable (dependent variable) could predict dementia progression with improved accuracy rate. These results were consistent with earlier reported studies (Nakata *et al.*, 2009; Daly *et al.*, 2000). Moreover, the classification accuracy with CDR orientation had better prediction rate of 93.67%. The comparison of CDR subscale with that of cognitive and functional measures generated in the present dataset showed an effective relationship with that of CDR. Also, as dementia disease has a low longevity, this is why data are scarcely available in the aged patient (Barrett & Burns, 2014; Prince *et al.*, 2014). This adversely affects the ML modeling and hence assessment of the disease. The findings of our study also displayed effective clinical power, concerning the evaluation of clinical dementia using the visualization and for determining the functional and cognitive impairment severity. The interrater reliability was lowest in the domain categories of judgment and problem-solving and community affairs (0.77 and 0.79, respectively) (Lim *et al.*, 2007). These domains test higher cognitive function and necessitate greater judgmental skills on the part of the assessor (Lim *et al.*, 2007).

We discovered that the background history information reveals that about 60% of the subjects were middle-educated while only 23% were found to be highly educated. In turn, this may perhaps reduce generalizability in several ways. Certain individuals might involve in complex activities that create small cognitive alteration much more apparent. The educational background of the subjects is reflected in neuropsychological performance in spite of the existence of cognitive alterations in complex activities of day-to-day life.

We found that the change from high socio-economic status (Score 1) to low status (Score 5) ensued in a substantial decrease in the occurrence of dementia, i.e. from 77 to 0.68%. Specifically, education and SES are related to each other with a greater extent. Higher to moderate correlation level has been found amongst education and occupation-based SES (Karp *et al.*, 2004). Numerous modules of SES, such as income, education and working status, might affect Alzheimer's development in elderly patients (Evans *et al.*, 1997).

Folstein test for MMSE is a comprehensive analysis measure of cognitive impairment, which has been used widely in dementia detection. Arevalo-Rodriguez *et al.* reported the MMSE accuracy for Alzheimer's detection in mild MCI subjects (Arevalo-Rodriguez *et al.*, 2015). They discovered that the MMSE score is not an essential tool in categorizing demented or nondemented people (Cacchione *et al.*, 2003). On the same line, we discovered that the nondemented group acquired a higher score than the demented group.

The CDR Scale usefulness has been found to be in part correlated with the CDR assessment process (Morris, 1993). The CDR evaluation depends on concise cognitive assessment and the information acquired at the time of interview with collateral informant concerning functional and cognitive changes compared to the earlier usual level (Hughes *et al.*, 1982; Kim *et al.*, 2017). Because of this evaluation process, the CDR scores are not much affected by factors like age, depression, education and practice effect, but these may affect cognitive test scores (Morris, 1993; Kim *et al.*, 2017; Shaji, Sivakumar, Rao, & Paul, 2018). Several studies in line to this suggest that factors, like age, education, informant's mental well-being, relationship type with the patient, informant–patient relationship quality, common or different domicile, frequency of contact, aid in predicting the quality of collateral information source (Cacchione *et al.*, 2003). Thus, the ML classifier and hence its entire modeling increases the fraction of cases, which can be evaluated for severity of dementia in usual community settings, at the same time, reducing the expenditure and time needed for acquiring this information also.

Once the CDR Scale is categorized into normal, very mild, moderate and severe groups, the classification accuracies obtained from ML modeling comes out to analogous to that achieved from clinical experts employing a broad interview process. In many of the clinical settings, this kind of classification method is satisfactorily acceptable. In fact, it builds the essential distinctions amongst normal and demented patients and amongst mild and moderate to severe dementia (Shaji *et al.*, 2018). Data reduction, a technique provided by ML, if not done, it is more likely that medical experts would approximate dementia severity lower than the 80% interrater consistency achieved by specialists (Shankle, Mani, Dick, & Pazzani, 1998). Thus, the ML classifiers can play an essential role in the practical evaluation of the severity of dementia in normal community settings.

In concordance with the previous studies, we found that ML algorithms can enhance the clinical procedural guidelines and can restructure expenses of healthcare (Kim *et al.*, 2017; Ahmed, Mohamed, Zeeshan, & Dong, 2020). A study performed by Battineni *et al.* reported to have a 98.6% accuracy using the hybrid modeling approach (Battineni *et al.*, 2020a, b). The study was based on OASIS longitudinal dataset, and it did not consider CDR as a target variable (Battineni *et al.*, 2020a, b). Although they reached to attain high accuracy while building a predictive model, we cannot compare our results solely on the basis of accuracy (Battineni *et al.*, 2020a, b). In a similar scenario, Khan & Zubair achieved an accuracy of 87% on Random Forest classifier employing the LDA as a dimensionality reduction technique (Khan & Zubair, 2020a, b, c, d). They performed the analysis on the same OASIS cross-sectional data, but the approach of predictive modeling was different (Khan & Zubair, 2020a, b, c, d). Also, the capability of these algorithms is often constrained by inadequate patient numbers. Thus, this problem emphasized the significance of creating simple and structured data gathering methods so that the real potential of ML for rapid decision-support is not retained back because of scarce data.

Our results show the advantage of employing the CDR Scale as a dementia prognosis tool in clinical settings. Previous studies have also established that the CDR score has been broadly used as a principle standard in multi-center clinical AD-related tests (Morris *et al.*, 1998; Sano *et al.*, 1997), and its inter-rater consistency has also been established (Burke *et al.*, 1998; Schafer *et al.*, 2004; Khan & Zubair, 2020a, b, c, d).

5.1 Limitations
We were able to access a limited number of demented cases in this study due to the small number of total subjects (<500). The sample size was limited to demonstrate statistical validity between mild and moderate to severe dementia groups. This was a significant limitation of the study presented.

6. Conclusion
We studied the use of demographic and neuropsychological data in the prediction of CDR cognitive assessment scale by training, cross-validating and testing ML model. We employed 20 ML classifiers to differentiate subjects between normal, very mild and moderately impaired with dementia. Our findings showed Bagging classifier (with Random Forest algorithm as a base estimator) to provide the highest accuracy amongst all 20 applied ML classifiers. Also, in terms of other performance metrics, i.e. precision, recall and F1 measures, the Bagging classifier resulted in improved performance. With proper and sequential data preprocessing techniques, training, cross-validating and testing approach, dealing with the imbalanced classification and further tweaking with the right set of hyperparameters, our ML modeling gave a higher accuracy as a result. We employed SMOTE analysis for handling imbalanced classification and repeated stratified cross-validation methodology. Through this study, we were able to determine that the CDR Scale may provide beneficial information to predict dementia in the individuals in actual clinical settings. The experimental analysis reported in this study implies that artificial intelligence can be used effectively, in particular, to computerize the variables of clinical diagnosis. This can further aid in providing significant patterns and insights, where features are indispensable for such identification. In the future, hybrid modeling would be used in research. In a moderate clinical setting, an analytical approach like this will be useful in detecting the prognosis of suspected or diagnosed dementia or AD patients. Thus, with proper training of medical practitioners, this can be employed in the community follow-up of afflicted patients.

References

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, *2020*, baaa010. doi:10.1093/database/baaa010.

Alanazi, A. (2022). Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked*, *30*. doi: 10.1016/j.imu.2022.100924.

Amancio, D. R., Comin, C. H., Casanova, D., Travieso, G., Bruno, O. M., Rodrigues, F. A., & Costa, L. F. (2014). A systematic comparison of supervised classifiers. *PLoS One*, *9*(4), 1–13.

Amoroso, N., Rocca, M. L., Bruno, S., Maggipinto, T., Monaco, A., Bellotti, R., & Tangaro, S. (2017). Brain structural connectivity atrophy in Alzheimer's disease. *arXiv*:1709.02369 *[physics.med-ph]*, 7(7), 1-16. doi:10.48550/arXiv.1709.02369.

Arevalo-Rodriguez, I., Smailagic, N., Figuls, M. R. I., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., . . . & Cullum, S. (2015). Mini-mental state examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD010783.pub2.PMID: 25740785.

Balsis, S., Miller, T. M., Benge, J. F., & Doody, R. S. (2011). Dementia staging across three different methods. *Dementia and Geriatric Cognitive Disorders*, *31*(5), 328–333.

Barrett, E., & Burns, A. (2014). Dementia revealed: what primary care needs to know. Available from: https://www.readkong.com/page/dementia-revealed-what-primary-care-needs-to-know-6196949 (accessed 28 October 2021).

Battineni, G., Chintalapudi, N., & Amenta, F. (2019). Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, 16, 1–8. doi: https://doi.org/10.1016/j.imu.2019.100200.

Battineni, G., Sagaro, G. G., Chinatalapudi, N., & Amenta, F. (2020a). Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of Personalized Medicine*, 10(2), 21, 1-11. doi: 10.3390/jpm10020021.

Battineni, G., Chintalapudi, N., Amenta, F., & Traini, E. (2020b). A comprehensive machine-learning model applied to magnetic resonance imaging (MRI) to predict Alzheimer's disease (AD) in older subjects. *Journal of Clinical Medicine*, 9(7), 2146, 1-14. doi:10.3390/jcm9072146.

Burke, W. J., Miller, P., Rubin, E. H., Morris, J. C., Coben, L. A., Duchek, J., Wittels, I. G., & Berg, L. (1998). Reliability of the Washington university clinical dementia rating. *Archives of Neurology*, 45, 31–32.

Cacchione, P. Z., Powlishta, K. K., Grant, E. A., Buckles, V. D., & Morris, J. C. (2003). Accuracy of collateral source reports in very mild to mild dementia of the Alzheimer type. *Journal of the American Geriatrics Society*, 51, 819–823.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence*, 16, 321–357.

Coley, N., Andrieu, S., Jaros, M., Weiner, M., Cedarbaum, J., & Vellas, B. (2011). Suitability of the clinical dementia rating-sum of boxes as a single primary endpoint for Alzheimer's disease trials. *Alzheimers Dement*, 7(6), 602–610, e2. doi: 10.1016/j.jalz.2011.01.005.PMID: 21745761.

Crammer, K., Dekel, O., Keshet, J., Shwartz, S. S., & Singer, Y. (2006). Online passive-aggressive algorithms. *Journal of Machine Learning Research*, 7, 551–585.

Daly, E., Zaitchik, D., Copeland, M., Schmahmann, J., Gunther, J., & Albert, M. S. (2000). Predicting conversion to Alzheimer disease using standardized clinical information. *Archives of Neurology*, 57(5), 675–680.

Deng, L., & Yu, D. (2014). Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(3-4), 197–387. doi: 10.1561/2000000039.

Denisko, D., & Hoffman, M. M. (2018). Classification and interaction in random forests. *Proceedings of the National Academy of Sciences of the United States of America*, 115(8), 1690–1692.

Emanuele, B., Alessandro, T., Victòria, B. R., Federica, D. A., & Lavinia, A. (2021). Intercepting dementia: Awareness and innovation as key tools. *Frontiers in Aging Neuroscience*, 13. doi: 10.3389/fnagi.2021.730727.

Evans, D. A., Hebert, L. E., Beckett, L. A., Scherr, P. A., Albert, M. S., Chown, M. J., . . . & Taylor, J. O. (1997). Education and other measures of socioeconomic status and risk of incident Alzheimer disease in a defined population of older persons. *Archives of Neurology*, 54(11), 1399–1405. doi: 10.1001/archneur.1997.00550230066019.PMID:9362989.

Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J. (2008). Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, 9, 1871–1874.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63, 3–42.

Gomar, J. J., Bascaran, M. T. B., Goldberg, C. C., Davies, P., & Goldberg, T. E. (2011). Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. *Archives of General Psychiatry*, 68(9), 961–969.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140(6), 566–572.

Jayatilake, S. M. D. A. C., & Ganegoda, G. U. (2021). Involvement of machine learning tools in healthcare decision making. *Journal of Healthcare Engineering*, 2021, 6679512. doi:10.1155/2021/6679512.

Karp, A., Kåreholt, I., Qiu, C., Bellander, T., Winblad, B., & Fratiglioni, L. (2004). Relation of education and occupation-based socioeconomic status to incident Alzheimer's disease. *American Journal of Epidemiology*, 159(2), 175–183. doi: 10.1093/aje/kwh018.PMID:14718220.

Khan, A., & Zubair, S. (2018). Machine learning tools and toolkits in the exploration of big data. *International Journal of Computational Science and Engineering*, 6(12), 570–575.

Khan, A., & Zubair, S. (2019). Usage of random forest ensemble classifier based imputation and its potential in the diagnosis of Alzheimer's disease. *International Journal of Scientific & Technology Research*, 8(12), 271–275.

Khan, A., & Zubair, S. (2020a). An improved multi-modal based machine learning approach for the prognosis of Alzheimer's disease. *Journal of King Saud University - Computer and Information Sciences*, 34(6, Part A), June 2022, 2688-2706. doi: 10.1016/j.jksuci.2020.04.004.

Khan, A., & Zubair, S. (2020b). Expansion of regularized kmeans discretization machine learning approach in prognosis of dementia progression. *11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, (pp. 1–6), Kharagpur. doi: 10.1109/ICCCNT49239.2020.9225397.

Khan, A., & Zubair, S. (2020c). Longitudinal magnetic resonance imaging as a potential correlate in the diagnosis of Alzheimer disease: exploratory data analysis. *JMIR Biomedical Engineering*, 5(1), 1–13.

Khan, A., & Zubair, S. (2020d). A machine learning-based robust approach to identify dementia progression employing dimensionality reduction in cross-sectional MRI data. *First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, (pp. 237–242), Riyadh.

Khan, A., Zubair, S., & Sabri, M. A. (2019). An improved pre-processing machine learning approach for cross-sectional MR imaging of demented older adults. In *2019 First International Conference of Intelligent Computing and Engineering (ICOICE)*, (pp. 1–7).

Kim, J. W., Byun, M. S., Sohn, B. K., Yi, D., Seo, E. H., Choe, Y. M., . . . & Lee, D. Y. (2017). Clinical dementia rating orientation score as an excellent predictor of the progression to Alzheimer's disease in mild cognitive impairment. *Psychiatry Investigation*, 14(4), 420-426. doi: 10.4306/pi.2017.14.4.420.

Lim, W. S., Chin, J. J., Lam, C. K., Lim, P. P., & Sahadevan, S. (2005). Clinical dementia rating: Experience of a multi-racial Asian population. *Alzheimer Disease and Associated Disorders*, 19(3), 135–142. doi: 10.1097/01.wad.0000174991.60709.36.PMID:16118530.

Lim, W. S., Chong, M. S., & Sahadevan, S. (2007). Utility of the clinical dementia rating in asian populations. *Clinical Medical and Research*, 5(1), 61–70.

Linz, N., Troger, J., Alexandersson, J., Konig, A., Robert, P., & Wolters, M. (2017). Predicting dementia screening and staging scores from semantic verbal fluency performance. *ICDM 2017 – IEEE International Conference on Data Mining, Workshop on Data Mining for Aging, Rehabilitation and Independent Assisted Living*, (pp. 719–728), New Orleans, Nov 2017. doi: 10.1109/ICDMW.2017.100.hal-01672590.

Lynch, J. W., & Kaplan, G. A. (2000). Socioeconomic position. In Berkman, L. F. & Kawachi, I. (Eds.), *Social epidemiology*, (pp. 13–35). New York: Oxford University Press.

Magni, E., Binetti, G., Padovani, A., Cappa, S. F., Bianchetti, A., & Trabucchi, M. (1996). The mini-mental state examination in Alzheimer's disease and multi-infarct dementia. *International Psychogeriatrics*, 8(1), 127–134. doi: 10.1017/s1041610296002529.PMID:8805093.

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9), 1498–1507. doi: 10.1162/jocn.2007.19.9.1498.PMID:17714011.

Morris, J. C. (1993). The clinical dementia rating (CDR): Current version and scoring rules. *Neurology*, 43(11), 2412–2414.

Morris, J. C., Ernesto, C., Schafer, K., Coats, M., Leon, S., Sano, M., . . . & Woodbury, P. (1997). Clinical dementia rating training and reliability in multicenter studies: The Alzheimer's disease

cooperative study experience. *Neurology*, *48*(6), 1508–1510. doi: 10.1212/wnl.48.6.1508.PMID: 9191756.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., Belle, G. V., Fillenbaum, G., . . . & Clark, C. (1998). The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, *39*, 1159–1165.

Morris, J. C., Storandt, M., Miller, J. P., McKeel, D. W., Price, J. L., Rubin, E. H., & Berg, L. (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology*, *58*(3), 397–405. doi: 10.1001/archneur.58.3.397.PMID:11255443.

Nakata, E., Kasai, M., Kasuya, M., Akanuma, K., Meguro, M., Ishii, H., . . . & Meguro, K. (2009). Combined memory and executive function tests can screen mild cognitive impairment and converters to dementia in a community: The Osaki-Tajiri project. *Neuroepidemiology*, *33*(2), 103–110.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*, 1–21.

NIST/SEMATECH. (2003). NIST/SEMATECH e-handbook of statistical methods. Available from: http://www.itl.nist.gov/div898/handbook/.

Nori, V. S., Hane, C. A., Martin, D. C., Kravetz, A. D., & Sanghavi, M. (2019). Identifying incident dementia by applying machine learning to a very large administrative claims dataset. *PLoS One*, *14*(7), 1–15. doi:10.1371/journal.pone.0203246.

O'Bryant, S. E., Lacritz, L. H., Hall, J., Waring, S. C., Chan, W., Khodr, Z. G., . . . & Cullum, C. M. (2010). Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. *Archives of Neurology*, *67*(6), 746–749.

Onan, A. (2018a). An ensemble scheme based on language function analysis and feature engineering for text genre classification. *Journal of Information Science*, *44*(1), 28–47. doi: 10.1177/0165551516677911.

Onan, A. (2018b). Biomedical text categorization based on ensemble pruning and optimized topic modelling. *Computational and Mathematical Methods in Medicine*, *2018*, 22, 2497471. doi: 10.1155/2018/2497471.

Onan, A. (2019). Consensus clustering-based undersampling approach to imbalanced learning. *Scientific Programming*, *2019*, 5901087. doi:10.1155/2019/5901087.

Onan, A. (2020). Mining opinions from instructor evaluation reviews: A deep learning approach. *Computer Applications in Engineering Education*, *28*, 117–138. doi: 10.1002/cae.22179.

Onan, A. (2022). Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. *Journal of King Saud University – Computer and Information Sciences*, *34*(5), 2098–2117. doi: 10.1016/j.jksuci.2022.02.025.

Onan, A., & Korukoğlu, S. (2017). A feature selection model based on genetic rank aggregation for text sentiment classification. *Journal of Information Science*, *43*(1), 25–38. doi: 10.1177/0165551515613226.

Onan, A., & Toçoğlu, M. A. (2021). A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. *IEEE Access*, *9*, 7701–7722. doi: 10.1109/ACCESS.2021.3049734.

Onan, A., Korukoğlu, S., & Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. *Expert Systems with Applications*, *57*, 232–247. doi: 10.1016/j.eswa.2016.03.045.

Onan, A., Korukoğlu, S., & Bulut, H. (2017). A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. *Information Processing and Management*, *53*(4), 814–833. doi: 10.1016/j.ipm.2017.02.008.

Otoyama, W., Niina, R., Homma, A., Sanada, J., Takahashi, M., Kamimura, N., . . . & Takeuchi, H. (2000). Inter-rater reliability of the Japanese version of CDR (in Japanese). *Journal of Geriatric Psychiatry*, *11*, 521–527.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Porsteinsson, A. P., Isaacson, R., Knox, S., Sabbagh, M. N., & Rubino, I. (2021). Diagnosis of early Alzheimer's disease: Clinical practice in 2021. *Journal of Prevention of Alzheimer' Disease*, *8*, 371–386. doi:10.14283/jpad.2021.23.

Prince, M., Albanese, E., Guerchet, M., & Prina, M. (2014). World Alzheimer report 2014 – Dementia and risk reduction: An analysis of protective and modifiable factors. Available from: https://www.alzint.org/u/WorldAlzheimerReport2014.pdf.

Prince, M., Herrera, A. C., Knapp, M., Guerchet, M., & Karagiannidou, M. (2016). World alzheimer report 2016 – Improving healthcare for people living with dementia: Coverage, quality and costs now and in the future. Available from: https://www.alzint.org/u/WorldAlzheimerReport2016.pdf.

Ranson, J. M., Rittman, T., Hayat, S., Brayne, C., Jessen, F., Blennow, K., . . . & European Task Force for Brain Health Services. (2021). Modifiable risk factors for dementia and dementia risk profiling. A user manual for Brain Health Services – Part 2 of 6. *Alzheimer's Research and Therapy*, *13*. doi: 10.1186/s13195-021-00895-4.

Ribeiro, P. C. C., Lopes, C. D. S., & Lourenço, R. A. (2013). Prevalence of dementia in elderly clients of a private health care plan: A study of the FIBRA-RJ, Brazil. *Dementia and Geriatric Cognitive Disorders*, *35*, 77–86.

Robinson, L., Tang, E., & Taylor, J. P. (2015). Dementia: Timely diagnosis and early intervention. *BMJ*, *350*, 1–6.

Sano, M., Ernesto, C., Thomas, R. G., Klauber, M. R., Schafer, K., Grundman, M., . . . & Thal, L. J. (1997). A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *New England Journal of Medicine*, *336*, 1216–1222.

Schafer, K. A., Tractenberg, R. E., Sano, M., Mackell, J. A., Thomas, R. G., Gamst, A., . . . & Morris, J. C. (2004). Reliability of monitoring the clinical dementia rating in multicenter clinical trials kimberly. *Alzheimer Disease and Associated Disorders*, *18*(4), 219–222.

Shaji, K. S., Sivakumar, P. T., Rao, G. P., & Paul, N. (2018). Clinical practice guidelines for management of dementia. *Indian Journal of Psychiatry*, *60*, S312–S328.

Shankle, W. R., Mani, S., Dick, M. B., & Pazzani, M. J. (1998). Simple models for estimating dementia severity using machine learning. *MEDINFO'98*. IOS Press, pp. 472-476. Available from https://pubmed.ncbi.nlm.nih.gov/10384501/.

Sheehan, B. (2012). Assessment scales in dementia. *Therapeutic Advances in Neurological Disorder*, *5*(6), 349–358. doi:10.1177/1756285612455733.

Wessels, A. M., Dowsett, S. A., & Sims, J. R. (2018). Detecting treatment group differences in Alzheimer's disease clinical trials: A comparison of Alzheimer's disease assessment scale – Cognitive subscale (ADAS-Cog) and the clinical dementia rating – Sum of boxes (CDR-SB). *Journal of Prevention of Alzheimer's Disease – JPAD*, *5*(1), 15–20.

Wimo, A., Winbald, B., Torres, H. A., & Strauss, E. V. (2003). The magnitude of dementia occurrence in the world. *Alzheimer Disease and Associated Disorders*, *17*, 63–67.

Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, *4*(11), 1–7.

Corresponding author

Swaleha Zubair can be contacted at: swalehazubair@yahoo.com