



# Comprehensive Performance Analysis of Neurodegenerative disease Incidence in the Females of 60-96 year Age Group

Afreen Khan<sup>a</sup>, Swaleha Zubair<sup>a</sup> and Samreen Khan<sup>b</sup>

<sup>a</sup> Department of Computer Science, Aligarh Muslim University, Aligarh, India

<sup>b</sup> Department of Community Medicine, Integral Institute of Medical Sciences and Research, Lucknow, India

afreen.khan2k13@gmail.com, swalehazubair@yahoo.com, drsamreen2k4@gmail.com

## KEYWORDS

Alzheimer's disease; dementia, female; machine learning; neurodegenerative disease; performance

## ABSTRACT

Neurodegenerative diseases such as Alzheimer's disease (AD) and dementia are gradually becoming more prevalent chronic diseases, characterized by the decline in cognitive and behavioral symptoms. Machine learning (ML) is revolutionising almost all domains of our life, including the clinical system. The application of ML has the potential to enormously augment the reach of neurodegenerative care thus building it more proficient. Throughout the globe, there is a massive burden of AD and dementia cases; which denotes an exclusive set of difficulties. This provides us with an exceptional opportunity in terms of the impending convenience of data. Harnessing this data using ML tools and techniques, can put scientists and physicians in the lead research position in this area. The objective of this study was to develop an efficient prognostic ML model with high-performance metrics to better identify female candidate subjects at risk of having AD and dementia. This paper portrays our latest contribution to the advancement in neurodegenerative disorders. The study was based on two diverse datasets. The results have been discussed employing seven performance evaluation measures i.e. accuracy, precision, recall, F-measure, Receiver Operating Characteristic area, Kappa statistic, and Root Mean Squared Error. Also, comprehensive performance analysis has been carried out later in the study. The experiment had shown a high accuracy of 98.90 for the AD recognition and 99.60 for the dementia prognosis.



# 1. Introduction

Machine learning (ML), a subset and an application of artificial intelligence is a term defined as the theory and construction of such systems that provide the competence to learn and improve from subjected knowledge without being overtly programmed (Team, 2020). The developed systems can be as easy as rule-based or might be powered by complicated statistical techniques. ML can be one of four kind i.e. supervised, semi-supervised, unsupervised, or reinforcement-based. Significantly, ML tools and techniques are being employed in all scientific domains and thus, are liable for transforming organizations all around the globe (Khan & Zubair, 2018). In contrast, healthcare corporations, have been slow in implementing ML-based systems.

At present, ML is being used in the field of several chronic diseases in myriad ways likewise diagnosis, prognosis, early detection and its management. There is a scarcity of specific data on all these aspects of ML. Specifically, Alzheimer's disease (AD) and dementia are the two neurodegenerative diseases (ND) that are incurable and at the same time extremely challenging to treat. Dementia is an umbrella term for several diseases affecting mental health i.e. memory and related cognitive abilities, the performance of day- to-day tasks, thinking, communication and problem-solving abilities (Medical Research Council, 2019). On the other side, AD affects memory, thought process and language competency.

AD is the commonest type of dementia; it is the 4<sup>th</sup> largest cause of death for people > 65 years of age (Sonkusare et al., 2005). In general, dementia is the root cause for the extreme public burden of disease, where AD represents > 60-70 percent of the cases (Sonkusare et al., 2005). Neurodegenerative diseases are very much linked with age. But, neither of the two are considered a normal part of aging. As many studies have been performed and found that the younger population too have developed symptoms of AD and dementia (Khan & Zubair, 2020a). The severity of ND symptoms is generally evaluated through brain metrics such as the Clinical Dementia Rating (CDR) scale and Mini Mental State Exam (MMSE) (Folstein et al., 1975). Thus, diagnosing the disease at the early stage aids in reducing medical overheads and the risk of patients getting inflicted with more complex health issues.

As of current times, the treatment used for ND is designed at improving both behavioral and cognitive symptoms and thus, aiming at providing a better quality of life for the patients. At the same time, the diverse behavior of AD and dementia causes these diseases hard to diagnose, treat and manage. This thereby leads to demand for improved systems to predict the disorder timely and monitor its progression easily (Cummings et al., 2016). There are tremendous reviews on compiling studies that have used the ML approach in the diagnosis and prognosis of ND. Also, ML techniques have been widely used in ND research to explore the related risk factors. Several applications of ML to ND data have utilized methods from supervised learning to predict definite endpoints (Lasko et al., 2013; Myers et al., 2017). Instead of building separate supervised ML models to predict every single attribute, there is a pressing need to build a comprehensive model that could predict the evolution of numerous attributes simultaneously, that too with higher prediction accuracy.

A number of AD and dementia progression ML models have been built using clinical (Ito et al., 2013; Kennedy et al., 2016; Rogers et al., 2012; Szalkai et al., 2017) and imaging data (Hinrichs et al., 2011; Mueller et al., 2005; Risacher et al., 2009; Suk et al., 2014; Weiner et al., 2013). Even though earlier methodologies of predicting disease advancement have demonstrated to be useful, they have concentrated on forecasting a particular endpoint only (Romero et al., 2015). As both AD and dementia are multifactorial and heterogeneous (Khan & Zubair, 2020a), in

this study we set out to build a model specifically for female candidate patients of AD and dementia progression. This is the major motivation of the performed study. Also, the contribution of this study is two-fold. We present predictive models using several supervised ML techniques to predict the probability of patients having AD and dementia based on the clinical data, which includes demographic, brain metrics and related scores employing two diverse sets of data. Moreover, we also compare these methods later in our study.

This paper is organized as follows. Section 2 describes the ML modeling in detail, including the employed classifiers and performance evaluation metrics, Section 3 includes data description, Section 4 describes the results of the experiments, Section 5 communicates the discussion followed by conclusion and future pursuits in Section 6.

## 2. Theoretical Background

### 2.1. ML Modeling

ML process can be both supervised and unsupervised. The supervised ML operates in those settings where the machine learns the function with known inputs and outputs (Khan & Zubair, 2019a). In the unsupervised ML setup, the machine learns where the outputs are not known. Since ML require supervised tasks, in this study we focused on the classification approach to achieve the correct classification of healthy and inflicted subjects directed towards attaining improved accuracy.

Classification is one of the key domain of ML and has been widely employed for different functions (Ahmed et al., 2016). In particular, the ML model is a mathematical portrayal of a real-life system. When a ML model is built, it consists of several stages. The various stages are data preprocessing, data segregation, modeling, prediction and performance metrics evaluation. The third stage i.e. ML modeling itself consists of four sub-stages viz. model training, model evaluation, cross-validation and model validation (Khan & Zubair, 2020b). This includes training a ML classifier/algorithm that could predict the independent variable (label) amongst the pool of various dependent attributes. This is then further tuned using a combination of hyperparameters and is validated on the test data.

Subsequently, a trained model is obtained from the ML modeling stage, which is further used for interpretation and predictions on test data. The main idea of this phase is to fulfil the necessities of the agenda for creating the model and that could be deployed on factual data in real-life process (Khan & Zubair, 2020c). The key task required in classification is locating the set of class i.e. binary class (consists of only two labels) or multi-class (includes more than two labels). At an initial stage, a classification model is generated, known as classifiers. This defines the relationship amongst attributes and classes (labels). In the next stage, the accuracy of a classifier is evaluated that has been produced in the first stage.

In particular, we aimed to perform a comprehensive study on the two set of neurodegenerative diseases i.e. AD and dementia. Also, the objective of this study was to build a ML classifier system for an unambiguous job of input and output alteration where one can find out the classification of a healthy and inflicted group of patients effectively. In this study, two datasets have been selected (described in Section 3.). One relates to AD and the other belongs to the dementia class of adults. In the first one, we have labeling based on the group of AD class i.e. AD or non-AD. Whereas, the other dataset is multi-class, which is based on brain metric CDR (Clinical Dementia Score) scale having four sets of values.

Therefore, to predict these set of label class, given the independent attributes, we applied twelve supervised machine learning classifiers for model training. Moreover, we built the model on a selected set of classifiers because it resulted in improved performance and hence, better accuracy. Those models that produced less accuracy were dropped and were excluded from the study. The classifiers employed while building a ML model are described in Table 1.

Table 1: ML Classifiers

S.No.	Classifiers	Description
1.	AdaBoost (Adaptive Boosting)	It is based mainly on classification problems which include binary class. It is employed to boost the performance of a base ML classifier, in particular decision trees classifier. It operates well on weak classifiers, which generates a strong classifier after the ensembling technique. The weights that are assigned are of wrong categorized instances (Cao et al., 2013). These are altered in such a manner that the selected classifiers lay emphasis on challenging cases (Cao et al., 2013).
2.	Bagging	It is based on the ensembling technique. Ensemble learning is a learning technique that aids in improving the ML prediction results by combining numerous ML classifiers (Grassi et al., 2019). The bagging type of ensemble learning when employed focuses on decreasing the variance of the model. There are numerous ways of using the ensembling method. Reducing the variance signifies reducing the overfitting of the model. This is achieved by voting or averaging process. Besides, it fits the base classifier on subsets of the dataset. After fitting, it aggregates the results of individual classifiers, thereby producing a decisive prediction. Thus, this is an approach that generates improved predictive performance as compared to a single model. In this study, we applied the Random Forest classifier and REPTree classifier as base estimators for building the bagging model.
3.	Classification Via Regression	This is a classifier that does classification using regression techniques. It generates regression-based models that predict the true/positive class. After this, the probabilities are calculated by the transformation of the predictions employing a softmax function (Classification via regression wrapper, 2019). In this, the class is binarized and then, a single regression model is created for every class value (Frank et al., 1998).
4.	Decision Table	It is a precise approach for numeric prediction. This is achieved using a decision trees classifier. It is based on if-then rules and holds an ability to be more compact, thereby leading to have more understandable results. A decision table classifier along with a certain default rule which is mapped to the majority class is used to classify discrete spaces with relatively improved accuracy, sometimes more than advanced induction algorithms (Kohavi, 1995).

S.No.	Classifiers	Description
5.	J48	It is a decision tree based classifier. The task of this classifier is to split each aspect of the data into small subsets to centre on a decision. It builds a decision node using the probable class estimations. Moreover, it deals with specific characteristics such as missing or lost attribute data estimations and altering attribute costs. The prediction accuracy is extended by performing pruning (Venkatesan and Velmurugan, 2015). In addition, the J48 classifier is the implementation of the ID3 classifier (Iterative Dichotomiser 3).
6.	Logistic Regression	It is the fundamental classifier in ML, used in classification. This classifier is based on a logistic function. And the logistic function is programmed on the rules of the sigmoid function. Also, the set of attributes (dependent variables) works on the concept of Bernoulli distribution. While the approximation of the set of labels (independent variables) is accomplished via maximum probability.
7.	Naive Bayes	It is a classification algorithm that is based on Bayes' theorem. Naive Bayes classifier is centred on the concept of class-conditional independence. In this, the effect of one attribute remains independent of other attributes.
8.	Random Forest	It works on the principle of ensemble technique. The ensembling method is a method that combines numerous single classifiers to build a ML model, which is primarily improved and more precise than the single classifier alone (Biau et al., 2008). A Random Forest classifier is an ensemble classifier that advances the entire performance of the model (Khan & Zubair, 2019b). It includes several decision tree classifiers and learns decision trees randomly. The decision tree consists of nodes. These nodes contain the information of individual attributes respectively. The decision trees are faster but are vulnerable to get adjusted disproportionately to the training data set or might result into a reduced performance via pruning of tree for generalization (Tin Kam Ho, 1995). Contrary to this, Random Forest, while building trees, enhances the performance of the model at every node by randomization of attributes.
9.	Random Tree	As stated earlier, a single decision tree alone can easily be conceptualized but undergoes high variance. Thus because of this reason, they do not yield high accuracy. This limitation of decision trees is handled by generating several variants of a single decision tree. This is achieved by selecting a dissimilar subset of the training data set every time. The selection is done based on the randomization of ensemble methods (Breiman, 2001). The Random Tree classifier is based on a class of ML algorithms that performs ensemble classification (Random Trees classifier, 2020). It builds a tree that includes varied attributes which are chosen randomly at each node. Also, it performs no tree pruning.

S.No.	Classifiers	Description
10.	REPTree (Reduced Error Pruning Tree)	This classifier is a fast decision tree learner. It creates a decision or regression tree by employing information gain while building a ML model. Further, it prunes the tree by using a reduced error pruning approach along with backfitting (REPTree, 2020).
11.	SGD (Stochastic Gradient Descent)	It is based on discriminative-based learning of linear ML classifiers. It functions better with the data characterized as a dense or sparse matrix of floating-point values (Ruder, 2016). SGD classifier is effective and can be implemented easily. Nevertheless, it includes several hyperparameters, regularization parameters and iterations.
12.	ZeroR	It is the simplest classification approach that depends on the label class while ignoring dependent attributes. It predicts for the majority category i.e. the label class. The predictability power of this classifier is less as compared to other classifiers. On the contrary, ZeroR classifier is beneficial in determining a standard (baseline) performance which acts as a benchmark for the rest of the ML classification methods.

## 2.2. Performance Evaluation Metrics

The performance of a classifier is evaluated by examining it with a reliable metric. In our study, to determine the effect of the applied methods in building an accurate ML model, we calculated seven extensively used metrics, namely, accuracy, precision, recall, F-measure, Receiver Operating Characteristic (ROC) area, Kappa statistic, and Root Mean Squared Error (RMSE). Next, we interpreted this with the AD and dementia problem accordingly.

1. Accuracy: It is the ability to calculate correct prediction data values out of all the concerned data values.
2. Precision: Precision is a measure that calculates the actual correct instances from amongst the positive identifications.
3. Recall: Recall is a measure that calculates how much of the given actual positive predictions were calculated correctly.
4. F-measure: It comprises of both precision and accuracy. Perhaps, it is regarded as the weighted average of both values.
5. Receiver Operating Characteristic (ROC) area: It shows the diagnostic ability of the ML model. The ROC analysis gives the means to choose for the optimal model.
6. Kappa statistic: It is a robust method of measuring inter-rater reliability, which assesses qualitatively and calculates the reliability amongst the provided data.
7. Root Mean Squared Error (RMSE): It is the square root of the mean of squared errors. RMSE is a measure that is used to calculate the differences among the data values which is calculated by the ML classifier.

### 3. Data Description

In this study, the ML models were trained and tested on the data obtained from the below two described datasets.

1. Dataset1 - OASIS Longitudinal: This is a longitudinal pool of Open Access Series of Imaging Studies (OASIS) MRI data. The class of subjects is segmented amongst the group of AD and non- AD older adults<sup>1</sup>. It consists of 150 subjects and 15 varied features for a total of 373 MRI sessions.
2. Dataset2 - OASIS Cross-Sectional: This is based on a cross-sectional pool of OASIS MRI scans<sup>1</sup>. We studied 416 subjects for a total of 434 MRI sessions. It includes 12 diverse sets of features. The class of subjects is segmented amongst 4 classes, from no dementia to very mild, mild and moderate dementia.

Denoting the datasets Dataset1 (OASIS Longitudinal) as D1 while Dataset2 (OASIS Cross-Sectional) as D2, before moving ahead in our study.

### 4. Experimental Results

#### 4.1. Demographic Characteristics

The demographic status of both the datasets D1 and D2 are illustrated in Table 2.

We can comprehend from both D1 and D2 that the sample size of a female is higher than that of a male. The subjects that have been studied were clinically identified with a very mild to moderate state of AD in D1. While the subjects in D2 were clinically diagnosed with very mild to mild and moderate dementia. The D1 consisted of 78 AD and 72 non-AD patients. Out of which, 38 females were diagnosed with AD and 50 with non-AD. Besides, D2 consisted of 316 subjects with no dementia while 100 subjects were diagnosed with dementia. Out of 316 subjects, 197 were females while 119 were found to be males. Out of the 100 demented subjects, 39 females were diagnosed with very mild, 19 with mild and 1 female with moderate dementia. Moreover, 31 males were detected with very mild, 9 with mild and 1 with moderate dementia.

Table 2: Demographic Profile

Dataset → Characteristic ↓	OASIS Longitudinal (D1)	OASIS Cross-Sectional (D2)
Subjects	150	416
Male	62	160
Female	88	256
Age range	60-96	18-96
Mean (SD) age	77.01 (7.64)	51.36 (25.27)
Median age	77.0	54.0

<sup>1</sup> <https://www.oasis-brains.org/>

## 4.2. Statistical Summary

The summary statistics of the studied attributes of both the datasets are described in Table 3 and Table 4.

The data illustrated in Table 3 and Table 4 provides the dispersion and central tendency of attributes of both D1 and D2 respectively. Such kind of statistical analysis further aided in our study in data pre-processing, mostly when dealing with missing values, the occurrence of outliers (extreme data values) and other related data-wrangling procedures.

The results that we presented in Table 2, Table 3 and Table 4 belong to all the male and female subjects. This was performed to look for the insights and patterns that remain among both the candidate subjects. Further in our study, we have presented the results of only the female subjects to scrutinize the effect of neurodegenerative diseases on females.

*Table 3: Statistical Description of OASIS Longitudinal Dataset*

<b>OASIS Longitudinal (D1)</b>					
	<b>Mean</b>	<b>Standard Deviation</b>	<b>Minimum</b>	<b>Median</b>	<b>Maximum</b>
<b>Age</b>	77.01	7.64	60.00	77.00	98.00
<b>ASF<sup>a</sup></b>	1.20	0.14	0.88	1.19	1.59
<b>CDR<sup>b</sup></b>	0.29	0.37	0.00	0.00	2.00
<b>Education</b>	14.60	2.88	6.00	15.00	23.00
<b>eTIV<sup>c</sup></b>	1488.13	176.14	1106.00	1470.00	2004.00
<b>MMSE<sup>d</sup></b>	27.34	3.68	4.00	29.00	30.00
<b>nWBV<sup>e</sup></b>	0.73	0.04	0.64	0.73	0.84
<b>SES<sup>f</sup></b>	2.46	1.13	1.00	2.00	5.00

*Table 4: Statistical Description of OASIS Cross-Sectional Dataset*

<b>OASIS Cross-Sectional (D2)</b>					
	<b>Mean</b>	<b>Standard Deviation</b>	<b>Minimum</b>	<b>Median</b>	<b>Maximum</b>
<b>Ag<sup>c</sup></b>	51.36	25.27	18.00	54.00	96.00
<b>ASF<sup>a</sup></b>	1.20	0.13	0.88	1.19	1.56
<b>CDR<sup>b</sup></b>	0.29	0.38	0.00	0.00	2.00
<b>Education</b>	3.18	1.31	1.00	3.00	5.00
<b>eTIV<sup>c</sup></b>	1481.92	158.74	1123.00	1475.50	1992.00
<b>MMSE<sup>d</sup></b>	27.06	3.70	14.00	29.00	30.00
<b>nWBV<sup>e</sup></b>	0.79	0.06	0.64	0.81	0.89
<b>SES<sup>f</sup></b>	2.49	1.12	1.00	2.00	5.00

<sup>a</sup>ASF: Atlas Scaling Factor; <sup>b</sup>CDR: Clinical Dementia Rating; <sup>c</sup>eTIV: estimated Total Intracranial Volume; <sup>d</sup>MMSE: Mini Mental State Examination; <sup>e</sup>nWBV: normalized Whole Brain Volume; <sup>f</sup>SES: Socio Economic Status



### 4.3. ML Model Evaluation

Amongst the ML modeling executed for several classifiers, we here present the results of only 12 selected classifiers. The classifiers that gave the improved performance, have been reported in the present study. Table 5 and Table 6 present the performance evaluation measurements of both datasets.

Each classifier was trained for cross-validation of 10 folds. Ahead of testing, every classifier went through hyperparameter tuning respectively. This, in general, aid in optimizing and building the ML model precisely. Also, it assists in evaluating the model performance with the help of related metrics accordingly.

From the above results, we can infer that the prediction accuracies range from 57.30 percent to 98.90 percent for D1 and for D2, the accuracies range from 78.00 percent to 99.60 percent. As again, these results are for female subjects only.

Among all the applied methods, several classifiers significantly outperform with higher accuracy and hence, improved performance. Such improvements were reached because of a proper data processing step was performed and also, during training and testing of the model, a correct set of hyperparameters were provided so that a robust model is generated. Moreover, we aimed to maintain a good balance between precision, recall and F-measure. We reached to attain this goal as we can see from the results in the table. High precision means less false positive rate while higher recall denotes to a less false-negative rate. Higher scores for both of these metrics signify that the built model is giving accurate results as an output i.e. higher precision. And also, the generated model is giving all positive results as a majority i.e. higher recall.

Table 5: Performance Results on OASIS Longitudinal Dataset

OASIS Longitudinal (D1)							
	Accuracy %	Precision %	Recall %	F-measure %	ROC %	Kappa Statistic %	RMSE %
<b>AdaBoost</b>	98.90	98.90	98.90	98.90	97.50	0.97	10.60
<b>Bagging (with Random Forest)</b>	98.90	98.90	98.90	98.90	99.90	0.97	28.07
<b>Bagging (with REPTree)</b>	57.30	50.00	57.30	53.40	49.90	0.00	49.14
<b>Classification Via Regression</b>	57.30	50.00	57.30	53.40	49.80	0.00	49.44
<b>Decision T able</b>	98.90	98.90	98.90	98.90	97.50	0.97	10.96
<b>J48</b>	98.90	98.90	98.90	98.90	97.50	0.97	10.71
<b>Logistic Regression</b>	92.10	92.20	92.10	92.10	97.10	0.84	27.94
<b>Na'ive Bayes</b>	97.80	97.90	97.80	97.80	99.90	0.95	12.73
<b>Random Forest</b>	98.90	98.90	98.90	98.90	99.90	0.98	26.88
<b>Random Tree</b>	73.00	79.30	73.00	69.90	80.00	0.41	40.12
<b>SGD</b>	98.90	98.90	98.90	98.90	98.70	0.98	10.60
<b>ZeroR</b>	57.30	50.00	57.30	53.40	47.00	0.00	49.50

Table 6: Performance Results on OASIS Cross-Sectional Dataset

OASIS Cross-Sectional (D2)							
	Accuracy %	Precision %	Recall %	F-measure %	ROC %	Kappa Statistic %	RMSE %
<b>AdaBoost</b>	99.60	98.60	99.60	99.60	99.60	0.98	12.12
<b>Bagging (with Random Forest)</b>	99.60	98.60	99.60	99.60	99.80	0.98	7.74
<b>Bagging (with REPTree)</b>	99.60	98.33	99.60	98.96	99.80	0.98	4.41
<b>Classification Via Regression</b>	99.60	99.60	99.60	99.60	99.70	0.98	7.43
<b>Decision T able</b>	99.60	98.33	99.60	98.96	99.90	0.98	4.93
<b>J48</b>	99.60	98.33	99.60	98.96	99.80	0.98	4.44
<b>Logistic Regression</b>	99.60	93.47	99.60	96.44	99.80	0.98	4.32
<b>Na'ive Bayes</b>	97.40	92.41	97.40	94.84	98.80	0.92	9.76
<b>Random Forest</b>	99.60	99.60	99.60	99.60	99.80	0.98	8.00
<b>Random Tree</b>	92.90	92.80	92.90	92.80	93.50	0.80	18.83
<b>SGD</b>	91.20	91.20	91.20	91.20	92.38	0.78	16.24
<b>ZeroR</b>	78.00	50.00	78.00	60.93	48.70	0.00	30.24

Also, the F-measure is the harmonic mean of precision and recall. Since we have both precision and recall on the same scale, the F-measure also show similar value results, if one of the precision or recall would have gone up and the other down, then it has been affected the F-measure accordingly. The next metric is the ROC score. As stated before, it describes the performance of both binary classifier (for D1) and multiclass classifier (for D2). We can infer from our results that as a higher ROC score has been achieved in building the ML model, it signifies an overall higher accuracy. Next, we can see from the results that the model accomplished to have a higher Kappa value. This means that there was a stronger degree of agreement. If it would have been low, then additional training of the dataset would be required. The RMSE shows to have a lower value. And this is a good indication because the lower the value of RMSE, the better the model has been fitted. Thus, we can say that our results attained a better advancement for the diagnosis of ND in female subjects.

## 5. Discussion

Specifically, clinical data exhibit several challenges which are not overcome easily with the present methods of ML (Goldstein, et al., 2017). Likewise, many clinical datasets comprise multiple kinds of data i.e. multimodal, or contain a lesser amount of samples relatively or encompass numerous missing data values. Handling such types of glitches normally involves data preprocessing extensively or dropping those attributes that cannot be modeled easily (Khan et al., 2019). Thus, developing such methodologies that can overcome these shortcomings is a vital step in developing robust ML models

in precision medicine. In this study, we examined AD and dementia diagnosed in female subjects. As both of these disorders are complicated neurodegenerative diseases with manifold cognitive and behavioral symptoms, we reached to attain the goal of obtaining improved performance of the entire ML model using ML tools and techniques.

In our results, we found that no one specific classifier resulted in higher accuracy. Rather, several applied classifiers were modeled in such a way that they resulted in overall better and improved performance. This study was entirely focused on female cases of neurodegenerative diseases. In the first place, from the D1 dataset, we observed that males (51.28 percent) were more prevalent to have AD than females (48.72 percent). While from the D2 dataset, we discovered that women (59.0 percent) were more likely to have dementia as compared with men (41.0 percent). According to several studies, it has been found that numerous key differences often emerge amongst women and men in the incidence, appearance, and development of psychiatric conditions (Khan & Zubair, 2020d; Sukel, 2018). Previous studies performed suggests that females are more susceptible to acquire ND as they are at a larger threat of getting depression than males (Hara, Y., 2018). Moreover, the APOE-e4 gene (Apolipoprotein E) affects both male and female in a different manner (Hara, 2018). Another study stated that age, gender and APOE-e4 gene are the major risk factors in the progression of neurodegenerative diseases (Riedel et al., 2016). Also, the rate of these disorders is almost similar in both genders. It is when the incidence becomes more noticeable in females at a late age (Riedel et al., 2016).

The comprehensive analysis performed in this study validates the prevalence of AD and dementia in women. Subsequently, neurodegenerative diseases hold a lower survival rate. Because of this reason, there is scarce data available of the aged adults. Additionally, the presented ML models hold an unbiased comparison. We cannot say that a particular classifier outperformed any other specific classifier. This was not the basis of our study. We accomplished to have an overall robust model for both the disorders for female subjects. Thus, we can assert that our presented method of ML modeling in the diagnosis of AD and dementia ascertains to perform better. All seven performance measures showed an improved score significantly. Therefore, it can be well stated that our methodology is effective which can essentially identify, execute an improved diagnosis and thus, classify healthily and inflicted AD and dementia patients individually and respectively.

## 6. Conclusion

This work has shown the classification (both binary and multi-class) problem in diagnosing neurodegenerative diseases in female subjects. The study highlights that when such problems are handled correctly i.e. a proper ML process is carried out, then they can be successfully be solved by building robust ML models. The as-developed approach outperforms state-of-the-art learning classifiers on problems with both binary and multivalued attributes. We achieved prediction accuracies ranging from 57.30 percent to 98.90 percent for longitudinal data and 78.00 percent to 99.60 percent for cross-sectional data. The experimental results had shown a high accuracy of 98.90 for the AD recognition and 99.60 for the dementia prognosis respectively. Amongst all the employed approaches, several classifiers outperformed considerably with high accuracy, thereby, an enhanced performance was observed. Lastly, we infer that the MRI biomarkers stand to have a key role in the prognosis of AD and dementia. Altogether, our results were obtained on a definite sample size and might not be applied to the broad population range. Hence, extensive studies may perhaps be necessitated to support our findings. Our

future study will be based on large-scale data. We will further be employing hybrid-based ML modeling to build operative decision systems for the prediction of AD and dementia.

## 7. References

- Ahmed, F., Samorani, M., Bellinger, C., & Zaiane, O. R. (2016). Advantage of Integration in Big Data: Feature Generation in Multi- Relational Databases for Imbalanced Learning. In *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016* (pp. 532–539). <https://doi.org/10.1109/BigData.2016.7840644>
- Biau, G., Devroye, L., & Lugosi, G. (2008). Consistency of Random Forests and Other Averaging Classifiers. *Journal of Machine Learning Research*, 9, 2015–2033.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cao, Y., Miao, Q. G., Liu, J. C., & Gao, L. (2013). Advance and prospects of AdaBoost algorithm. *Acta Automatica Sinica*, 39(6), 745–758. <https://doi.org/10.3724/SP.J.1004.2013.00745>
- Classification via regression wrapper (2019). Retrieved from <https://mlr.mlr-org.com/reference/makeClassificationViaRegressionWrapper.html>
- Cummings, J., Aisen, P. S., Dubois, B., Frolich, L., Jack, C. R., Jones, R. W., ... Scheltens, P. (2016). Drug development in Alzheimer's disease: The path to 2025. *Alzheimer's Research and Therapy*, 8(1), 1–12. <https://doi.org/10.1186/s13195-016-0207-9>
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198. [https://doi.org/10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6)
- Frank, E., Wang, Y., Inglis, S., Holmes, G., & Witten, I. H. (1998). Using Model Trees for Classification. *Machine Learning*, 32(1), 63–76. <https://doi.org/10.1023/A:1007421302149>
- Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, 24(1), 198–208. <https://doi.org/10.1093/jamia/ocw042>
- Grassi, M., Rouleaux, N., Caldirola, D., & Loewenstein, D. (2019). A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion From Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures. *10(July)*, 1–15. <https://doi.org/10.3389/fneur.2019.00756>
- Hara, Y. (2018, July 2). How does Alzheimer's affect women and men differently? Retrieved from <https://www.alzdiscovery.org/cognitive-vitality/blog/how-does-alzheimers-affect-women-and-men-differently>
- Hinrichs, C., Singh, V., Xu, G., & Johnson, S. C. (2011). Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *NeuroImage*, 55(2), 574–589. <https://doi.org/10.1016/j.neuroimage.2010.10.081>
- Ito, K., Corrigan, B., Romero, K., Anziano, R., Neville, J., Stephenson, D., & Lalonde, R. (2013). Understanding placebo responses in Alzheimer's disease clinical trials from the literature meta-data and CAMD database. *Journal of Alzheimer's Disease*, 37(1), 173–183. <https://doi.org/10.3233/JAD-130575>

- Kennedy, R. E., Cutter, G. R., Wang, G., & Schneider, L. S. (2016). Post Hoc Analyses of ApoE Genotype-Defined Subgroups in Clinical Trials. *Journal of Alzheimer's Disease*, 50(4), 1205–1215. <https://doi.org/10.3233/JAD-150847>
- Khan, A., & Zubair, S. (2018). Machine Learning Tools and Toolkits in the Exploration of Big Data. *International Journal of Computer Sciences and Engineering*, 6(12), 570–575. <https://doi.org/10.26438/ijcse/v6i12.570575>
- Khan, A., Zubair, S., & Sabri, M. Al. (2019a). 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). IEEE.
- Khan, A., & Zubair, S. (2019b). 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). A Machine Learning-based robust approach to identify Dementia progression employing Dimensionality Reduction in Cross-Sectional MRI data. In *2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia* (pp. 237–242). <https://doi.org/0.1109/SMART-TECH49988.2020.00060>
- Khan, A., & Zubair, S. (2020b). An Improved Multi-Modal based Machine Learning Approach for the Prognosis of Alzheimer's Disease. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2020.04.004>
- Khan, A., & Zubair, S. (2020c). Expansion of Regularized Kmeans Discretization Machine Learning Approach in Prognosis of Dementia Progression. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India* (pp. 1–6). <https://doi.org/10.1109/ICCCNT49239.2020.9225397>.
- Khan, A., & Zubair, S. (2020d). Longitudinal Magnetic Resonance Imaging as a Potential Correlate in the Diagnosis of Alzheimer Disease: Exploratory Data Analysis. *JMIR Biomedical Engineering*, 5(1), 1–13. <https://doi.org/10.2196/14389>
- Kohavi, R. (1995). The Power of Decision Tables. In *ECML'95: Proceedings of the 8th European Conference on Machine Learning* (pp. 174–189). [https://doi.org/10.1007/3-540-59286-5\\_57](https://doi.org/10.1007/3-540-59286-5_57)
- Lasko, T. A., Denny, J. C., & Levy, M. A. (2013). Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data. *PLoS ONE*, 8(6). <https://doi.org/10.1371/journal.pone.0066341>
- Medical Research Council (2019). Neurodegeneration, dementia, and mental health. Retrieved from <https://mrc.ukri.org/successes/investing-for-impact/priority-challenges/neurodegeneration-dementia-and-mental-health/>
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., ... Beckett, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's and Dementia*, 1(1), 55–66. <https://doi.org/10.1016/j.jalz.2005.06.003>
- Myers, P. D., Scirica, B. M., & Stultz, C. M. (2017). Machine Learning Improves Risk Stratification after Acute Coronary Syndrome. *Scientific Reports*, 7(1), 1–12. <https://doi.org/10.1038/s41598-017-12951-x>
- Random Trees classifier. (2020). Retrieved from [https://www.pcigeomatics.com/geomatica-help/concepts/focus\\_c/oa\\_classif\\_intro\\_rt.html](https://www.pcigeomatics.com/geomatica-help/concepts/focus_c/oa_classif_intro_rt.html)
- REPTree (2020). Retrieved from <https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/REPTree.html>

- Riedel, B. C., Thompson, P. M., & Brinton, R. D. (2016). Age, APOE and Sex: Triad of Risk of Alzheimer's Disease. *J Steroid Biochem Mol Biol.*, 134–147. <https://doi.org/10.1016/j.jsbmb.2016.03.012>.
- Risacher, S., Saykin, A., Wes, J., Shen, L., Firpi, H., & McDonald, B. (2009). Baseline MRI Predictors of Conversion from MCI to Probable AD in the ADNI Cohort. *Current Alzheimer Research*, 6(4), 347–361. <https://doi.org/10.2174/156720509788929273>
- Rogers, J. A., Polhamus, D., Gillespie, W. R., Ito, K., Romero, K., Qiu, R., . Corrigan, B. (2012). Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a beta regression meta-analysis. *Journal of Pharmacokinetics and Pharmacodynamics*, 39(5), 479–498. <https://doi.org/10.1007/s10928-012-9263-3>
- Romero, K., Ito, K., Rogers, J. A., Polhamus, D., Qiu, R., Stephenson, D., . Corrigan, B. (2015). The Future Is Now: Model-Based Clinical Trial Design for Alzheimer's Disease. *Clinical Pharmacology and Therapeutics*, 97(3), 210–214. <https://doi.org/10.1002/cpt.16>
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. Retrieved from <http://ruder.io/optimizing-gradient-descent/>
- Sonkusare, S. K., Kaul, C. L., & Ramarao, P. (2005). Dementia of Alzheimer's disease and other neurodegenerative disorders — memantine, a new hope. *Pharmacological Research*, 51(1), 1–17. <https://doi.org/10.1016/j.phrs.2004.05.005>
- Suk, H. II, Lee, S. W., & Shen, D. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*, 101, 569–582. <https://doi.org/10.1016/j.neuroimage.2014.06.077>
- Sukel, K. (2018, November 15). Figuring Out Why Alzheimer's Disease Strikes More Women Than Men. Retrieved from <https://www.brainfacts.org/diseases-and-disorders/topic-center-alzheimers-and-dementia/figuring-out-why-alzheimers-di-sease-strikes-more-women-than-men-1115183>
- Szalkai, B., Grolmusz, V. K., & Grolmusz, V. I. (2017). Identifying combinatorial biomarkers by association rule mining in the CAMD Alzheimer's database. *Archives of Gerontology and Geriatrics*, 73, 300–307. <https://doi.org/10.1016/j.archger.2017.08.006>
- Team, E. (2020, May 6). What is Machine Learning? A definition - Expert System. Retrieved from <https://www.expert.ai/blog/machine-learning-definition/>
- Tin Kam Ho. (1995). Random Decision Forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (pp. 278–282). Retrieved from <https://ieeexplore.ieee.org/abstract/document/598994/>
- Venkatesan, E., & Velmurugan, T. (2015). Performance Analysis of Decision Tree Algorithms for Breast Cancer Classification. *Indian Journal of Science and Technology*, 8(29), 1–8. <https://doi.org/10.17485/ijst/2015/v8i29/84646>
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., . Trojanowski, J. Q. (2013). The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's and Dementia*, 9(5), e111–e194. <https://doi.org/10.1016/j.jalz.2013.05.1769>