Dementia Prediction Using Machine Learning Algorithms

Habiba Elsherbiny
Department of Electrical and Computer Engineering
Queen's University
Kingston, Ontario, Canada
Habeeba.564@gmail.com

***Abstract*—Dementia is a category of brain disorders that results in declining cognitive functioning and behavioral abilities in a way that severely affects a person's daily life activities. These functions and abilities include memory, visual perception, and the ability to focus and pay attention [1]. A patient's level of dementia can range from a mild stage, where some of the patient's activities are affected, to a very severe stage, where the patient can't survive without the constant aid of others. People with dementia may have problems with shot-term memory, keeping track of their belongings, and following up during conversations. Dementia diagnosis consists of a series of physical examinations and laboratory tests, as well as medical history review. Over the past years, machine learning has been successfully applied to multiple healthcare problems. This paper discusses the use of different machine learning techniques for dementia prediction in old adults. Various machine learning algorithms were trained and tested in order to asses and compare their classification accuracies and determine the optimal one for dementia prediction.**

***Keywords*—Dementia, machine learning algorithms and healthcare applications.**

I. Introduction

Dementia is considered one of the world's fastest growing diseases. According to Alzheimer's Disease International, someone in the world develops dementia every 3 seconds [2]. The disorder is caused by a damage of the brain cells, which inhibits the brain cells from communicating with each other properly. Current techniques for dementia diagnosis include medical history review, physical and neurological examinations and diagnostic tests. The doctors cannot determine for sure if the person has dementia, but the diagnosis can help them have a high level of certainty about it. Research shows that the majority of people living with dementia at the present time have not been formally diagnosed with the disorder, and therefore don't receive the treatment and medical care they need to survive [2]. Currently, there are various known causes of dementia, with Alzheimer's disease being the most common one. Although treatment cannot cure dementia, it can reduce some of its effects, such as thinking and memory problems. Therefore, it is important to ensure early diagnosis of dementia and provide treatment to demented patients as soon as possible.

This clearly shows that we need to provide more powerful techniques that can easily diagnose dementia and even predict it before it has some serious effects. These techniques should be accessible to everyone in the world, in order increase the awareness of the disorder and ensure that everyone with the disorder is diagnosed and treated.

Machine learning techniques have drawn a lot of academic and industrial interest and have been successful in a variety of application areas, such as: computer vision, speech recognition and natural language processing. Currently, machine learning plays an important role in the healthcare field, performing different tasks ranging from medical image segmentation to disease diagnosis [3]. Machine learning algorithms build a mathematical model of training data in order to make predictions or decisions without being explicitly programmed to perform the task [4]. Generally, machine learning algorithms fall into three categories: supervised learning, unsupervised learning and reinforcement learning.

The objective of this paper is to discuss the application of supervised learning algorithms on MRI data to perform dementia prediction in old adults. For optimal performance, the data must be prepared before it's fed to the machine learning classifiers. Therefore, several data preprocessing steps such as data cleaning and data transformation were performed. By training different machine learning classifiers and comparing their results, the optimal classifier for dementia prediction could be determined.

The rest of this paper is organized as follows. Section II presents and discusses the dataset. Section III explains the methodology used for preparing the data, as well as training and testing the machine learning classifiers. Section IV demonstrates and compares the experimental results. Finally, concluding remarks and future work are described in Section V.

II. Dataset

The dataset used is the OASIS-2: Longitudinal MRI Data in Nondemented and Demented Older Adults [5].

"This set consists of a longitudinal collection of 150 subjects aged 60 to 96. Each subject was scanned on two or more visits, separated by at least one year for a total of 373 imaging sessions. For each subject, 3 or 4 individual T1-

weighted MRI scans obtained in single scan sessions are included. The subjects are all right-handed and include both men and women. 72 of the subjects were characterized as nondemented throughout the study. 64 of the included subjects were characterized as demented at the time of their initial visits and remained so for subsequent scans, including 51 individuals with mild to moderate Alzheimer's disease. Another 14 subjects were characterized as nondemented at the time of their initial visit and were subsequently characterized as demented at a later visit." [5]

The dataset contains demographic information such as gender, age, years of education, socioeconomic status and handedness. In addition, the dataset includes some derived anatomic volumes such as the estimated total intracranial volume (eTIV), normalized whole-brain volume (nWBV), and atlas scaling factor (ASF). Furthermore, the dataset contains clinical information such as mini-mental state examination score (MMSE) and clinical dementia rating (CDR)

A. *Understanding the data attributes*

The Mini–Mental State Examination (MMSE) or Folstein test is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment [6]. It is used to test five areas of cognitive function: orientation, registration, attention and calculation, recall, and language and is commonly used in medicine for dementia diagnosis. In addition, it is used to estimate the severity and progression of cognitive impairment and to track a person's cognitive status changes over time [7]. Therefore, it can be used to capture how the individual responds to treatment. A score that is higher than or equal to 24 points (out of 30) corresponds to a normal cognition. On the other hand, scores below 24 can indicate severe (⩽9 points), moderate (10 – 18 points) or mild (19 – 23 points) cognitive impairment. Low to very low scores can indicate the presence of dementia, although they could also be the result of the presence of other mental disorders.

The CDR is a numeric scale that is used to characterize a patient's cognitive and functional performance in six domains. These domains are: memory, orientation, judgment & problem solving, community affairs, home & hobbies, and personal care [8]. Each of these domains has a single score, and all of the scores are combined to obtain a single score ranging from 0 to 3. The CDR is based on information obtained through a structured clinician-rated interview of the patient and a reliable informant such as a family member. The CDR Scoring Table provides descriptive anchors that guide the clinician in making appropriate ratings based on interview data and clinical judgment [9].

This scale can be used for assessing a patient's level of dementia:

- 0= Normal
- 1.5 = Very Mild Dementia
- 1 = Mild Dementia
- 2 = Moderate Dementia
- 3 = Severe Dementia

B. *Visualizing the data*

Since the CDR value determines the patient's level of dementia, we need to understand how it depends on the other data attributes. Fig. 1 and Fig. 2 display the distribution of different data variables by CDR Rating.

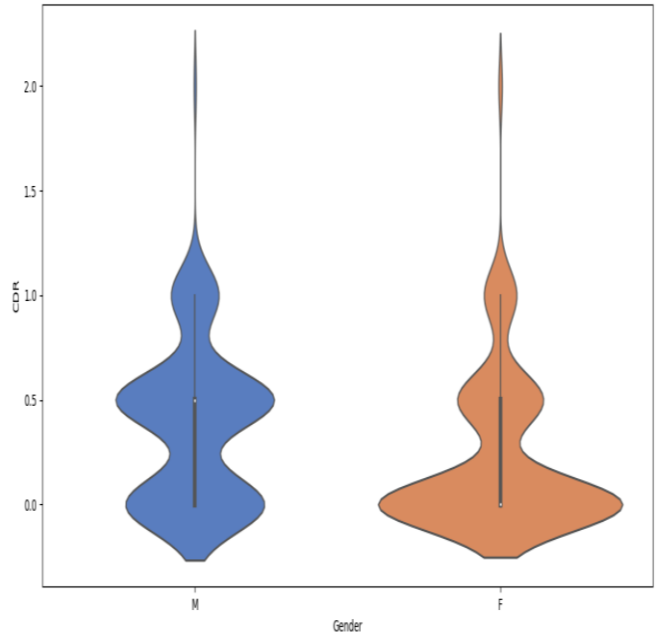

Fig. 1. The distribution of gender by CDR

As shown in Fig.1, there is no apparent connection between the CDR and the gender. Therefore, we can conclude that the CDR is independent of the patient's gender.

The scatterplots in Fig.2 (a), Fig.2 (b), Fig.2 (c), and Fig.2 (d) show that the CDR is also independent of the patient's age, years of education, SES and eTIV. As shown in Fig.2 (e), the MMSE of demented people is dispersed across the whole axis, while that of undemented people is concentrated around 20-30 point rates. Moreover, the mean and median of the MMSE are lower for demented subjects than for undemented subjects. On the other hand, Fig.2 (f) shows that the nWBV of patients characterized with dementia is concentrated around 0.65-0.75, while that of undemented patients is more dispersed.

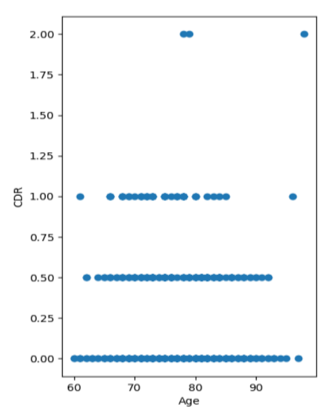

Fig. 2 (a)

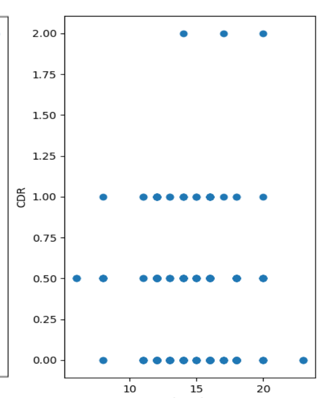

Fig. 2 (b)

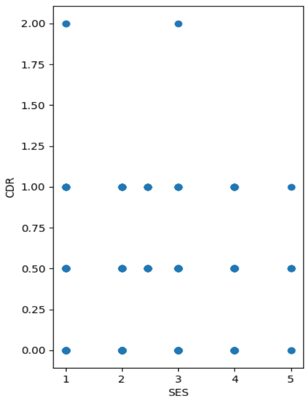

Fig. 2 (c)

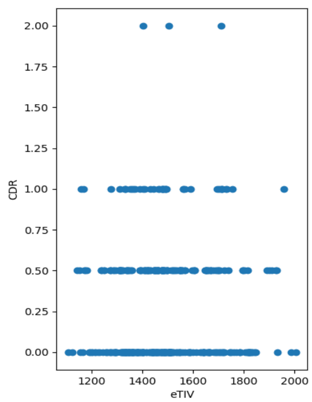

Fig. 2 (d)

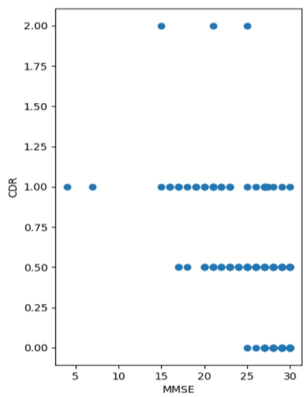

Fig. 2 (e)

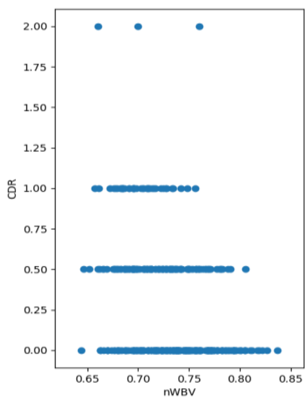

Fig. 2 (f)

Fig. 2 The distribution of each variable by the CDR

III. Methodology

A. *Data Preprocessing*

The dataset was preprocessed in order to deal with missing values. Handling missing values in the datasets is commonly done by replacing the missing values by the mean or the median of the other values of that variable. A median imputation is a better choice when the data contains outliers. Because the dataset is small, and therefore more sensitive to noise and outliers, imputation was performed by replacing the missing values with the median of the column. This was done using the Sklearn's library SimpleImputer [10].

Because all of the subjects in the dataset are right-handed, the handedness column was dropped from the dataset and therefore not used in the model.

A threshold value was set up for prediction, so that if the CDR of the subject is above 0.5, the subject would be characterized as demented. Otherwise, it would be characterized as undemented.

Supervised learning algorithms require a training set and a test set. Therefore, the dataset was divided into 70% for the training set and 30% for the test set. Subsequently, different machine learning classifiers were trained using the training set and then their results on the test set were compared in order to determine the optimal one for the problem.

B. *Machine Learning Classifiers Training*

After data preprocessing, the following six supervised learning classifiers were trained and tested on the OASIS dataset: Logistic Regression [11], K- Nearest Neighbors (KNN) [12], Support Vector Machines (SVM) [13], Decision Trees [14], Gradient Boosting [15] , and Adaboost [16].

1) *Logistic Regression*

The first model trained was the logistic regression with L1 penalty. Using L1 penalty helped prevent the overfitting problem that would occur because of the small dataset size.

2) *KNN*

A KNN classifier was then trained on the data. Different k values were tried in order to tune the classifier to produce the highest accuracy, and the best tune was with k = 4.

3) *SVM*

SVMs are considered a powerful technique of supervised learning. An SVM model was trained using various cost values, and the optimal value was cost =8.

4) *Decision Trees*

Another classification algorithm used was a decision tees classifier. As well as being used for classification, the decision trees classifier could also determine the feature importance. According to the feature importance produced from the classifier, the MMSE is the most important feature.

5) *Gradient Boosting*

Ensemble techniques are known to boost the power of machine learning models. A type of ensemble learning is the Gradient Boosting algorithm. A gradient boosting classifier was trained and tuned for best performance. The optimal tune was of depth = 3 and learning rate = 0.1.

6) *Adaboost*

Another ensemble learning algorithm used was the Adaboost. Various "number of estimators" values were tried and tested, and a value of 100 resulted in the highest accuracy.

C. *Overfitting Prevention using K-fold Cross Validation*

Because the size of the dataset is fairly small, K-fold cross validation was tried as a preventative measure against overfitting. A K-fold cross validation [17] with k = 10 resulted in a higher classification accuracy than splitting the data into 70%/30% training set and test set.

IV. Experimental Results

The six classification algorithms, namely logistic regression, KNN, SVM, decision trees, gradient boosting and adaboost, were used to classify the data. Table 1 and Table 2 display the accuracy, precision, recall and F1-score of each classifier. As shown in the table, the Adaboost achieved the best performance, with an accuracy of 87%. With an accuracy of 85%, the Gradient Boosting classifier was found to be comparable to the Adaboost classifier. In addition, the SVM and Decision Trees classifiers performed fairly well, with classification accuracies of 80% and 75% respectively. On the other hand, the KNN classifier had the lowest classification accuracy of 67%.

Table 1. The accuracy and precision results

Classifier	Accuracy	Precision
Logistic Regression	0.80	0.84
KNN	0.67	0.70
SVM	0.80	0.83
Decision Trees	0.75	0.76
Gradient Boosting	0.85	0.86
Adaboost	0.87	0.88

Table 2. The recall and F1-score results

Classifier	Recall	F1-score
Logistic Regression	0.80	0.79
KNN	0.68	0.66
SVM	0.80	0.79
Decision Trees	0.76	0.76
Gradient Boosting	0.86	0.86
Adaboost	0.88	0.88

In order to visualize the performance of each classification algorithm, the confusion matrices were computed (see Fig.3-Fig. 8). Instances in the actual class are represented by rows, while the instances in a predicted class are represented by columns.

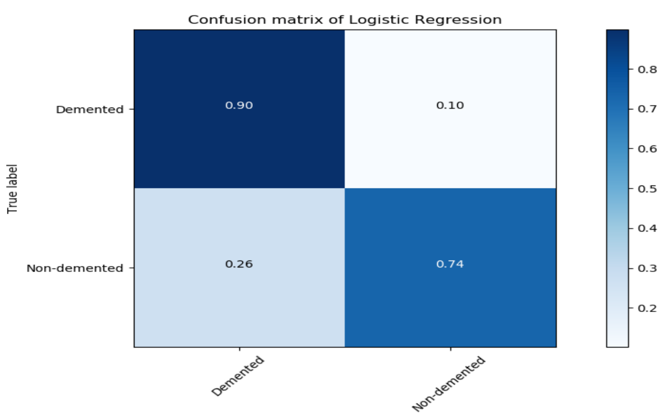

Fig. 3. The confusion matrix of Logistic Regression

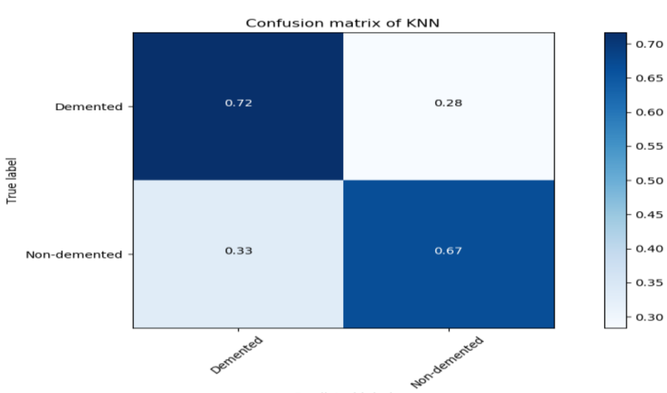

Fig. 4. The confusion matrix of KNN

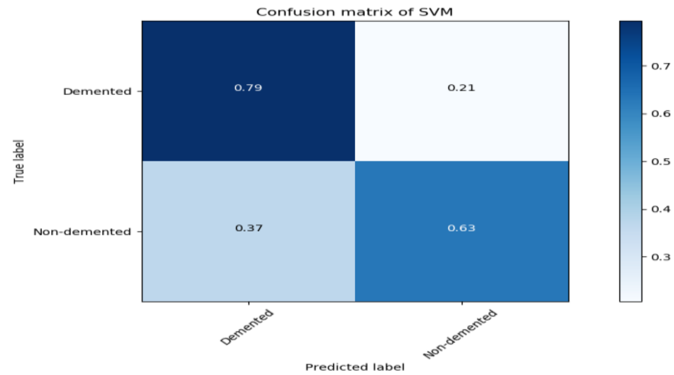

Fig. 5. The confusion matrix of SVM

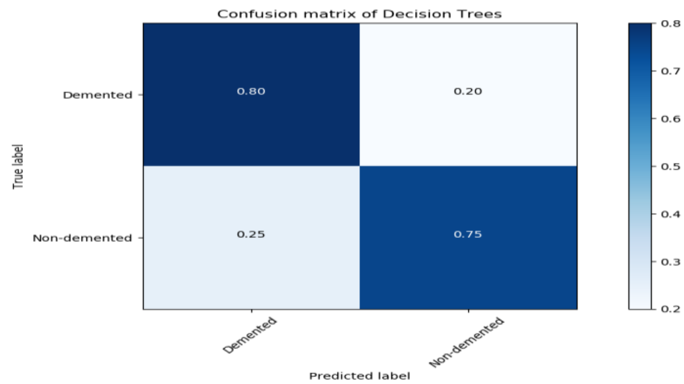

Fig. 6 The confusion matrix of Decision Trees

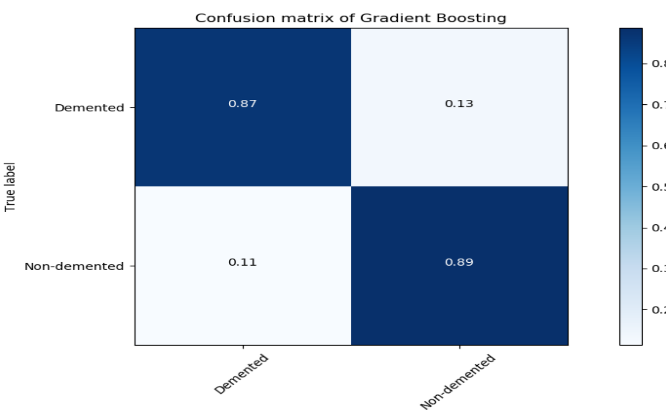

Fig. 7. The confusion matrix of Gradient Boosting

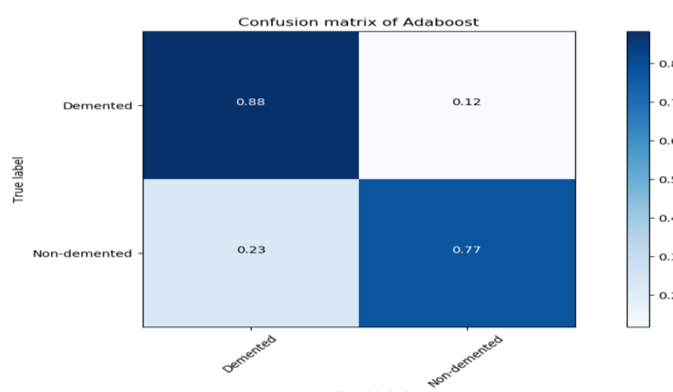

Fig. 8. The confusion matrix of Adaboost

In addition, the ROC curves were computed in order to compare the performance of the classifiers. Fig. 9 displays the ROC curve of the Logistic Regression classifier, Fig. 10 illustrates the ROC curve of the KNN classifier, and Fig. 11 shows the ROC curve of the SVM classifier.

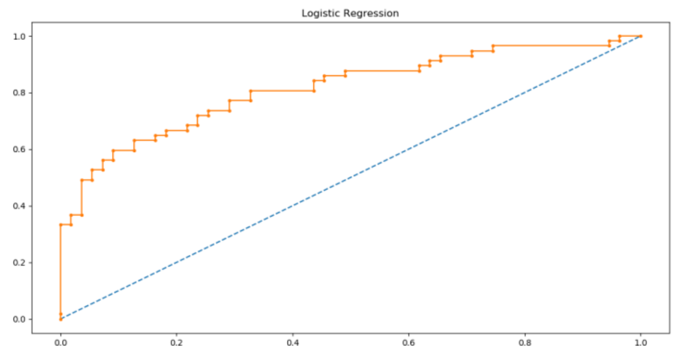

Fig. 9. The ROC curve of the Logistic Regression classifier

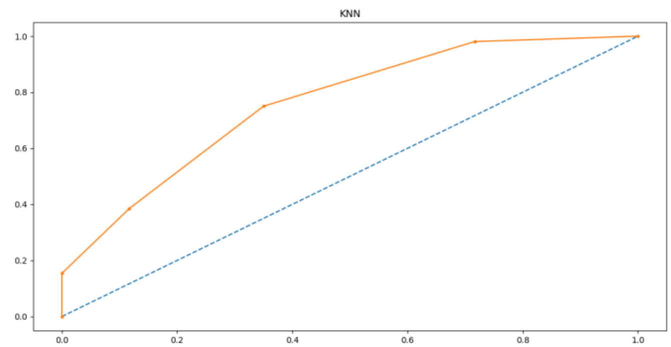

Fig. 10. The ROC curve of the KNN classifier

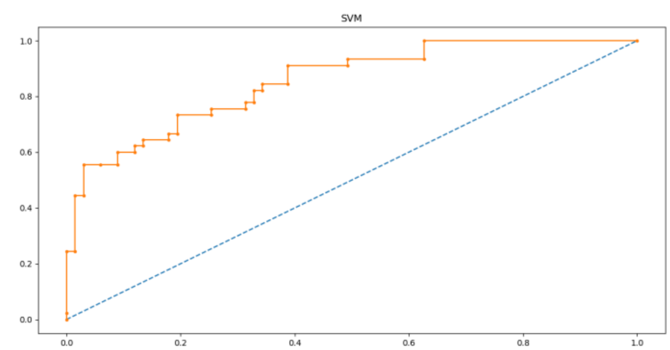

Fig. 11. The ROC curve of the SVM classifier

Moreover, the AUC of the ROC curves was computed. The following table compares between the different AUC values of the classifiers. As shown in Table 3, the AUC values of logistic Regression and SVM were higher than that of the KNN.

Table 3 The AUC values of the ROC curves

Classifier	**AUC**
Logistic Regression	0.85
SVM	0.85
KNN	0.73

V. Conclusion and Future Work

In this paper, I have applied six different machine learning algorithms for predicting dementia in old adults using the OASIS dataset. Out of the six machine learning classifiers trained, the Adaboost classifier had the best performance, followed directly by the Gradient Boosting classifier. On the contrary, the KNN classifier had the worst performance.

For future work, obtaining more data could lead to the development more accurate and robust classification models. Furthermore, it would be interesting to investigate the use of deep learning algorithms for dementia classification and prediction, since deep learning models have been shown to have higher prediction accuracies than traditional machine learning models.

References

[1] "What Is Dementia?", December 2017, National Institute on Aging https://www.nia.nih.gov/health/what-dementia.

[2] "Dementia statistics", https://www.alz.co.uk/research/statistics.

[3] A. Khare, M. Jeon, I. Sethi, B. Xu, "Machine Learning Theory and Applications for Healthcare", Journal of Healthcare Engineering, Sep. 2017.

[4] Wikipedia contributors. (2018, December 8). Machine learning. In Wikipedia, The Free Encyclopedia. Retrieved 04:55, December 9, 2018, from https://en.wikipedia.org/w/index.php?title=Machine_learning&oldid=872618259.

[5] Marcus, DS, Fotenos, AF, Csernansky, JG, Morris, JC, Buckner, RL, "Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults", Journal of Cognitive Neuroscience, 2010, vol. 22, pp.2677-2684.

[6] Wikipedia contributors. (2018, December 3). Mini–Mental State Examination. In Wikipedia, The Free Encyclopedia. Retrieved 05:35, December 9, 2018, from https://en.wikipedia.org/w/index.php?title=Mini%E2%80%93Mental_State_Examination&oldid=871821416.

[7] L. Kurlowiez, M. Wallace, "The Mini Mental State Examination (MMSE)", The Hartford Institute for Geriartric Nursing, Jan, 1999.

[8] M. Kogan, H. S. Jeong, Chapter 11 - Alzheimer Disease, Editor(s): David Rakel, Integrative Medicine (Fourth Edition), Elsevier, 2018, pp. 95-107.e4, ISBN 9780323358682, https://doi.org/10.1016/B978-0-323-35868-2.00011-6.

[9] "Clinical Dementia Rating (CDR) Scale", Massachusetts Alzheimer's Disease Research Center.

[10] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

[11] S. Menard, (2010). Logistic regression: From introductory to advanced concepts and applications. Thousand Oaks, CA: SAGE Publications, Inc.

[12] T. Srivastava, Oct 10, 2014, "Introduction to k-Nearest Neighbors: Simplified (with implementation in Python)", Analytics Vidhya.

[13] Support Vector Machines. (2008). Springer New York.

[14] H. K. Sok., M. P. Ooi, Y. C. Kuang, & S. Demidenko, (2016). Multivariate alternating decision trees. Pattern Recognition, 50(Complete), 195-209.

[15] J. H. Friedman, (2002). Stochastic gradient boosting. Computational Statistics and Data Analysis, 38(4), 367-378.

[16] T. An and M. Kim, "A New Diverse AdaBoost Classifier," 2010 International Conference on Artificial Intelligence and Computational Intelligence, Sanya, 2010, pp. 359-363.

P. Refaeilzadeh, L. Tang, H. Liu, (2016) Cross-Validation. In: Liu L., Özsu M. (eds) Encyclopedia of Database Systems. Springer, New York,