ISSN 1870-4069

Stacking Ensemble for Cognitive Impairment and Alzheimer's Disease Classification Using the ADNI Database

Sergio Vega-Guzmán[1], Gerardo Ramírez-Nava[1,2], Mariel Alfaro-Ponce[1,2]

[1] Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Mexico

[2] Tecnológico de Monterrey, Instituto de Materiales Avanzados para la Manufactura Sostenible, Mexico

{a01194108, gerardo.j.ramirez, mariel.alfaro}@tec.mx

Abstract. Dementia is a medical condition encompassing a broad spectrum of cognitive impairments, including a progressive decline in cognitive, motor, and memory skills. Although numerous types of dementia have been identified to date, Alzheimer's disease is still the most extensively studied due to its high prevalence and impact on individuals and society. The Alzheimer's Disease Neuroimaging Initiative (ADNI)is a collaborative research effort dedicated to studying Alzheimer's Disease neuropathology. ADNI has collected clinical data through different study phases, such as laboratory analysis, biomarkers, genetic information, brain imaging, volumetric information, cognitive tests, and other clinical measurements. This information allowed to conform a database that has contributed to the development of multiple scientific studies and clinical trials, including those that have implemented machine learning and deep learning algorithms to classify cognitive impairment stages and the severity of dementia symptoms. Stacked ensemble methods are an interesting alternative that fuses the strengths of several classification base models. This approach has provided flexible frameworks for combining multiple models, leveraging their strengths, and thus making more accurate classifications and predictions. This paper reports a stacking ensemble of classic machine-learning models to classify Alzheimer's disease, normal cognition, and mild cognitive impairment. The stacked ensemble comprises three Gradient Boosting Machine, two Extreme Gradient Boosting models, and two Distributed Random Forests that reached an overall accuracy of 86.9% in the classification process.

Keywords: Cognitive impairment, dementia, stacked ensemble.

1 Introduction

Dementia, as a clinical term, encapsulates a broad spectrum of cognitive impairments where an individual progressively deviates from their normative behavioural patterns to

the extent that they can no longer accomplish tasks routinely expected from a person in their respective age group [11]. This neurodegenerative disease manifests itself most commonly as memory loss but can also come as motor function reduction, spatial awareness decline, and general disorientation and confusion. As per recent studies, it is estimated that around 50 million individuals globally are affected by dementia, a number that showcases the threat of this global health concern [17].

The rising prevalence of dementia worldwide is predicted to escalate further in the future, attributed primarily to the consistent increase in the average lifespan and the consequent growth of the elderly population [12]. Numerous types of dementia have been identified to date, each representing unique facets of neurodegenerative pathologies. These include Vascular dementia, Lewy body dementia, Parkinson's disease, and Alzheimer's disease (AD), each with distinct symptomatology and progression patterns [10].

AD, the most prevalent neurodegenerative pathology, accounts for approximately 70 percent of all dementia occurrences. The alarming rate of its incidence, which is said to double every 5 to 10 years, implies that people in age brackets of 65-69, 70-74, 75-79, 80-84 are at a continually increasing risk, with likelihoods of 0.6%, 1.0%, 2.0%, 3.3%, and 8.4%, respectively [5]. It is pertinent to mention that AD often does not begin with severe symptoms. In many cases, the early stages manifest as Mild Cognitive Impairment (MCI), a condition considered a transitional stage between the expected cognitive decline of normal aging and the more serious decline of dementia.

Individuals with MCI often experience noticeable cognitive changes to the people around them and to themselves, but not severe enough to interfere with their daily life or independent function to a concerning point. Despite not all people with MCI developing AD, a significant proportion do, with studies suggesting that MCI patients progress to AD at a rate of approximately 10-15% per year. Therefore, the importance of the MCI denomination lies in its strong correlation with the progression to AD, making its early detection and study crucial for understanding, preventing, and treating this neurodegenerative condition [14].

Several risk factors contributing to Alzheimer's have been identified in scientific literature, including a family history of dementia, a history of head trauma, certain genetic factors, the presence of two X chromosomes, lower education levels, and vascular disease. These factors, in turn, have led to the identification of several biomarkers that have shown to produce accurate classification results when incorporated into machine learning and deep learning algorithms [5].

These algorithms have been trained on clinical and imaging data to produce an acceptable model capable of classifying AD stages or forecasting the progression from MCI to AD. In the relevant literature, several examples of this can be found. For instance, the work of Beltrán [2] used the ADNI database to predict the transition from MCI to AD. To do so, several machine learning models were implemented, with Random Forests (RF) and Gradient Boosting Machines (GBM) being the most successful of them, achieving an AUC of 0.77 in the forecasting task. Similarly, Dimitriadis [8] also used the ADNI database to create a new and unique four-class AD-based problem. By integrating morphological MRI-based features such as cortical thickness, subcortical volumes, and hippocampal subfields within a Random Forest

framework, the study achieved a 61.9% classification performance in distinguishing between four groups: Healthy Control, MCI, converted MCI, and AD. Doyle [9] forecasted the development of AD using multivariate ordinal regression to model the ordered brain deterioration from normal aging (CTL) to MCI to AD. Wang [16] developed a hybrid machine learning system that combines multiple convolutional neural networks and a linear support vector classifier. According to clinical evidence, convolutional neural networks were used to automatically extract image features from brain segments related to cognitive decline.

The linear support vector classifier then used the extracted image features and non-image information to make the final predictions. Recently, stacked ensembles have been successfully implemented in medical diagnostics and a variety of other fields. Stacked ensemble methods have improved the predictive performance of a model by combining the strengths of several base models and feeding their predictions into a higher-level, secondary model (meta-learner) to produce the final prediction.

The primary purpose of this technique is to blend the capabilities of numerous diverse models to mitigate individual model weaknesses, improve generalization, and enhance the overall predictive accuracy [13]. For example, stacked ensemble models have been utilized for neuropathologies to predict AD onset [1] by combining different machine learning algorithms.

In this project, a novel methodology for classifying Alzheimer's disease, normal cognition, and mild cognitive impairment was proposed using a stacking ensemble of classic machine learning models. This paper is structured as follows: Section 2 Methodology describes the database and the clinical data considered in the study, the processing and feature selection of the data. The stacked ensemble model is also reported in this section. Section 3 reports the performance and accuracy of the stacked ensemble. Finally, the conclusions are outlined in the last section of the document.

2 Methodology

The data analysis and model training for this study were carried out on a virtual computer with the following specifications: the operating system was a Linux distribution, the virtual machine architecture was x86_64, the full platform description was Linux-5.15.107+-x86_64-with-glibc2.31, the processor was an x86_64, the total CPU count was 2, and the system was equipped with a total memory of approximately 12.68 GB. Figure 1 depicts the general pipeline of the proposed stacking ensemble algorithm.

2.1 Database

In this project, multiple datasets from the ADNI database [15] were considered. The primary dataset of the study, the ADNIMERGE dataset encapsulates critical information from various phases of the ADNI project (ADNI1, GO, 2, 3) and it comprises 16,345 rows and 42 columns, capturing a broad spectrum of participants information across different stages of the disease.

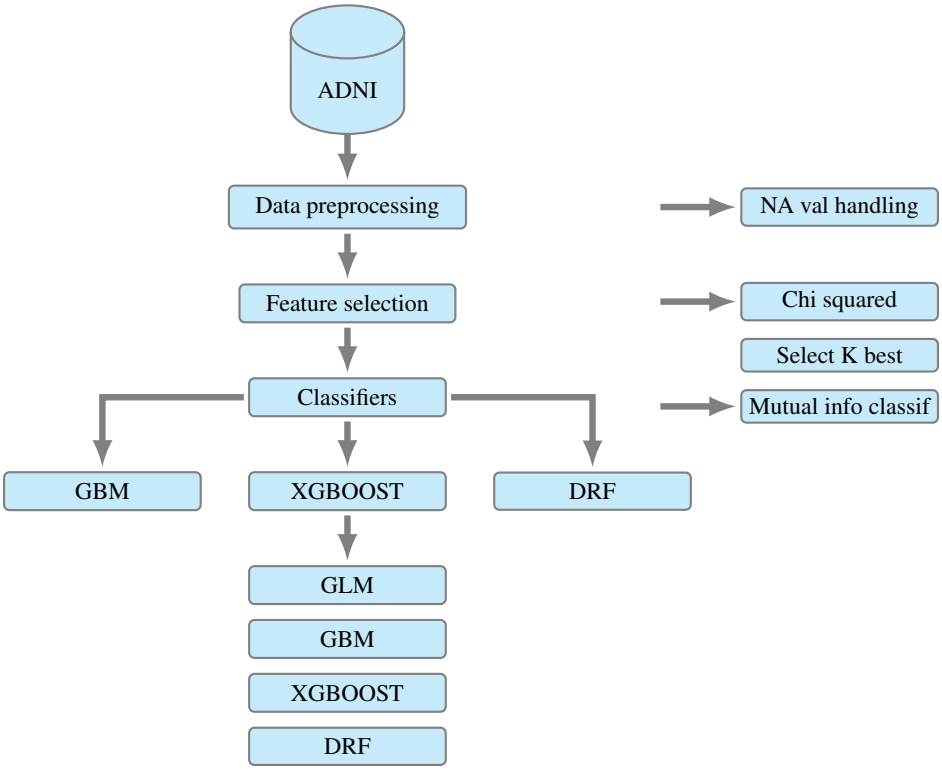

Fig. 1. General pipeline of the proposed stacking ensemble algorithm.

However, this dataset has the disadvantage of having a significant amount of missing data particularly in the 'ABETA,' 'TAU,' and 'PTAU' columns, with over 13,975 missing values each. In order to improve the dataset, an integration of the TOMM40 PolyT Variant Data and the Desikan Lab Polygenic Hazard Score (PHS) was made. The TOMM40 dataset, consisting of 1,520 rows and five columns, provides a clean and focused view of the TOMM40 gene. The PHS dataset, on the other hand, containing 757 rows and five columns, reveals minor data inconsistencies with 11 missing values each in the 'TOMM40_A1' and 'TOMM40_A2' columns.

These columns potentially represent alleles of the TOMM40 gene, enhancing the understanding of the genetic influence on Alzheimer's disease progression. After this process, the unified dataset encompassing approximately 2000 participants categorizes individuals into three cognitive states: Normal cognition, Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD). This diverse dataset improved the external validity and reliability of the machine learning models.

It is worth mentioning that a significant part of the study is centered around the analysis of Mild Cognitive Impairment (MCI), a transitional stage between normal cognitive aging and dementia, since the analysis of MCI helps identify the early stages of cognitive decline, capturing the subtle yet significant shifts that a person undergoes when they drift away from normal cognition.

The features utilized in this study are categorized into five primary groups: genetic data, protein data, radiopharmaceutical data, brain volumetric data, demographic data, and cognitive assessment data.

2.2 Data Preprocessing

During the initial stages of data preprocessing, a significant amount of missing values in the dataset was identified. This posed a significant problem, especially considering the requirement for developing a classifier that is not reliant on incomplete or artificially augmented data. Upon closer inspection, it was found that certain fields, such as 'FDG,' 'ABETA,' 'TAU,' 'PTAU,' and, at certain stages of the ADNI dataset, the brain volume data, contained fewer complete records compared to other variables.

For the data cleanup stage, which in part involved the elimination of rows with missing data, the loss of two entire classes: 'SMC' and 'EMCI,' was observed. Regarding genetic data, the 'APOE4' column was considered. This captures information about the presence of the APOE4 allele, which has been associated with an increased risk of Alzheimer's disease. In the category of proteic data, the columns 'ABETA,' 'TAU,' and 'PTAU' were included. These columns contain information about various Alzheimer-related proteins, which serve as biochemical markers for the disease.

For radiopharmaceutical data, the 'FDG' column was considered. Regarding imaging data, the columns 'Ventricles,' 'Hippocampus,' 'WholeBrain,' 'Entorhinal,' 'Fusiform,' 'MidTemp,' and 'ICV' were selected. These columns contained volumetric measurements of various brain regions and the overall intracranial volume. The demographic data considered the columns: 'AGE,' 'PTGENDER,' 'PTEDUCAT,' 'PTETHCAT,' 'PTRACCAT,' and 'PTMARRY .' Lastly, for cognitive assessment data, the following columns were included: 'CDRSB,' 'ADAS11', 'ADAS13', 'ADASQ4', 'MMSE,' 'RAVLT_immediate,' 'RAVLT_learning,' 'RA- VLT_forgetting,' 'RAV LT_perc_forgetting,' 'LDELTOTAL,' 'DIGITSCOR,' 'TRABSCOR,' 'FAQ.' and there were no experimental configurations discovered that could retain these two classes without a severe hindrance on the model's performance.

Another issue related to protein data and how it was stored in the CSV files was that the three relevant columns 'ABETA,' 'TAU,' and 'PTAU' had information written as a string when concentrations exceeded or did not reach a certain value. The adjustments performed were simply the swapping of specific string values to their closes numerical representation. In detail, the process was done as follows:

1. The data associated with the 'ABETA' protein was transformed by replacing any instances of values greater than 1700 and less than 200 with 1700 and 200, respectively.
2. Similarly, the 'TAU' protein data was adjusted by modifying the instances of values above 1300 and below 80, with 1300 and 80, respectively.
3. Finally, the data associated with the 'PTAU' protein was updated by replacing the occurrences of values exceeding 120 and falling below 8 with 120 and 8, respectively.

2.3 Feature Selection

For the feature selection stage, the 'SelectKBest' function combined with the 'mutual_info_classif' method, both from Scikit-learn's feature selection module were used. 'SelectKBest' is a univariate feature selection method that identifies the 'k' highest scoring features. The 'mutual_info_classif' method is designed to compute the Mutual Information (MI) between each feature and the target variable, in this case, 'DX_bl'. The mathematical formula [7] underpinning this method is as follows:

$$\mathrm{MI}(X,Y) = \sum_{x}\sum_{y} p(x,y)\log\left(\frac{p(x,y)}{p(x)\,p(y)}\right). \tag{1}$$

In this equation, X represents a feature, and Y symbolizes the target variable. $p(x,\, y)$ is the joint probability distribution function of X and Y, whereas $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y, respectively. Mutual information is beneficial as it measures the dependency between the variables and only returns a zero when two variables are found to be independent.

The relevance of each feature was determined by calculating Mutual Information (MI), which indicated the strength of its relationship with the target variable. These scores were then assessed, focusing especially on categorical and numerical features with non-zero mutual information. This non-zero value signified a degree of correlation with the target variable. This feature selection process allowed for a focus on the most relevant attributes, thereby improving the precision and efficiency of the predictive model.

2.4 Classification Model

The study implemented an ensemble model that utilized a meta learner algorithm based on generalized linear models (GLM) with a logit transformation. The ensemble was constructed using the following base models:

- **GBM_4 Model (Gradient Boosting Machine):**
  - Trained with 48 trees, a maximum depth of 10, and a learning rate of 0.1.
  - Utilized a multinomial distribution and employed a UniformAdaptive histogram type.
- **DRF_1 Model (Distributed Random Forest):**
  - Trained with 32 trees and a maximum depth of 20.
  - Utilized a multinomial distribution and employed a UniformAdaptive histogram type.
- **XGBoost_3 Model (eXtreme Gradient Boosting.):**
  - Trained with 40 trees, a maximum depth of 5, and a learning rate of 0.3.
  - Utilized a multinomial distribution and employed the exact tree method with a depthwise grow policy.

– **GLM_1 Model (Generalized Linear Model):**
  – Trained using a multinomial family with coordinate descent as the solver.
  – Employed lambda search with early stopping.
– **XRT_1 Model (Extremely Randomized Trees, Treated as DRF):**
  – Trained with 26 trees and a maximum depth of 20.
  – Utilized a multinomial distribution and employed a random histogram type.

XGBoost, short for "eXtreme Gradient Boosting," is an optimized implementation of a Gradient Boosting Machine (GBM). XGBoost improves upon the classic GBM framework by introducing regularization to avoid overfitting, as well as several system optimizations to speed up and improve the model's performance. In essence, the XGBoost algorithm works by iteratively adding new models to the ensemble that predict the errors of the previous models. The prediction [6] at each step is given by:

$$\mathcal{F}_m(x_i) = \mathcal{F}_{m-1}(x_i) + \langle_m(x_i), \tag{2}$$

where $\mathcal{F}_m(x_i)$ is the predicted output after the mth model, $\mathcal{F}_{m-1}(x_i)$ is the prediction from the previous step, and $\langle_m(x_i)$ is the current model that's added to improve the prediction by predicting the residuals of the previous model. Distributed Random Forest (DRF) models operate similarly to the standard Random Forest algorithm, with variations in their configuration to ensure diversity among the predictions of individual trees in the ensemble. A Random Forest or DRF model can be abstractly represented [3] as:

$$Y = \frac{1}{n}\sum_{i=1}^{n} T_i(X), \tag{3}$$

where Y is the output variable, X is the vector of input variables, $T_i(X)$ represents the prediction of the $i-th$ decision tree in the ensemble, and n is the total number of trees in the ensemble. The final output of the stacked ensemble model is a weighted sum of the individual model predictions and can be represented [4] as:

$$F(x_i) = \sum_{m=1}^{M} w_m\, F_m(x_i), \tag{4}$$

where, $F(x_i)$ is the final output, $F_m(x_i)$ is the output of the mth model, and w_m is the weight for the mth model. These weights are learned during training to optimize the ensemble's performance.

K-fold cross validation. Is a statistical method used for estimating the performance of predictive models. This type of validation is mostly used when a model's goal is prediction, and one wants to estimate its accuracy with as little bias as possible. This study employed a specific form of cross-validation called 5-fold cross-validation. In this approach, the dataset was divided into five equal-sized folds, and then the predictive model was trained and validated five times, with each iteration using a different fold for validation while the remaining folds were used for training.

Table 1. Feature importance and categories for the ensemble model.

Feature	Score_MI	Category
cdrsb	0.596834	Clinical
ldeltotal	0.439072	Neuropsychological
mmse	0.423307	Clinical
faq	0.422767	Clinical
adas13	0.418954	Clinical
adasq4	0.405963	Clinical
cir	0.345453	Biomarkers
adas11	0.329069	Clinical
phs	0.282552	Biomarkers
ravlt_perc_forgetting	0.261960	Neuropsychological
ravlt_immediate	0.236800	Neuropsychological
ravlt_learning	0.190921	Neuropsychological
fdg	0.187301	Imaging
ptau	0.174670	Biomarkers
trabsor	0.173784	Clinical
ravlt_forgetting	0.157009	Neuropsychological
digitscor	0.140957	Neuropsychological
hippocampus	0.133758	Imaging
apoe4_0.0	0.130713	Genetic
abeta	0.118638	Biomarkers
fusiform	0.115856	Imaging
tau	0.097753	Biomarkers
entorhinal	0.084071	Imaging
midtemp	0.079739	Imaging
tomm40_a1	0.078894	Genetic
apoe4_1.0	0.071801	Genetic
icv	0.059267	Imaging
ptmarry_married	0.053313	Demographic
tomm40_a2	0.051060	Genetic
wholebrain	0.034028	Imaging
ptraccat_white	0.029367	Demographic
ptmarry_widowed	0.024984	Demographic
ventricles	0.023164	Imaging
ptgender_female	0.021457	Demographic
ptethcat_not hisp/latino	0.014322	Demographic
ptmarry_never married	0.002360	Demographic
ptmarry_divorced	0.001434	Demographic

A mathematical representation for the average performance in 5-fold cross-validation can be expressed as:

$$E = \frac{1}{5}\sum_{i=1}^{5} E_i, \tag{5}$$

where E is the average performance across the folds, and E_i is the performance metric.

Table 2. Performance metrics of various classifiers.

Classifier	Mean Accuracy	SD	CV1	CV2	CV3	CV4	CV5
Logistic Regression	0.835486	0.038173	0.773585	0.865385	0.884615	0.826923	0.826923
Random Forest	0.724311	0.046802	0.679245	0.769231	0.788462	0.673077	0.711538
Support Vector Machine	0.853661	0.030135	0.828125	0.831169	0.888889	0.836066	0.884058
Gradient Boosting	0.770174	0.039931	0.754717	0.730769	0.846154	0.769231	0.750000
XGBoost	0.839260	0.034568	0.792453	0.884615	0.846154	0.865385	0.807692
Ensemble Classifier	0.869884	0.021199	0.830189	0.884615	0.884615	0.884615	0.865385

3 Results

3.1 Data Preprocessing

As previously mentioned, the data employed for this study is made of multiple datasets from the ADNI database. In order to be able to work with it, the clean-up of the data plays an important role in the implementation of the staking ensemble. After carrying out the preprocessing, the database went from having 16345 incomplete rows, 42 columns, and five classes (LMCI, CN, AD, EMCI, SMC) to having 411 rows, 36 columns, and three classes (LMCI, CN, AD). Of those additional columns, 4 correspond to integrating the TOMM40 PolyT Variant Data and the Desikan Lab Polygenic Hazard Score (PHS) associated information.

3.2 Feature Selection

As a second step, a feature selection was performed on the 'clean' database to improve the performance of the ML algorithm. According to the results that can be seen in Table 1, cognitive test data (labeled as clinical) was shown to have the highest MI scores, meaning that these features have a substantial impact on the model's predictive accuracy, follow up by neuropsychological data.

Conversely, demographic features demonstrated the lowest MI scores, indicating a lesser contribution to the model's predictions, and although these factors did contribute to some extent, their impact was not as pronounced as that of the cognitive tests. This result is consistent with what is reported by medical specialists, which gave a higher weight to clinical, neuropsychological, imaging, biomarkers, and generic data.

3.3 Classification Model

After the feature selection stage, a stacked ensemble model was implemented; this model ensured a robust prediction method by leveraging the strengths of different machine learning algorithms. For this specific case a GBM, XGBoost, and DRF, combined with a powerful meta-learner (GLM), optimally enhanced these base models predictions. Table 2 shows the performance of the stacked ensemble as a validation method a k-fold cross-validation with $k = 5$ was employed, obtaining an accuracy of 86.9%.

4 Conclusion

This study was done using the ADNI database and its multiple datasets. Two main preprocessing steps were performed due to several data utilization problems. Specifically, a large portion of incomplete rows needed to be eliminated, and several string records in the protein columns needed to be replaced by their closest numerical representation.

In the feature ranking analysis conducted on the preprocessed dataset using the random forest algorithm, cognitive examination data emerged as the most significant predictor for Alzheimer's disease. This was closely followed by indicators such as the presence of the ptau protein and volumetric measurements of the hippocampus, a region notably affected in Alzheimer's pathology.

The pronounced significance of the examination data can be attributed to its direct and intrinsic nature. While various biomarkers and neuroanatomical measurements provide valuable insights into the disease's progression and manifestations, direct cognitive assessments capture the immediate and functional impact of the disease on an individual's cognitive abilities. As such, by their very nature, these examinations are poised to inherently possess greater diagnostic relevance than indirect predictors.

In the experimental phase, several configurations were tested. The most promising results were obtained when all features were considered. The optimal stacking ensemble architecture consisted of seven foundational models: three Gradient Boosting Machines (GBM), two Extreme Gradient Boosting models (XGBoost), and two Distributed Random Forests (DRF). Evaluated using a 5-fold cross-validation method, this model configuration achieved an overall accuracy of 86.9%.

Acknowledgments. The authors acknowledge ADNI for database access; CONAHCYT (grant number 1239365) and Tecnologico de Monterrey for their financial support.

References

1. Alatrany, A. S., Hussain, A., Jamila, M., Al-Jumeiy, D.: Stacked machine learning model for predicting Alzheimer's disease based on genetic data. In: 14th International Conference on Developments in eSystems Engineering, IEEE, pp. 594–598 (2021) doi: 10.1109/DeSE54285.2021.9719449
2. Beltrán, J. F., Wahba, B. M., Hose, N., Shasha, D., Kline, R. P.: Inexpensive, non-invasive biomarkers predict Alzheimer transition using machine learning analysis of the Alzheimer's disease neuroimaging (ADNI) database. Public Library of Science ONE, vol. 15, no. 7 (2020) doi: 10.1371/journal.pone.0235663
3. Breiman, L.: Random forests. Machine Learning, vol. 45, no. 1, pp. 5–32 (2001) doi: 10.1023/A:1010933404324
4. Brownlee, J.: How to develop a weighted average ensemble with python. Ensemble Learning (2021) machinelearningmastery.com/weighted-average-ensemble-with-python/
5. Castellani, R. J., Rolston, R. K., Smith, M. A.: Alzheimer disease. Disease-a-Month, vol. 56, no. 9, pp. 484–546 (2010) doi: 10.1016/j.disamonth.2010.06.001

6. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd Association for Computing Machinery, Special Interest Group on Knowledge Discovery in Data, International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016) doi: 10.1145/2939672.2939785
7. Cover, T. M., Thomas, J. A.: Elements of information theory. John Wiley and Sons, Inc (2005) doi: 10.1002/047174882x
8. Dimitriadis, S. I., Liparas, D., Tsolaki, M. N.: Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate among healhy elderly, MCI, cMCI and Alzheimer's disease patients: From the Alzheimer's disease neuroimaging initiative (ADNI) database. Journal of Neuroscience Methods, vol. 302, pp. 14–23 (2018) doi: 10.1016/j.jneumeth.2017.12.010
9. Doyle, O. M., Westman, E., Marquand, A. F., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Lovestone, S., Williams, S. C. R., Simmons, A.: Predicting progression of Alzheimer's disease using ordinal regression. Public Library of Science ONE, vol. 9, no. 8 (2014) doi: 10.1371/journal.pone.0105542
10. Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., Cummings, J. L., de-Leon, M., Feldman, H., Ganguli, M., Hampel, H., Scheltens, P., Tierney, M. C., Whitehouse, P., Winblad, B.: Mild cognitive impairment. The Lancet, vol. 367, no. 9518, pp. 1262–1270 (2006) doi: 10.1016/s0140-6736(06)68542-5
11. Geldmacher, D. S., Whitehouse, P. J.: Evaluation of dementia. New England Journal of Medicine, vol. 335, no. 5, pp. 330–336 (1996) doi: 10.1056/NEJM199608013350507
12. Lee, M., Chodosh, J.: Dementia and life expectancy: What do we know? Journal of the American Medical Directors Association, vol. 10, no. 7, pp. 466–471 (2009) doi: 10.1016/j.jamda.2009.03.014
13. Pavlyshenko, B.: Using stacking approaches for machine learning models. In: IEEE Second International Conference on Data Stream Mining and Processing, pp. 255–258 (2018) doi: 10.1109/DSMP.2018.8478522
14. Petersen, R. C.: Mild cognitive impairment. CONTINUUM: Lifelong Learning in Neurology, vol. 22, no. 2, pp. 404–418 (2016) doi: 10.1212/con.0000000000000313
15. University of Southern California: ADNI database. image and data archive laboratory of neuro imaging (2023) ida.loni.usc.edu/
16. Wang, C., Li, Y., Tsuboshita, Y., Sakurai, T., Goto, T., Yamaguchi, H., Yamashita, Y., Sekiguchi, A., Tachimori, H.: A high-generalizability machine learning framework for predicting the progression of Alzheimer's disease using limited data. npj Digital Medicine, vol. 5, no. 1, pp. 43 (2022) doi: 10.1038/s41746-022-00577-x
17. World Health Organization: Risk reduction of cognitive decline and dementia: WHO guidelines (2019) www.who.int/publications/i/item/9789241550543