Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 1 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i01.58189

PAPER

Symmetry-Aware Machine Learning for the Diagnosis of Alzheimer's Disease and Detection of Mild Cognitive Impairment Using Biomarkers

Orlando Iparraguirre-Villanueva[1] (✉),
José Luis Herrera Salazar[1],
Gloria Castro-Leon[1],
Hernán Ochoa-Carbajal[1],
Henry Chero-Valdivieso[2],
Rosalynn Ornella Flores-Castañeda[3]

[1]Universidad Nacional Tecnológica de Lima Sur, Lima, Peru

[2]Universidad César Vallejo, Lima, Peru

[3]Universidad San Ignacio de Loyola, Lima, Peru

oiparraguirre@untels.edu.pe

ABSTRACT

Alzheimer's disease (AD) is one of the most common causes of dementia in older adults, and there is currently no cure for this disease. Early detection can be very beneficial for patients, as it allows them to slow the progression of symptoms and improve their quality of life. This is where technology comes into play, especially artificial intelligence (AI), which can help doctors and nurses work faster and make better diagnoses. The goal is to test the performance of six machine learning (ML) algorithms—k-nearest neighbors (KNN), decision tree (DT), logistic regression (LR), random forest (RF), support vector machines (SVM), and Naive Bayes (NB)—to examine biomarkers and help diagnose AD and mild cognitive impairment (MCI). The dataset included 212 patients, 91 of whom had AD, 86 had MCI, and 35 showed no signs of the disease. The stages of the process were preprocessing, exploratory analysis, training, testing, and validation. DT and RF models achieved the best performance, with accuracy of 0.75 and 0.73, sensitivity of 0.75 and 0.72, and F1-scores of 0.75, respectively. LR obtained the highest MCC with 0.54. This demonstrates that ML models can be very useful for making better diagnoses of AD and MCI, especially when medical resources are limited. Finally, the DT and RF models demonstrate that applying symmetry in model training and performance metrics results in tools that can accelerate clinical translation.

KEYWORDS

Alzheimer's disease (AD), machine learning (ML), diagnosis, cognitive degeneration

1 INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disease that significantly affects key cognitive functions such as memory, thinking, and the ability to perform daily activities [1]. AD has become a major public health challenge, affecting millions of people and placing an increasing burden on healthcare

Iparraguirre-Villanueva, O., Salazar, J. L. H., Castro-Leon, G., Ochoa-Carbajal, H., Chero-Valdivieso, H., Flores-Castañeda, R. O. (2026). Symmetry-Aware Machine Learning for the Diagnosis of Alzheimer's Disease and Detection of Mild Cognitive Impairment Using Biomarkers. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(1), pp. 21–39. https://doi.org/10.3991/ijoe.v22i01.58189

Article submitted 2025-08-09. Revision uploaded 2025-09-18. Final acceptance 2025-09-19.

systems [2], [3]. The Alzheimer's Foundation Europe and Alzheimer's International Network of Associations argue that as the population ages, the risk of dementia increases. It is therefore essential to identify people at risk of developing dementia by implementing interventions in the early stages of the disease to safeguard them from damage that cannot be reversed [4], [5]. It is also important to note that, according to the World Health Organization (WHO), more than 55 million people worldwide live with some form of dementia, and more than 60% of cases are concentrated in low-income countries [6].

The growing need to address AD has led the medical community to explore new treatment techniques, many of which are supported by information technologies. Current advances aim to achieve more accurate diagnoses in the early stages of AD, where timely intervention can become decisive [7].

Alzheimer's disease, as such has become a priority for both health professionals and researchers [7]. The search, for new diagnostic methods is leading scientists to explore unknown scenarios, using advanced tools such as structural and functional MRI, electroencephalography (EEG), and blood biomarker analysis. This breakthrough is not only a major step forward for medicine but also a major leap in the fight with AD [8], [9]. Detecting AD in its early stages offers a double advantage: on the one hand, it allows us to intervene in time, and on the other hand, it helps us during the process to discover the mechanisms that cause the disease [10], [11]. In this work we focused on performing a comprehensive analysis of biomarkers and evaluating how machine learning (ML) algorithms (decision tree (DT), logistic regression (LR), random forest (RF), k-nearest neighbors (KNN), support vector machines (SVM), and Naive Bayes (NB)) can improve the diagnosis of AD and mild cognitive impairment (MCI). With this, we seek to have a direct impact on the lives of patients, allowing early and efficient interventions.

Now-a-days, AI tools and ML algorithms play a very important role in the healthcare field, mainly in the fight against AD. For example, in the study [12], they demonstrated how ML techniques manage to detect with great ease hidden patterns in neuroimages, allowing specialists to classify more clearly the different stages of MCI. However, the true value of these tools is harnessed when they are integrated with other types of data, creating strategies to identify Alzheimer's more effectively. A clear example of this case is presented in [13], where they used ML models for early detection of AD. This study demonstrated how the combination of structured and functional data improves the accuracy of disease diagnosis, making it possible to offer options for blood-based diagnostic tests for Alzheimer's disease.

Also, the authors in the paper [14] integrated ML algorithms with biomarkers to diagnose AD. After analysis, the researchers were able to demonstrate that certain ML algorithms can improve both the sensitivity and specificity of diagnosis. Following this line of research, in [15] they evaluated an ML algorithm based on EEG, a non-invasive technique to record the electrical activity of the brain, to differentiate Lewy bodies and MCI in Japanese patients. With a sample of 18 patients with MCI and 21 with Alzheimer's disease, the algorithm achieved a sensitivity of 72.2%, a specificity of 85.7%, and an accuracy of 79.5%. The results highlighted the great potential of EEG as a diagnostic tool. But it should be kept in mind that ML is not only limited to diagnosis. And this was reflected in the work [16] that used ML algorithms to define AD subtypes and genes associated with the immune microenvironment. The findings showed that the genes were tightly linked to disease progression, offering clues for future treatments. Similarly, in [17] they investigated how new biomarkers and ML techniques could improve the diagnosis of AD. Finally, the researchers concluded that the combination of innovative biomarkers and ML could significantly improve the sensitivity and specificity of diagnosis.

Other important findings can be seen in the work [18], where they explored the potential of five metabolites in blood plasma. Their objective was to identify MCI and Alzheimer's disease, for which they used metabolomic and ML techniques. They collected samples from different regions of the United States, and by integrating proton magnetic resonance imaging and ML, they were able to identify biomarkers with significant results. The metrics obtained in the area under the curve (AUC) were significant: the accuracy was between 72% and 76%, the sensitivity was between 75% and 85%, and finally, the specificities were in between the range of 69% and 81%. These metrics support the robustness of ML models for this type of task. Finally, in [19] they evaluated the effectiveness of six supervised learning ML models for predicting AD progression. For this, they used EEG sensors to analyze brain signals. The results were more than encouraging, the DT algorithm achieved an accuracy of 0.785 in detecting AD, while RF achieved an accuracy of 0.863 in detecting frontotemporal dementia. These results confirm the potential offered by models such as DT and RF for the early diagnosis of Alzheimer's disease.

Mild cognitive impairment is frequently considered an early stage of AD; therefore, the joint study of both conditions allows for the identification of crucial diagnostic indicators throughout the disease's progression. This study uses ML models to evaluate the effectiveness of biomarker information in distinguishing between AD and mild cognitive impairment.

The aim of this work is to perform an in-depth analysis of biomarkers and to evaluate how ML algorithms can improve the diagnosis of AD and MCI. Also, it seeks to explore the most recent advances in Alzheimer's research.

For ease of reading, the paper is organized as follows. The architecture of the ML models to be used in the study is described in Section 2. In addition, data processing techniques are presented, and the metrics to be used to evaluate the models are described. Section 3 presents the results that can be drawn from research. Section 4 provides a detailed analysis of the results, emphasizing the context of the research. Finally, the conclusions of the study are discussed, emphasizing their importance for future research.

2 MATERIALS AND METHODS

This section presents the theoretical basis of the ML models that are key to the development of this case: DT, LR, RF, KNN, SVM, and NB. With the aim of providing a clear understanding of how these models work in their practical application in the diagnosis of Alzheimer's disease.

The six ML algorithms selected for this study come from different algorithmic families: DT and RF (tree-based), LR (linear), NB (probabilistic), KNN (instance-based), and SVM (margin-based). They were chosen to give a robust benchmark comparison for diagnostic modeling.

Advanced techniques such as ensemble amplification (e.g., XGBoost) or neural networks can make predictions more accurate but often do so at the expense of ease of understanding. This study focused on models that balance performance with transparency due to clinical focus and the possibility of using them in low-resource settings. A comparison of the models can be seen in Table 1.

Table 1. Model comparison

Model	Algorithm Type	Interpretability	Computational Cost	Use in Biomedical Research
DT	Tree-based (Single)	High	Low	Common
RF	Ensemble of trees	Moderate	Moderate	Very Common
LR	Linear (Parametric)	High	Low	Very Common
NB	Probabilistic (Bayes theorem)	Moderate	Low	Common
KNN	Instance-based (distance)	Low	High	Common
SVM	Margin-based	Moderate	Moderate	Common

2.1 Model based on hierarchical decisions

The hierarchical DT model, also known as a DT, is one of the most widely used tools in the field of ML, mainly in classification tasks [20]. The popularity of this model lies in its simplicity when it comes to organizing information into a series of nodes, where each hierarchical structure represents a decision based on the attributes of the data [21]. Like all other ML models, the DT model uses metrics for its evaluation, such as accuracy, precision, recall, f1-score, and receiver operating characteristics (ROC) curve. Building a DT seems very simple; however, it has some important aspects, such as the calculation of entropy, which is a measure of the disorder in the data. For this purpose, two types are used: 1) the first type of entropy, which is calculated using equation (1), which considers the frequency of an attribute; 2) the second type of entropy: equation (2) is used for the calculation, which considers the frequency of two attributes.

$$H(D) = \sum_{k=1}^{k} - P_k \log 2 P_k \tag{1}$$

$$H(T, X) = \sum_{c \varepsilon X} P(c) H(c) \tag{2}$$

2.2 Logistic regression as a classification technique

The LR model, such as DT, is a widely popular model in the analysis of large data sets and in predicting binary outcomes. That is, in situations where there are only two outcomes (sick or healthy). What makes this model interesting is its ability to exploit the outcomes of the predictor variables and convert them into probabilities [22]. The selection of the predictor variables is of utmost importance for the processing; the quality and relevance of these variables must be considered in the selection phase, since they directly influence the results [23]. The LR model relies on a mathematical function to calculate the probability of a specific event. This function allows transforming any value to a range of 0 and 1, which is ideal for working with probabilities [24]. An important element to be considered in the LR model is the linear relationship between the predictor variables and the probability algorithm. This relationship is important to understand how the model analyzes the data and makes predictions [25]. In mathematical terms, equation (3) describes how the LR model works.

$$\int(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

2.3 Classification methods: RF and KNN

The RF model, like the other previously mentioned models, has great popularity in the ML field, specifically in classification tasks. What makes this model special is that it combines multiple DTs to find the most appropriate accuracy, thus reducing overfitting if it occurs [26]. Now, to determine whether the model is working well, the same symmetric metrics are used as in other ML models (precision, accuracy, recall, F1 score, and ROC curve) [27]. A very important advantage of RF is its ability to handle a wide variety of variables, including nonlinear ones. A further advantage is its ability to tolerate outliers, since it eliminates them primarily without the need for data preprocessing [28].

The KNN algorithm is also a widely used algorithm in classification tasks such as regression. It works quite simply to predict values at data points. What this algorithm does is to search for the k nearest neighbors within a set of data based on a distance measure such as Euclidean [29]. The most relevant feature of this algorithm is that it does not make assumptions about the distribution of the data [30], this makes it more flexible and adaptable to different types of data, allowing gradual learning, i.e., it does not require prior training, but it learns from the data in real time during the prediction [31], [32]. It is very important to note that this algorithm, despite its flexibility and adaptability, if working with large volumes of data, could become very expensive computationally. This is because, for the prediction, the algorithm must calculate the distance between the queried point and the other points in the data set [33]. The KNN algorithm uses the Euclidean distance formula, as represented in equation (4).

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2} \tag{4}$$

2.4 SVM and naïve bayes

Support vector machines is among the most widely used techniques in ML, particularly in classification and regression problems. It is very flexible in its operation, as its main objective is to find the optimal hyperplane that best separates the classes within a multidimensional plane [34]. Algorithms such as SVM have multiple uses, including image and text analysis. They are also characterized by their ability to process complex, high-dimensional data, making them the best candidate for this type of task. Such algorithms as SVM are used in a wide range of applications, mainly image and text classification in the medical field. For example, SVM has been used to analyze medical data and predict disease [35].

Meanwhile, the NB algorithm, named after Bayes' theorem, is a classic in the field of ML. Although its approach seems simple, the truth is that NB over time has proven to obtain very good results in classification tasks, mainly when dealing with text [36]. An important feature of the NB algorithm is its computational efficiency in handling large volumes of data with great ease, which makes it ideal for applications that are managed in real time [37]. For example, NB is very popular in SPAM filtering, sentiment analysis, and document classification tasks.

2.5 Performance metrics

Evaluation metrics are key tools for measuring how well ML models perform in classification, regression, or other tasks. These metrics allow us to present the results in four main categories, known as the confusion matrix; among them we have. True Positives (TP), Cases in which the model correctly classifies an instance as part of the positive class. True Negatives (TN), When the model correctly recognizes an instance belonging to the negative class. False Positives (FP): It occurs when the model incorrectly assigns an instance to the positive class. False Negatives (FN): These are situations in which a negative instance is mistakenly classified as part of the positive class. From these categories several metrics are derived, such as: Recall, Precision, Accuracy, AUC and F1-Score. These metrics help us to better understand the performance of the models and are represented by equations (5)–(10), and (11).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{6}$$

$$Recall = \frac{TP}{TP + FP} \times 100 \tag{7}$$

$$F1 - Score = 2 * \frac{Pre + Rec}{Pre * Rec} * 100 \tag{8}$$

$$AUC = 0.5 * (TPR + TNR) * 100 \tag{9}$$

$$Macro\ avg = \frac{1}{N} \sum_{i=1}^{N} (Metric_i) \tag{10}$$

where N is the number of classes, and Metric$_i$ refers to the metric value for class i.

$$Weighted\ avg = \sum_{i=1}^{N} (W_i)\ Metric_i \tag{11}$$

where W_i is the proportion of samples from class i in the test set.

2.6 Description of the data set

A dataset obtained from the kaggle repository, a platform recognized for hosting research databases, was used for this work. This dataset is composed of 212 patients diagnosed with AD and MCI and includes ten key variables for analysis. The objective is to analyze the relationship between variables such as age, Mini-Mental State Examination (MMSE) scores, cerebrospinal fluid (CSF) biomarkers, amyloid concentration, total tau protein, phosphorylated tau, and the presence of the APOE4 allele in relation to these conditions.

It is very important to highlight the importance of these variables. For example, age: it is no secret that neurodegenerative diseases such as Alzheimer's are closely related to aging, analyzing age differences between AD and MCI groups can help us identify patterns and thus understand how these diseases are age dependent. Another variable analyzed is the MMSE, which, is a widely used test that helps to assess cognitive status by comparing scores between AD and MCI patients, and thus it is possible to have a broader look at the degree of cognitive impairment in patients. Also, there is the CSF biomarker variable, which, when studied allows us to identify the biochemical patterns that differentiate AD and MCI. Amyloid levels indicate the accumulation of amyloid plaques, while total tau and phosphorylated tau reflect the formation of neurofibrillary tangles, both of which are key markers of AD. Finally, the variable APOE4 allele—it is essential to analyze its presence in patients with AD and MCI to have a better understanding of its role. Table 2 presents in detail the characteristics that make up the dataset.

Table 2. Description of the characteristics of the dataset

#	Features	Description
1	Diagnostic labels: 3 categories	AD, Mild Cognitive Impairment, Control
2	Sex–gender of the subject, 2 categories	Female/Male
3	Age	Patient's age (in years)
4	MMSE test	MMSE is a cognitive assessment used to measure cognitive functions such as memory, language, attention, and visuospatial skills.
5	CSF Amyloid	CSF amyloid concentration
6	CSF Total tau	Total tau protein (t-tau) level in CSF
7	CSF Phosphorylated tau	Phosphorylated tau (p-tau) level in cerebrospinal fluid
8	APOE4 - APOE ε4 allele (ApoE4), 2 categories:	Yes/No
9	Progression to AD	Indicates whether the patient is progressing towards the disease.
10	Progression time	Months to progression

Table 3 presents the structure of the dataset with some patient records, where the following characteristics are considered: Diagnostic, Sex, Age, MMSE (Mini-Mental State Examination), CSF Amyloid (pg/mL), CSF Total tau (pg/mL), CSF Phosphorylated tau (pg/mL), APOE4, Progression to AD and Progression time (months).

Table 3. Data set information

	Diagnostic	Sex	Age	MMSE	CSF Amyloid (pg/mL)	Total tau in CSF (pg/mL)	Phosphorylated Present in CSF (pg/mL)	APOE4
01	Alzheimer's Disease	Female	68	26	688	369	107	Yes
02		Female	66	22	489	482	176	Yes
03		Female	72	23	509	329	114	Yes
04		Male	73	25	544	482	80	No
05		Female	75	15	303	806	120	No
…	…	…	…	…	…	…	…	…
212	Control	Male	69	22	291	288	45.2	Nan

2.7 Exploratory data analysis

Figure 1 shows the frequency of each disease concerning the total number of samples. As evidenced in Figure 1, AD is the most common, with 91 samples, followed by MCI, with 86 samples. The control, representing people with no disease, has 35 samples. These results from the data set indicate that AD is a degenerative disease that affects older people. MCI is a condition that can precede AD, and it is also more common in older people.

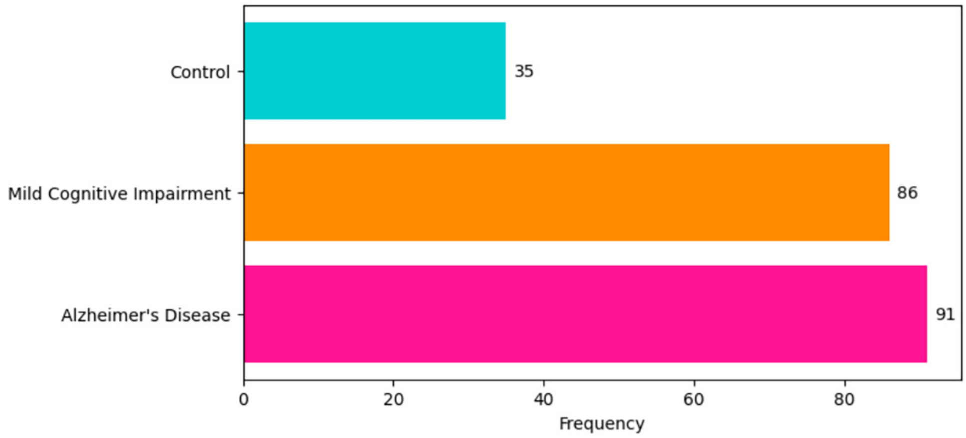

Fig. 1. Sample distribution

Although the original data set has three diagnostic classes (AD, MCI, and Control), for the final analysis the Control class is excluded. This choice was based on two factors: (1) the low number of samples in the Control group, which caused an imbalance in the partitioning of the data; and (2) the observation that its inclusion led to a bias in the models toward simpler classifications, making it much less sensitive to MCI detection. By examining only AD and MCI, we want to get a clearer picture of the models' ability to differentiate closely related clinical states, which is more useful for early diagnosis.

In Figure 2, a comparison of the age between patients with AD and patients with MCI is made. In this, the mean age of patients with AD is 75.4 years, while the mean age of patients with MCI is 72.7 years. These results show that patients with AD tend to be a little older than patients with MCI. This may be because AD is a neurodegenerative disease that progresses over time. As the disease progresses, patients may lose the ability to think and remember efficiently, which can make it difficult to care for themselves and perform daily activities. It is important to note that these data are based on a median sample. Further studies are needed to confirm that there are significant differences in age between AD and MCI patients. Also, it is appreciated that the two groups showed equal variances, which provides confidence in the validity of the t-test results.

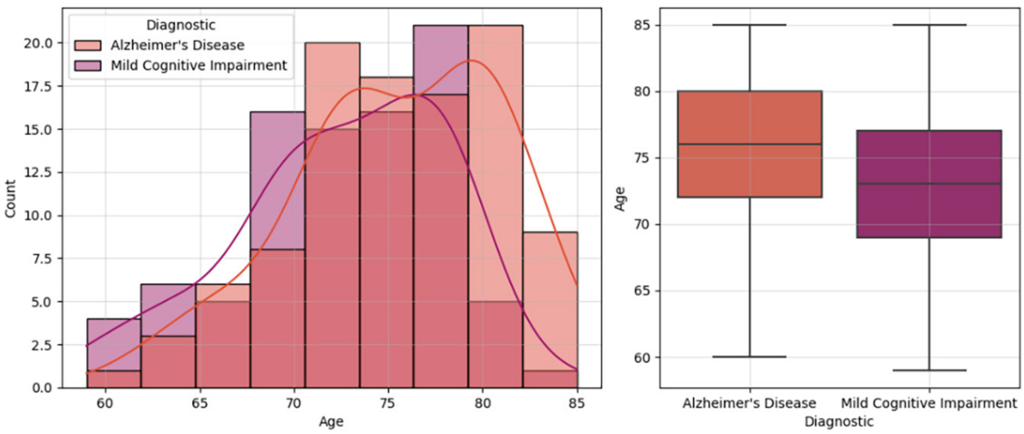

Fig. 2. Age comparison between AD and MCI patients

Figure 3 presents the relationship between CSF concentration and a memory test score in patients with AD. Figure 3 shows that patients with higher CSF amyloid concentration tend to have lower memory test scores. In other words, there is a positive correlation between amyloid concentration in CSF and cognitive impairment in AD patients. This proposes that CSF amyloid concentration may be a useful marker for the diagnosis and follow-up of Alzheimer's disease.

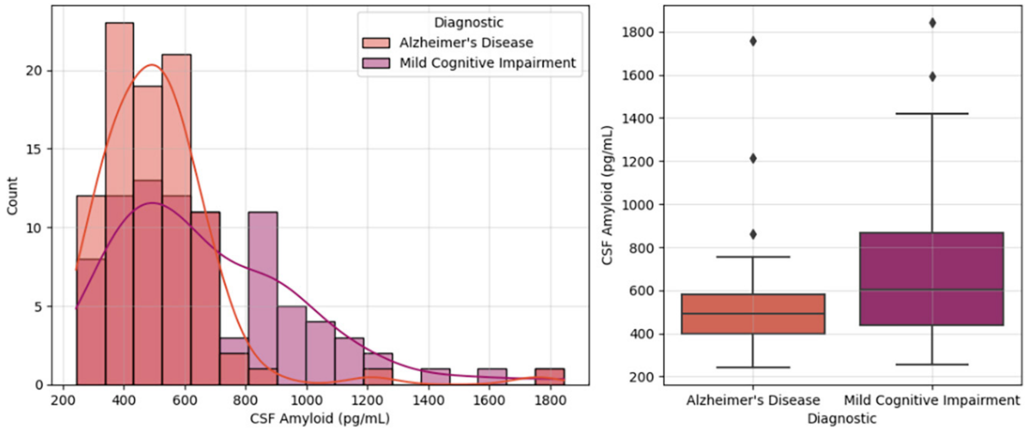

Fig. 3. Comparison of CSF amyloid concentration between AD and MCI patients

Also, it can be seen in Figure 3 that CSF amyloid concentrations in individuals with AD and MCI are significant. Table 4 presents the data, where it is determined that the mean CSF amyloid concentration in the ED group is 508.38 ± 201.80, while in the MCI group, it is 676.88 ± 309.80. The significant difference that exists in the variances highlights the heterogeneity in the distribution of CSF amyloid concentrations between the two diagnostic groups. This analysis provides evidence of the difference in CSF amyloid concentrations between individuals with AD and those with mild cognitive impairment.

Table 4. Statistical summary of AD with MCI

Diagnostic	Number of Cases	Mean	Standard Deviation	Mini Value	to 25%	to 50%	to 75%	Max Value
Alzheimer's Disease	91	508.38	201.7	243	396.5	493	580.5	1758
Mild Cognitive Impairment	87	675.82	308.16	257	438	604	868.5	1845

2.8 Correlation matrix of variables associated with diagnosis

After analysis, five characteristics of the dataset were identified as having significant relationships with the diagnostic labels. These characteristics include "Age," "MMSE score," "CSF amyloid level (pg/mL)," "Total CSF tau concentration (pg/mL)," and "CSF phosphorylated tau (pg/mL)." In Figure 4, each cell in the matrix shows the correlation coefficient between two variables. A value of 1 indicates a perfect correlation, while a value of 0 implies a complete absence of correlation. For example, the correlation between the variables age and MMSE score is 0.71. The correlation between age and amyloid level is 0.16, between age and total tau is 0.27, and between age and phosphorylated tau is 0.34.

The MMSE score and CSF amyloid concentration are 0.57, while its association with total CSF tau concentration reaches a value of 0.61. On the other hand, the correlation between MMSE and CSF phosphorylated tau concentration is 0.66. Likewise, a correlation of 0.63 is observed between CSF amyloid concentration and total CSF tau concentration, and a correlation of 0.69 between CSF amyloid concentration and phosphorylated tau concentration. Finally, it can be observed that the highest correlation is found between total tau concentration and CSF phosphorylated tau concentration, with a value of 0.71.

These five variables are the ones that will be used in training and testing, as they turned out to be the most important through correlation analysis.

The analysis also indicates that there is a relationship between age, MMSE score, CSF amyloid concentration, and CSF total tau concentration. Particularly, the variable age shows a negative correlation with the MMSE score, indicating that older people tend to obtain lower values in the cognitive test. Also, age correlates positively with CSF amyloid concentration. This indicates that levels of this biomarker increase with increasing age.

Cerebrospinal fluid amyloid concentration is also positively correlated with both total tau and phosphorylated tau. This clearly indicates that patients with higher levels of amyloid also tend to have higher levels of tau.

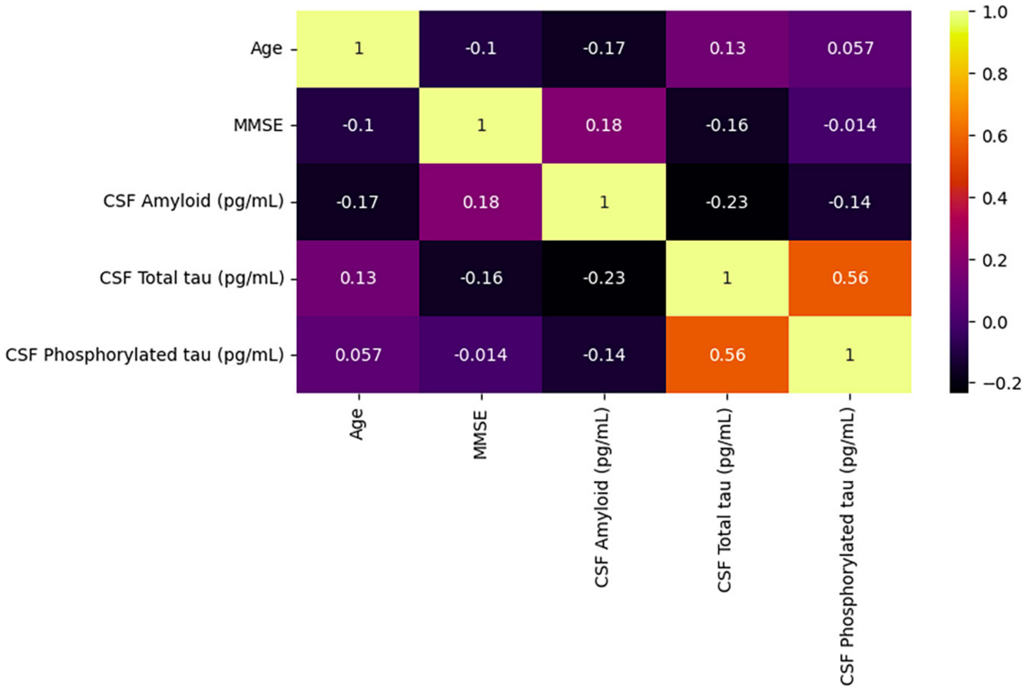

Fig. 4. Correlation matrix of statistically significant variables associated with diagnosis

2.9 Preprocessing and training of the dataset

The dataset preprocessing phase is crucial, as its goal is to prepare the data for model training. This includes 1) data cleaning, in this case, Python libraries were used to remove data. For example, three incomplete records were removed, initially there were 215, leaving 212 valid samples. 2) Normalization of the data, in this case, the variables were normalized to have a similar distribution. For example, the variables were transformed using Z-score normalization to ensure that they had a mean equal to zero and a standard deviation of one. The variables were then coded, transforming categorical variables into numerical variables. Thus, the binary variable "APOE4" was coded as 1 when present and 0 when absent. One-hot coding was not applied because no variables with more than two categories were found.

1. In the model training phase, the hyperparameters of each model were optimized to fit the training data. Then we proceeded to divide the data set into training data and test data, in 80/20. Next, we proceeded to write the code for each of the models: DT is a simple algorithm very useful for classification problems with relatively small data sets. RF is a variant of the DT algorithm, which is more accurate and robust. LR is a linear model useful for binary classification problems. NB classifier is a statistical model useful for classification problems with relatively small data sets. KNN is a model that is based on the distance between the training data and the test data. And finally, SVM is a model that is based on the separation of attributes in a hyperplane.
2. To improve the results and minimize the dependence on a single partition of the data set, a stratified K-fold cross-validation was applied. This technique to some extent ensures that each fold maintains the same proportion of classes (AD and MCI), which reduces imbalance-induced bias. In each iteration, the model is trained with 80% of the data and validated with the remaining 20%, repeating this process. The final metrics presented are the means of the results obtained in each fold, together with their corresponding standard deviation.

2.10 Performance evaluation and model comparison

At this stage, the performance of the models was analyzed using the test data; the metrics used were as follows: accuracy, sensitivity, F1 score, and AUC-ROC. Figure 5 shows the results of the classification of patients with AD and MCI using the six ML algorithms: DT, RF, LR, NB, KNN, and SVM.

Accuracy seeks to measure samples correctly classified as having AD or MCI. Sensitivity measures the percentage of patients with AD who were correctly classified, while specificity reflects the proportion of MCI patients correctly classified.

The findings presented in Figure 5 indicate that the DT and RF models were more accurate in this case, reaching an accuracy of approximately 75%. On the other hand, the LR and NB models obtained a slightly lower accuracy, with an accuracy of 72%. It is very important to note that these results are for this specific case, since, in another context, even with the same data set or other ML techniques, the results may be different. Although ML models can help in the classification of patients with AD and MCI, it is important to recognize their limitations in these results. For example, the DT model achieved a slightly higher accuracy rate than the others, but its simplicity makes it susceptible to variations in the symmetry of class distribution within the dataset.

Symmetry approach to model evaluation. In this paper, "symmetry" is used as a methodological principle to evaluate model performance. Symmetry is operationalized through the combined use of several metrics such as accuracy, recall, specificity, F1-score, AUC-ROC, and MCC. The objective of this method is to avoid biases that can be generated when optimizing a single metric, especially in problems with unbalanced classes such as the differential diagnosis between AD and mild cognitive impairment.

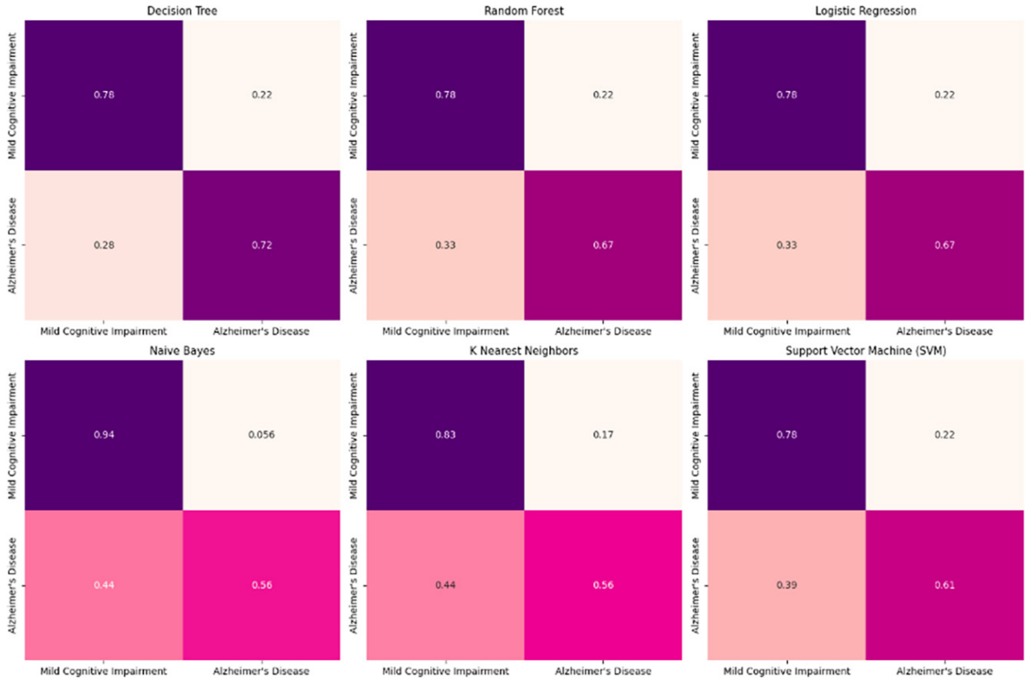

Fig. 5. Confusion matrix of 6 classifications models

2.11 Hyperparameter optimization

The GridSearchCV function of Scikit-learn was used to optimize the hyperparameters of each trained model. We used this method along with cross-validation to ensure that we chose the best parameters.

3 RESULTS

This section presents in detail the findings obtained in the classification of patients with AD and MCI. Also, to facilitate interpretation, Table 4 is presented with the performance metrics of each algorithm evaluated, allowing comparison of each metric in each model.

After training. Results are shown for the following metrics: Precision, which indicates the percentage of samples classified correctly. Recall indicates the percentage of positive samples identified by the model. The F1 score corresponds to the harmonic mean between precision and sensitivity. MCC is the Matthews concordance coefficient. The weighted mean is the precision and recovery of each category, where the weights are the proportions of samples of each category in the data set.

As shown in Table 5, the DT algorithm obtained the highest scores in all metrics. It is followed by the RF algorithm, which has a slightly lower performance than DT but is still positioned as a competitive algorithm. The LR, NB, and KNN algorithms show similar performance to each other, but lower than the algorithms mentioned above. The SVM algorithm is the lowest-performing algorithm in all metrics. For example, the Accuracy of DT and NB share the highest Accuracy at 75%, followed by RF and LR with 72%. In Accuracy, LR gets the highest Accuracy with 79%, followed by DT with 75%. In Recall, DT, RF, and LR share the highest recall with 75%. In MCC, the LR algorithm obtains the highest MCC with 54%, followed by DT with 50%.

The Macro average and Weighted average coincide with the F1-score results, with DT being the highest with 75%.

In the classification of AD and MCI, there are two key metrics that we cannot fail to consider. On the one hand, accuracy plays an important role, as it avoids misdiagnosing healthy people as sick, thus avoiding unnecessary and even harmful treatments. In this context, achieving high accuracy is the goal in this type of task. On the other hand, Recall is just as important; correctly identifying sick patients helps us to make a timely diagnosis, thus early access to treatment, significantly improving the quality of life of patients. Table 5 shows that the DT model obtained the best results in accuracy and recall, which makes it an excellent choice for this type of task. Also, the LR and RF models performed well but slightly inferior to DT. Overall, in this study, ML models have shown good results, indicating that they can be good tools to support clinicians, especially for classifying patients with AD and MCI. However, it is important to point out that the results may vary depending on the dataset, so it is crucial to perform further tests with different datasets and apply other techniques to validate and find the most appropriate model.

Table 5. Results of the models in the testing phase

Algorithm	DT	RF	LR	NB	KNNs	SVM
Accuracy	0.7500	0.7222	0.7222	0.7500	0.6944	0.6944
Precision	0.7508	0.7250	0.7945	0.7107	0.7000	0.5427
Recall	0.7500	0.7222	0.7500	0.6944	0.6944	0.5000
F1-Score	0.7498	0.7214	0.7402	0.6884	0.6923	0.5214
AUC-ROC	0.7954	0.7627	0.6633	0.6595	0.5664	0.7183
MCC	0.5008	0.4472	0.5427	0.4048	0.3944	0.2427
Macro avg	0.7498	0.7214	0.7402	0.6884	0.6923	0.5214
Weighted avg	0.7498	0.7214	0.7402	0.6884	0.6923	0.5214

4 DISCUSSION

The results of this study may be a useful and innovative tool in the diagnosis of AD and MCI [10]. Comparing our findings with studies employing deep learning models, such as those by Dwivedi et al. [9] and Wang et al. [11], these works show that they achieve superior accuracy levels, in some cases above 90%, by using advanced CNN architectures and multimodal approaches integrating imaging, genetics, and biomarkers. However, these models are computationally intensive and have limitations in terms of clinical interpretability. Unlike our study, which is based on classical, explainable, and computationally efficient models, such as DT and LR, making them suitable for application in resource-limited settings.

The DT and RF models have demonstrated significant performance in relation to accuracy, sensitivity, and F1-score metrics, reaching an average performance of 75%. These results, to some extent, relate to the findings of the papers analyzed in the literature review in the field of machine learning.

For example, when comparing our results with the findings of the Karaglani et al. paper [13], these results demonstrate that ML is advantageous for early detection of AD. They were able to achieve diagnostic accuracy levels of 85% using AutoML methods. This work focused on hematological factors, which opens new avenues to make a diagnosis without having to do anything invasive. On the other hand, our method focuses on comparing interpretable algorithms with standard CSF biomarkers and clinical data, such as MMSE. However, our study is valuable because it evaluates the potential of the correlation structure between cognitive and neurological biomarkers to improve the classification of MCI and AD cases, even in resource-limited situations. In the clinical setting, the decision must be clear to be made, and Karaglani et al. opt for automated optimization strategies. However, we prefer more interpretable models.

The works analyzed in the literature review support the great potential of ML algorithms for medical diagnosis, in this case AD and MCI. For example, the authors in the paper [15] employed automated encoders to detect patterns in neuroimages and thereby managed to minimize the differences between different techniques used to detect AD and MCI. These findings align with our results, in the sense that we use ML techniques to automate the identification of AD. Similarly, our results relate to the findings of the paper [16], where they proposed an ML model ensemble-based approach for early detection of AD. Moreover, in [17] the authors evaluated the effectiveness of AI models on innovative biomarker ensembles. Concluding that AI algorithms can improve accuracy, sensitivity, and specificity in the detection of AD. These results correlate with our findings, as it reaffirms that ML models are very useful tools in the field of medicine.

The results obtained in this work, as in other studies, to some extent depend on the data set used. This means that the results obtained are not absolute, so it is necessary to validate these models with diverse populations and larger data sets. This can be evidenced in the results of the work [18], where they obtained results slightly higher than those obtained in this work, with an accuracy of 79.5%, a sensitivity of 72.2%, and a specificity of 85.7%. This difference could be due to the use of larger datasets or optimization techniques, among other factors. In the same line of research, [19] and [22] reaffirm that ML algorithms are currently key tools in the prediction of Alzheimer's disease, through the analysis of biomarkers, with sensitivity metrics ranging between 0.75 and 0.85 and specificity between 0.69 and 0.81. These results strengthen our results. On the other hand, [20] supports the idea that combining ML algorithms with innovative biomarkers can significantly improve the identification of diagnostics.

A crucial aspect in the implementation of AI models is their explainability, mainly in medical contexts where decisions must be supervised by health specialists. In this sense, one of the contributions of this work is the choice of models such as DT and LR, whose structure allows a direct interpretation of the rules and coefficients used for classification.

It is important to mention that the DT and RF models are known for their good performance on small datasets such as the one in this study. However, the consistency of the classification task is reinforced by the inclusion of other classifiers with different mechanisms, such as SVM and KNN, which partly explains their high accuracy in this study.

Despite the confidence these results may generate, we cannot ignore the limitations of ML algorithms. One of them is the sensitivity to biases in the data during training, which significantly affects the generalization of the models. Also, we cannot ignore the size of the dataset of 212 samples, which affects the performance of the models and their applicability. Another important limitation is the interpretability of the algorithms, since in many cases it is difficult to understand the reason behind a specific classification. This last point is very important in the clinical field, where understanding the reasoning behind a decision is as important as the decision itself. Although there are limitations that need to be addressed, the results obtained in this work indicate that ML models have great potential to become tools to support the diagnosis of AD. Early identification can impede the progression of the disease by facilitating timely interventions, improving treatment planning, and reducing healthcare costs. Ultimately, its practical implementation could improve patients' quality of life and optimize strategies. Future research should address these limitations and explore more advanced variations of DTs to determine the most appropriate architecture in clinical settings to ensure that ML models can become increasingly accurate. Also, it is suggested to test the model using data from clinical institutions in different geographic regions. This cross-validation will allow the robustness of the model to be assessed with different types of individuals and in real clinical settings.

5 CONCLUSIONS

In this study we focus on biomarker analysis as a tool for the diagnosis of AD and MCI. The results obtained are promising, as they reinforce the hypothesis that ML algorithms can play a very important role in this type of task.

Particularly, the DT and RF models achieved the best metrics in terms of accuracy, with 0.75 and 073, sensitivity 0.75 and 0.72; and F1-Score, 0.75 and 0.72, respectively. Also, it is important to highlight the values obtained by the other models. For example, the LR model achieved an accuracy rate of 79% and obtained the best score in MCC with 54%, followed closely by the DT model with 50%. These findings are encouraging, as they indicate that ML models can contribute significantly to the early detection of AD and mild cognitive impairment.

However, much work remains to be done to optimize these models and ensure their implementation in the clinical field. To this end, it is recommended to evaluate the performance of the model with different data sets, preferably with large volumes that reflect a greater variety of patients. Also, it would be appropriate to experiment with and other techniques, other algorithms to optimize diagnostic accuracy.

On the other hand, to build more optimized, robust, and reliable models, it is very important to understand the relationship between biomarkers and the progression of AD and MCI. Further research in this field may pave the way for more accurate ML models and thus more effective tools for clinical diagnosis.

6 FUNDING

"This research has been funded by Universidad Nacional Tecnológica de Lima Sur, Lima, Peru, with resolution No. 026-2025-UNTELS -VRI" and "The APC has been funded by Universidad Nacional Tecnológica de Lima Sur, Lima, Peru."

7 DATA AVAILABILITY STATEMENT

"The raw data supporting the conclusions of this article will be made available by the authors on request."

8 ACKNOWLEDGMENTS

"Thanks to the entire team for their work and dedication, which have been key to the development of this project."

9 CONFLICTS OF INTEREST

"The authors declare no conflicts of interest."

10 ETHICAL APPROVAL

"The study used a dataset obtained from the Kaggle repository, a platform recognized for hosting fully anonymized research databases. In accordance with the ADNI data use agreement and institutional guidelines, no additional ethical approval was required."

11 REFERENCES

[1] Alzheimer's Disease International, "World Alzheimer Report 2023: Reducing Dementia Risk: Never too early, never too late," 2023. Accessed: Jan. 8, 2024. [Online]. Available: https://www.alzint.org/resource/world-alzheimer-report-2023/

[2] A. C. Garcia, "Study of brain imaging correlates of Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) with machine learning," Semantic Scholar, 2019. [Online]. Available: https://www.semanticscholar.org/paper/af5f68df8ff21e33f4e3a4e059bbcf1f42bc2c5e

[3] U. Khatri, U. Khatri, G. Kwon, and G.-R. Kwon, "Alzheimer's disease diagnosis and biomarker analysis using resting-state functional MRI functional brain network with multi-measures features and hippocampal subfield and amygdala volume of structural MRI," *Front. Aging Neurosci.*, vol. 14, 2022. https://doi.org/10.3389/fnagi.2022.818871

[4] International Network of Alzheimer's Associations, "The Global Alzheimer's association interactive network," 2024. Accessed: Jan. 8, 2024. [Online]. Available: https://www.gaain.org/

[5] Alzheimer Europe, "Understanding dementia research," 2024. Accessed: Jan. 8, 2024. [Online]. Available: https://www.alzheimer-europe.org/research/understanding-dementia-research

[6] World Health Organization, "Dementia," 2024. Accessed: Jan. 8, 2024. [Online]. Available: https://www.who.int/es/news-room/fact-sheets/detail/dementia

[7] G. I. A. Calub, E. Elefante, J. C. A. Galisanao, S. L. B. G. Iguid, J. C. Salise, and S. V Prado, "EEG-based classification of stages of Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI)," in *2023 5th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*, 2023. https://doi.org/10.1109/BioSMART58455.2023.10162117

[8] S. Leandrou *et al.*, "A cross-sectional study of explainable machine learning in Alzheimer's disease: Diagnostic classification using MR radiomic features," *Front. Aging Neurosci.*, vol. 15, 2023. https://doi.org/10.3389/fnagi.2023.1149871

[9] F. Salvetti, B. Bertagni, I. Contardo, R. Gardner, J. Rudolph, and R. Minehart, "AI-driven avatars in medical training: Personalized feedback for enhanced learning," *International Journal of Advanced Corporate Learning (iJAC)*, vol. 18, no. 3, pp. 46–59, 2025. https://doi.org/10.3991/ijac.v18i3.52595

[10] M. Kapoor, M. Kapoor, R. Shukla, and T. R. Singh, "Early diagnosis of Alzheimer's Disease using machine learning based methods," in *International Conference on Contemporary Computing*, 2021. https://doi.org/10.1145/3474124.3474134

[11] Y. Wang *et al.*, "Identification of novel diagnostic panel for mild cognitive impairment and Alzheimer's disease: Findings based on urine proteomics and machine learning," *Alzheimers Res. Ther.*, 2023. https://doi.org/10.1186/s13195-023-01324-4

[12] A. C. Garcia, "Study of brain imaging correlates of mild cognitive impairment and Alzheimer' s Disease with machine learning," Semantic Scholar, 2019. [Online]. Available: https://www.semanticscholar.org/paper/31834f294d95b8cb585ed53e2224b95646730a74

[13] M. Karaglani *et al.*, "Accurate blood-based diagnostic biosignatures for Alzheimer's disease via automated machine learning," *J. Clin. Med.*, vol. 9, no. 9, p. 3016, 2020. https://doi.org/10.3390/jcm9093016

[14] S. Yousafzai, G. Z. Khan, S. Ulhaq, and M. I. Butt, "Improved neural network-based system for early and accurate diagnosis of Alzheimer disease," *Journal of Computer Science and Technology Studies*, vol. 5, no. 4, pp. 32–40, 2023. https://doi.org/10.32996/jcsts.2023.5.4.4

[15] Y. Suzuki *et al.*, "A prospective multicenter validation study of a machine learning algorithm classifier on quantitative electroencephalogram for differentiating between dementia with Lewy bodies and Alzheimer's dementia," *PLoS ONE*, vol. 17, no. 3, p. e0265484, 2022. https://doi.org/10.1371/journal.pone.0265484

[16] Y. Lai *et al.*, "Identification of immune microenvironment subtypes and signature genes for Alzheimer's disease diagnosis and risk prediction based on explainable machine learning," *Front. Immunol.*, vol. 13, 2022. https://doi.org/10.3389/fimmu.2022.1046410

[17] C. Chang, C.-H. Chang, C. Lin, C.-H. Lin, H. Y. Lane, and H.-Y. Lane, "Machine learning and novel biomarkers for the diagnosis of Alzheimer's disease," *Int. J. Mol. Sci.*, vol. 22, no. 5, p. 2761, 2021. https://doi.org/10.3390/ijms22052761

[18] R. O. Bahado-Singh *et al.*, "A community-based study identifying metabolic biomarkers of mild cognitive impairment and Alzheimer's disease using artificial intelligence and machine learning," *Journal of Alzheimer's Disease*, vol. 78, no. 4, pp. 1381–1392, 2020. https://doi.org/10.3233/JAD-200305

[19] A. Miltiadous *et al.*, "Alzheimer's disease and frontotemporal dementia: A robust classification method of EEG signals and a comparison of validation methods," *Diagnostics*, vol. 11, no. 8, p. 1437, 2021. https://doi.org/10.3390/diagnostics11081437

[20] S. A. M. Mane and A. A. Shinde, "The development of machine learning method for analyzing EEG signals for diagnosisng Alzheimer diesease," in *2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM)*, 2023. https://doi.org/10.1109/ELEXCOM58812.2023.10370246

[21] B. Taha Jijo and A. Mohsin Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 1, pp. 20–28, 2021. https://doi.org/10.38094/jastt20165

[22] N. A. M. R. Senaviratna, T. M. J. A. Cooray, and M. M. Alberto Ferreira, "Diagnosing multicollinearity of logistic regression model," *Asian Journal of Probability and Statistics*, vol. 5, no. 2, pp. 1–9, 2019. https://doi.org/10.9734/ajpas/2019/v5i230132

[23] O. Iparraguirre-Villanueva *et al.*, "Comparison of predictive machine learning models to predict the level of adaptability of students in online education," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, 2023. https://doi.org/10.14569/IJACSA.2023.0140455

[24] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification," *Augmented Human Research*, vol. 5, no. 1, pp. 1–16, 2020. https://doi.org/10.1007/s41133-020-00032-0

[25] P. Schober and T. R. Vetter, "Logistic regression in medical research," *Anesth. Analg.*, vol. 132, no. 2, pp. 365–366, 2021. https://doi.org/10.1213/ANE.0000000000005247

[26] O. Iparraguirre-Villanueva, A. Epifanía-Huerta, C. Torres-Ceclén, J. Ruiz-Alvarado, M. Cabanillas-Carbonell, and N. Wiener, "Breast cancer prediction using machine learning models," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 2, 2023. https://doi.org/10.14569/IJACSA.2023.0140272

[27] G. Murdaca *et al.*, "Potential predictors for cognitive decline in vascular dementia: A machine learning analysis," *Processes*, vol. 10, no. 10, p. 2088, 2022. https://doi.org/10.3390/pr10102088

[28] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *Stata Journal*, vol. 20, no. 1, pp. 3–29, 2020. https://doi.org/10.1177/1536867X20909688

[29] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "Smart power consumption abnormality detection in buildings using micromoments and improved K-nearest neighbors," *International Journal of Intelligent Systems*, vol. 36, no. 6, pp. 2865–2894, 2021. https://doi.org/10.1002/int.22404

[30] H. Saadatfar, S. Khosravi, J. H. Joloudari, A. Mosavi, and S. Shamshirband, "A new K-nearest neighbors classifier for big data based on efficient data pruning," *Mathematics*, vol. 8, no. 2, p. 286, 2020. https://doi.org/10.3390/math8020286

[31] I. Triguero, D. García-Gil, J. Maillo, J. Luengo, S. García, and F. Herrera, "Transforming big data into smart data: An insight on the use of the K-nearest neighbors algorithm to obtain quality data," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 2, p. e1289, 2019. https://doi.org/10.1002/widm.1289

[32] X. He, P. Wang, and J. Cheng, "K-nearest neighbors hashing," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2834–2843. https://doi.org/10.1109/CVPR.2019.00295

[33] M. R. Islam, A. R. M. Kamal, N. Sultana, R. Islam, M. A. Moni, and A. Ulhaq, "Detecting depression using K-Nearest Neighbors (KNN) classification technique," in *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2 2018)*, 2018. https://doi.org/10.1109/IC4ME2.2018.8465641

[34] K.-T. Tsai, "Support vector machine," in *Machine Learning for Knowledge Discovery with R*, 2021, pp. 101–128. https://doi.org/10.1201/9781003205685-5

[35] S. G. Chen and X. J. Wu, "A new fuzzy twin support vector machine for pattern classification," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 9, pp. 1553–1564, 2018. https://doi.org/10.1007/s13042-017-0664-x

[36] R. Hua *et al.*, "Improved naive bayes classification algorithm for traffic risk management," *Research Square*, 2021. https://doi.org/10.21203/rs.3.rs-355037/v1

[37] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes," *Journal of Supercomputing*, vol. 77, no. 5, pp. 5198–5219, 2021. https://doi.org/10.1007/s11227-020-03481-x

12 AUTHORS

Orlando Iparraguirre-Villanueva is a Systems Engineer with a master's degree in information technology management and a PhD in Systems Engineering from the Universidad Nacional Federico Villarreal, Peru. He is certified in ITIL®, with a specialization in Business Continuity Management (SBCM) and SCRUM Certification. He is a Lecturer and panelist in national and international events, with extensive experience in undergraduate and graduate teaching in various universities across the country (E-mail: oiparraguirre@untels.edu.pe).

José Luis Herrera Salazar is a visionary leader in Systems Engineering in Peru, combining a robust academic, research, and technological innovation trajectory. D. in Systems Engineering from the Universidad Nacional Federico Villarreal, has an outstanding scientific production with publications in journals indexed in Scopus and Web of Science, contributing advances in artificial intelligence, academic data mining, and intelligent systems. He is currently a Senior Lecturer at the Universidad Nacional Tecnológica de Lima Sur (UNTELS) and a Contract Lecturer at the Universidad Autónoma de Ica. In addition, he has held senior academic management positions at major universities, leading engineering programs and interdisciplinary projects (E-mail: Jherreras@untels.edu.pe).

Gloria Castro-Leon is a Systems Engineer, bachelor's in computer science, Master in Administration, and has a PhD in Systems Engineering. She is a Senior Lecturer with 25 years of experience in teaching undergraduate and postgraduate courses, both in public and private universities, with extensive knowledge in project management and information technology, an ITIL Foundation Certificate in IT service, with extensive experience in administrative management of the Peruvian university. Director of the research institute of UNTELS (E-mail: Gcastrol@untels.edu.pe).

Hernán Ochoa-Carbajal is a University Professor–Appointed at the Universidad Nacional Tecnológica de Lima Sur–Associate Category in the Professional School of Systems Engineering with Academic Degree–master's in business management and Master in Education Administration. He has a Doctorate in Systems Engineering, 2nd Winner of the VII National Contest of Master's Thesis and Doctorates organized by the National Assembly of Rectors in the city of Lima (E-mail: hochoa@untels.edu.pe).

Henry Chero-Valdivieso has a degree in Mathematics Education from the National University of Trujillo–UNT, master's in education with a specialization in Research, Curriculum and University Teaching. He has a Doctorate in Education Administration from the Universidad César Vallejo and a Doctoral Candidate of the European Higher Education Area in the specialty of Communication and Education in digital environments at UNED-Spain. He did postgraduate studies in Systems Engineering with a mention of Information and Communication Technologies (ICT). He also has a Diploma in Advanced Studies in Research Proficiency (DEA) from the UNED University of Spain (E-mail: hacheroc@ucvvirtual.edu.pe).

Rosalynn Ornella Flores-Castañeda is a Systems Engineer by profession, with a master's degree in business administration and international Relations, a master's degree in information technology management, a Doctorate in Education Administration, and a Doctorate in Systems Engineering. She has served as the Director of Academic Records at a University. She teaches research courses, advise theses, and prepare articles to be published in high-impact journals (E-mail: rosalynn.flores@usil.pe).