

Article

Comparison of Machine Learning Algorithms in the Prediction of Hospitalized Patients with Schizophrenia

Susel Góngora Alonso ¹, Gonçalo Marques ², Deevyankar Agarwal ³, Isabel De la Torre Díez ^{1,*}
and Manuel Franco-Martín ⁴

¹ Department of Signal Theory and Communications, and Telematics Engineering, University of Valladolid, Paseo de Belén, 15, 47011 Valladolid, Spain; susel.gongora@uva.es

² Polytechnic of Coimbra, ESTGOH, Rua General Santos Costa, 3400-124 Oliveira do Hospital, Portugal; goncalosantosmarques@gmail.com

³ EEE Section, Department of Engineering, Higher College of Technology, Muscat 113, Oman; deevyankar.agarwal@hct.edu.om

⁴ Psiquiatry Service, Hospital Zamora, 49021 Zamora, Spain; mfrancom@saludcastillayleon.es

* Correspondence: isator@tel.uva.es; Tel.: +34-983-423-000 (ext. 3703)

Abstract: New computational methods have emerged through science and technology to support the diagnosis of mental health disorders. Predictive models developed from machine learning algorithms can identify disorders such as schizophrenia and support clinical decision making. This research aims to compare the performance of machine learning algorithms: Decision Tree, AdaBoost, Random Forest, Naïve Bayes, Support Vector Machine, and k-Nearest Neighbor in the prediction of hospitalized patients with schizophrenia. The data set used in the study contains a total of 11,884 electronic admission records corresponding to 6933 patients with various mental health disorders; these records belong to the acute units of 11 public hospitals in a region of Spain. Of the total, 5968 records correspond to patients diagnosed with schizophrenia (3002 patients) and 5916 records correspond to patients with other mental health disorders (3931 patients). The results recommend Random Forest with the best accuracy of 72.7%. Furthermore, this algorithm presents 79.6%, 72.8%, 72.7%, and 72.7% for AUC, precision, F1-Score, and recall, respectively. The results obtained suggest that the use of machine learning algorithms can classify hospitalized patients with schizophrenia in this population and help in the hospital management of this type of disorder, to reduce the costs associated with hospitalization.

Keywords: hospitalization; machine learning algorithms; predictive models; random forest; schizophrenia



Citation: Góngora Alonso, S.; Marques, G.; Agarwal, D.; De la Torre Díez, I.; Franco-Martín, M. Comparison of Machine Learning Algorithms in the Prediction of Hospitalized Patients with Schizophrenia. *Sensors* **2022**, *22*, 2517. <https://doi.org/10.3390/s22072517>

Academic Editor: Steve Ling

Received: 2 March 2022

Accepted: 23 March 2022

Published: 25 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Currently, data mining and machine learning techniques allow the exploration and analyzation of data patterns through statistical methods and artificial intelligence [1–3]. Researchers can obtain correlations and patterns from large data sets to create new knowledge with the help of machine learning and artificial intelligence [4–6].

Psychiatry is a field of medicine that specializes in studying and treating mental, emotional, or behavioral disorders [7,8]. Schizophrenia is a severe and debilitating chronic mental illness causing a high burden and healthcare utilization with the global age-standardized point prevalence of schizophrenia at 0.28%, which means 21 million cases worldwide [9]. The onset of the symptoms and diagnosis used to be during the second and third life decades and is controversial regarding the sex ratio, which is balanced between genders depending on the methodology, but in clinical groups has higher prevalence in men, as was found in the current Spanish study [10]. It is characterized by many different symptoms and signs such as thought disorder, delusions, emotional blunting, hallucinations, changes in volition, as well as cognitive deficits [11–14]. However, the main phenomenological feature is the variety of symptomatology and lack of a pathognomonic symptom or sign.

People with this disorder have higher rates of morbidity and mortality than the general population (adjusted hazard ratio (aHR) = 3.52). Most deaths are related to physical disorders, mainly metabolic syndrome, and its consequences (stroke, hypertension), and infectious diseases [15,16]. However, suicide is also a relevant cause of death in patients with schizophrenia, reaching 5% of them [17]. Indeed, it is considered an underestimated event because about 25–50% of these patients attempt suicide during their life [18]. The current challenge of schizophrenia is an early diagnosis of the illness and treatment to mitigate the progressive impairment of cognitive, skills, and social function of the natural course of the disease. It will also be useful for preventing suicide and improving comprehensive physical and mental treatment. Diagnostic and monitoring depend mainly on their clinical experience, and decisions are made based on the specific case [19,20].

On the other hand, and considering that schizophrenia patients use multiple healthcare services, with higher hospitalizations rates and longer mean duration, it is a relevant health problem. Indeed, this condition imposes a tremendous economic charge both for patients, their families, and society in general [21,22].

Data analysis and decision making are crucial steps, especially in mental illness [23,24]. Classification algorithms such as Logistic Regression, Decision Tree, Random Forest, Adaboost, Naïve Bayes, k-Nearest Neighbor (k-NN), and Support Vector Machine (SVM) are used in different studies [25–28] for the diagnosis of patients with Alzheimer's, Parkinson's, and mild cognitive impairment (MCI). The application of machine learning in mental health has allowed the prediction of genetic risk, the identification of candidates for biomarkers, or the exploration of etiological mechanisms [29]. Furthermore, the predictive precision of medical data by classifying people as healthy or not allows the development of new therapeutic and preventive strategies in mental health diseases [30–33].

In clinical activity, the identification of suicide risk in schizophrenia is very relevant and there are many proposals for its detection. In the study [34], the authors sought to find new available clinical resources that accurately identify the suicide risk in schizophrenia. In the study [35], the authors developed a clinically useful predictive model to identify schizophrenia patients who attempt suicide and those who do not, in a sample of 345 participants. The results showed the best metrics in the support vector classifier model and regularized logistic regression with an accuracy of 67.0% and an area under the ROC curve (AUC) of 0.70 and 0.71, respectively.

In studies such as [36], the authors analyzed a predictive system to preventively diagnose schizophrenia disease using SVM, Random Forest, ANN, and Naïve Bayes. The results showed the highest precision of 90.69% using SVM, Random Forest, and Naïve Bayes. In this work, the authors applied feature selection to understand if it was possible to improve the performance using recursive feature elimination.

In [37], the authors investigated a classification algorithm to predict schizophrenia using combined electroencephalogram (EEG) features, obtaining an average accuracy of 78.24% with the SVM algorithm. This work included a small data set with a total of 68 individuals and feature selection methods, and only SVM was tested and not compared with related machine learning approaches.

The authors of [38] have tested algorithms such as SVM, Naïve Bayes, Random Forest, and gradient boosting to predict schizophrenia patients and healthy controls, using 72 individuals corresponding 48 and 24 to schizophrenia patients and healthy controls, respectively. The experiments showed an accuracy of 58.2% using SVM and 68.6% through Random Forest. In [39], the authors with a sample of 21 features obtained accuracy values of 72.3% for Decision Tree, SVM with 78.7%, k-NN with 76.5%, and Random Forest with 85.1%. According to the authors, about 500 patient medical records were processed as the data set and cross-validation was used.

Machine learning algorithms such as Random Forest and SVM have been applied in the studies [40,41] to discriminate between schizophrenia patients and healthy patients using magnetic resonance imaging (MRI). The results of study [41] showed a high performance

rate using 504 features. Random Forest had sensitivity = 87.6% and specificity = 95.9%, and SVM had sensitivity = 89.5% and specificity = 94.5%.

In the study [42], the authors classified patients with schizophrenia from healthy patients using the messenger RNA expression level in peripheral blood. The authors compared machine learning algorithms such as artificial neural networks, extreme gradient boosting, SVM, Decision Tree, and Random Forest. The results showed SVM as the best model with AUC of 0.993, sensitivity = 100, and specificity = 0.895.

EEG signals are used in multiple studies to classify patterns between healthy and schizophrenia patients [43–45], using algorithms such as Random Forest, SVM, kNN, and Multilayer Perceptron (MLP). Other studies used a machine learning approach to predict cognitive function in schizophrenia [46], identify people with schizophrenia through social networks [47,48], and develop a predictive model to identify violence in patients with schizophrenia [49].

In the previous studies [50,51], an analysis of the hospitalization of patients with mental disorders in the region of Castilla and León was carried out, determining schizophrenia as the most prevalent disorder in this population. Therefore, the objective of this paper is to compare the performance of machine learning algorithms: Decision Tree, AdaBoost, Random Forest, Naïve Bayes, k-NN, and SVM in the prediction of hospitalized patients with schizophrenia, and to identify the appropriate classification algorithm for this research line that best classifies patients hospitalized with this disease. The authors have tested the previously mentioned low computational machine learning methods since several researchers have used the same methods and it is critical to compare the performance of the methods to understand if the results converge in this field of research. The authors expect Random Forest to provide the best performance based on the previous research activities of other authors [35–39]. These studies are different regarding the number of features, sample population, and algorithms they use to develop their research.

2. Materials and Methods

The purpose of the study is to analyze and compare the performance metrics of machine learning algorithms for the recognition of hospitalized patients with schizophrenia. According to previous studies [27,35,50], we selected six algorithms, such as Random Forest, k-NN, AdaBoost, Naïve Bayes, Decision Tree, and SVM, for the development of our research. The authors of [50] have made an initial review of this research topic. Moreover, by performing a more updated analysis of the literature, the authors find relevant studies of schizophrenia [27,35–44] that focus their research on these machine learning techniques.

2.1. Database

The database used in the study contains patient admission records from 11 public hospitals in Castilla and León, Spain, taking the minimum data set following the national rules for saving hospitalized patients' data. This database has been approved under the Project "Contribution to the Analysis and Development of Data Mining Techniques and Sources in an Internet of Things (IoT) Environment in the Field of Mental Health" by the Ethics Committee of the University of Valladolid (Ref. PI 20-1780).

Data includes the period from 2005 to 2015. We selected the acute units of the 11 hospitals in the region. The data set used contains a total of 6933 patients with different mental disorders, of which 3002 patients have schizophrenia and 3931 have other mental disorders. A total of 11,884 admission records corresponds to the 6933 patients, of which 5968 records belong to patients with schizophrenia and 5916 records belong to patients with other mental disorders. The most representative gender in the data set is male with 55.9% (6639 of 11,884 records), while females represent 44.14% (5245 of 11,884 records). The age of the patients is between 13 and 97 years.

The data are anonymized and follow the international coding standard for diseases ICD-9 [52]. The database collects minimum data set information such as the hospital to which the patient belongs, age, gender, dates of admission and discharge from the hospital,

days of stay, the main diagnostic diseases of psychiatric type, secondary diagnoses, and complementary therapies to each medical diagnosis. The procedures applied in the data set have been carried out, guaranteeing the patient's privacy and security.

2.2. System Architecture

The proposed predictive model includes data processing and feature selection before the application of machine learning techniques (see Figure 1).

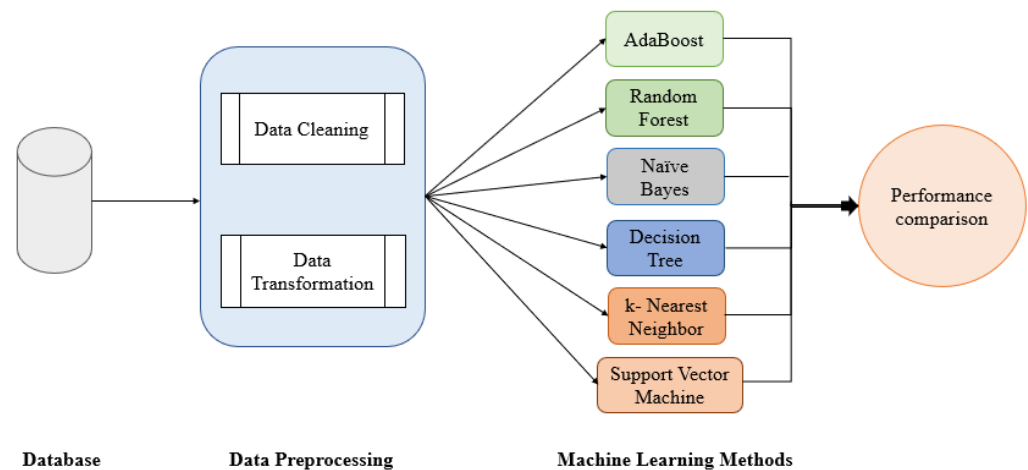


Figure 1. Flow diagram of the study. The diagram shows the first phase of pre-processing the database. Subsequently, the machine learning algorithms described are applied to the pre-processed dataset. In the final phase, the performance metrics obtained from the algorithms are compared.

The data pre-processing phase includes cleaning and data processing. For the development of this phase, the programming environment R version 3.6.3 was used, applying the dplyr [53] and tidyr [54] packages. The authors have performed a general exploratory analysis considering the type of data. In this phase, an analysis of nulls, zeros, and atypical values has been carried out, where null values, double white spaces, and special characters have been eliminated to avoid false classifications according to [55]. Furthermore, data normalization has been used with an interval (0–1) as suggested by [56]. Finally, a longitudinal and data coherence analysis was performed.

The original data set has a total of 27 different variables for the prediction of hospitalized patients with schizophrenia. The authors have classified them to obtain the best predictive features. For the selection of features, we have used the Gain Information, Gain Ratio, Gini, X^2 , and ReliefF methods. These techniques have been used to analyze the relationship between each input variable and the target variable according to previous research activities [15,29]. The threshold value chosen was Gain Information ≥ 0.009 , Gain Ratio ≥ 0.005 , and Gini ≥ 0.007 . Consequently, the input variables with better scores have been used in the model. Table 1 shows the classification and metrics of the subset of features obtained.

Once the selection of features with the supervised filter method has been carried out, 8 predictive variables are chosen to use in the study. The **Diag_Sec02_Code**, **Diag_Sec03_Code**, **Diag_Sec04_Code**, **Diag_Sec05_Code**, and **Diag_Sec06_Code** features contain the disease codes established by ICD-9 [52], which correspond to the diagnoses of admitted patients. Considering the clinical aspect of these 5 features, each one of them includes the following classification of diseases: (1) mental health disorders (Codes 290.0-319), (2) nervous system (Codes 323.9-389.9), (3) infectious diseases (Codes 041.4-138), (4) neoplasias (Codes 141-239.6), (5) circulatory system (Codes 394.1-459.81), (6) endocrine diseases (Codes 240-279.11), (7) respiratory system (Codes 461.1-519.8), (8) digestive system (Codes 521.09-579.8), (9) genitourinary system (Codes 584.8-626.6), (10) blood diseases (Codes 280-289.81), (11) skin and subcutaneous tissue (Codes 681.02-709.01), (12) le-

sion and poisoning (Codes 801.30-999.2), (13) congenital anomaly (Codes 743.61-759.89), and (14) arthropathies and related disorders (Codes 710.1-737.39). Stays_Days corresponds to days of stay in the healthcare complex, and we have included personal data such as age and gender variables. The Diag_Schizophrenia variable is the target class. These variables are used in Decision Tree, AdaBoost, Random Forest, k-NN, Naïve Bayes, and SVM algorithms to predict hospitalized patients with schizophrenia. For the implementation of the model, a MacBook Pro (2018) running on a macOS Catalina version 10.15.3 operating system has been used to perform predictive analysis, develop the machine learning models, and data visualization.

Table 1. Metrics and ranking of the features.

Variables	Information Gain	Gain Ratio	Gini	χ^2	Relieff
Diag_Sec02_Code	0.047	0.023	0.032	128.578	0.012
Diag_Sec03_Code	0.014	0.007	0.010	0.044	0.006
Diag_Sec04_Code	0.011	0.006	0.008	91.753	0.004
Diag_Sec05_Code	0.016	0.010	0.011	269.514	0.003
Diag_Sec06_Code	0.014	0.012	0.010	331.083	0.019
Stays_Days	0.009	0.005	0.007	128.946	−0.0003
Age	0.025	0.012	0.017	310.541	0.012
Gender	0.069	0.070	0.047	623.212	-
Admission_Type	0.0004	0.002	0.0003	0.238	−0.002
Proc_Ppal_Code	0.005	0.003	0.004	49.338	0.015
Proc_Sec02_Code	0.005	0.003	0.003	100.140	−0.014
Proc_Sec03_Code	0.004	0.004	0.003	96.206	−0.005

Algorithms of Machine Learning

In the study, we evaluated different machine learning algorithms such as Random Forest, Naïve Bayes, SVM, Decision Tree, AdaBoost, and kNN. Table 2 shows the parameters of the machine learning algorithms used in the study.

Table 2. Parameters of the machine learning algorithms used in the study.

Algorithms	Parameters
Random Forest	Number of trees = 10
	Maximum number of considered features: unlimited
	Maximum tree depth: unlimited
AdaBoost	Stop splitting nodes with maximum instances = 5
	Base estimator: tree
	Number of estimators = 50
Decision Tree	Minimum number of instances in leaves = 2
	Minimum number of instances in internal nodes = 5
	Maximum depth = 100
kNN	Number of neighbours = 5
	Distance metric: Euclidean
	Weight: Uniform
Naïve Bayes	fL = 0
	usekernel: False
	adjust = 0
SVM	C = 1.0
	sigma = 0.5
	Numerical tolerance = 0.001
	Iteration limit = 100

Regarding the Decision Tree algorithm, the model induces a binary tree, the minimum number of instances used on the leaves is two instances, the model does not split subsets smaller than five, the maximum depth is 100, and the algorithm stops dividing when most nodes reach 95%. The AdaBoost algorithm is an iterative ensemble method that builds a

strong classifier by combining multiple low-performing classifiers. The base estimator that we used to train the model was Decision Tree with 50 estimators.

The Naïve Bayes algorithm uses the adjustment parameter usekernel with a value False to assume a Gaussian density distribution, the correction factor Laplace (fL) is 0, and the parameter adjust was kept constant at a value of 0. For the k-NN algorithm, the number of neighbors is 5 and the distance metric is Euclidean, while for SVM we use $C = 1.0$ and $\sigma = 0.5$, radial basis function as a kernel parameter, numerical tolerance = 0.001, and iteration limit = 100.

The Random Forest algorithm is described as a joint learning technique since it combines results from multiple decision trees, returning a single prediction. Random Forest models are made up of a set of individual decision trees, each trained on a different sample of the training data, generated by bootstrapping. Prediction of a new observation is obtained by adding the predictions of all the individual trees that make up the model [49,57]. The parameters used are 10 trees, the maximum number of features considered is unlimited, there is no replicable training, the maximum depth of the tree is unlimited, and the nodes stop dividing into five maximum instances (See Table 2).

The pseudocode of Random Forest algorithm is presented as follows:

1. Select “ k ” features from total “ m ” features at random where $k < m$;
2. Calculate the node “ d ” through the better split point among “ k ” feature;
3. Split the node into child nodes through the better split;
4. While “1” number of nodes has been done, redo steps 1 to 3;
5. Create forest by doing steps 1 to 4 by “ n ” times to build “ n ” set of trees.

Figure 2 shows the visual diagram of the Random Forest algorithm.

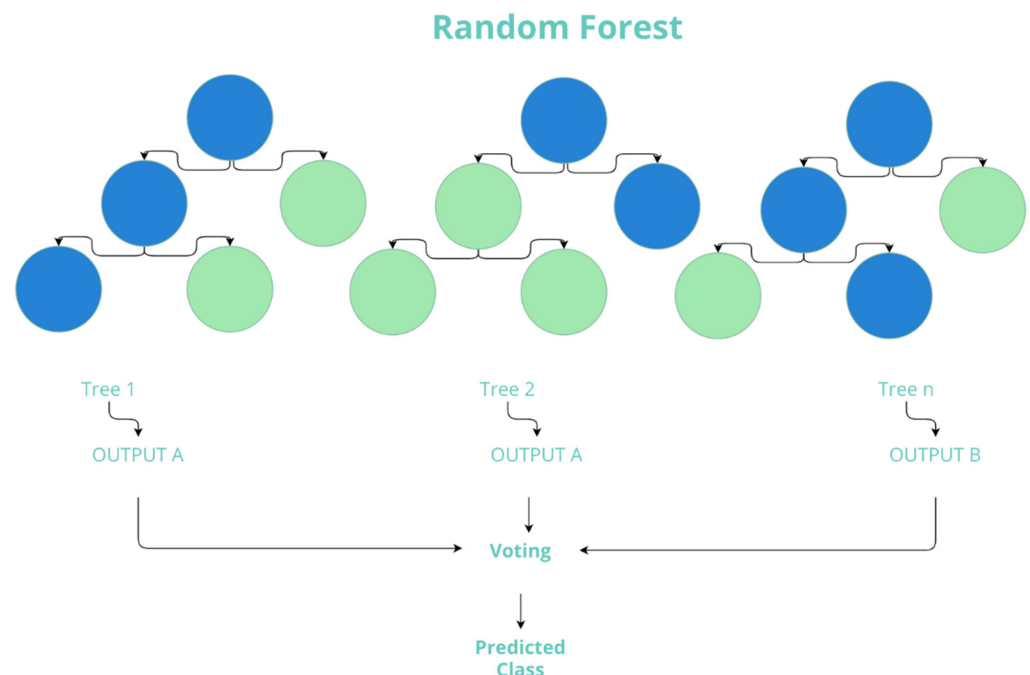


Figure 2. Diagram of the Random Forest algorithm. The algorithm generates multiple trees; in the figure, each set represents a tree. Each tree classifies a class, resulting in the class with the highest number of votes.

For the evaluation of the predictive model, we have followed the approach of other studies [33,58], using a k-fold stratified cross-validation procedure [35,38]. To avoid the overfitting problem, the authors have assigned a value to $k = 10$. This validation method divides the entire data into 9 training folds and 1 validation fold. Each fold is used once as the test set, while the remaining folds are used for training. Consequently, all the data is used both for training and testing. The results obtained with each algorithm are compared

with each other considering different machine learning performance metrics: accuracy, precision, F1-score, AUC, and recall.

3. Results

3.1. Data Analysis

The study includes a total of 11,884 records corresponding to 6933 hospitalized patients with several mental health disorders. Analysis of the clinical data shows significant differences regarding the gender of hospitalized patient records with schizophrenia and without the disease (See Table 3). In the records of schizophrenia, men represent 71.0% (4238 of 5968 records), being the most affected by this psychiatric disorder, compared to women who represent 29.0% (1730 of 5968 records). On the other hand, in records without schizophrenia, men represent 40.6% (2403 of 5916 records) and women represent 59.4% (3513 of 5916 records). The group from 31 to 60 years is the most representative age of the data set and represents 64.9% (7717 of 11,884 records). Regarding the days of stay and age mean, both groups of records have similar behavior. Records with schizophrenia show a mean age of 43 years and a mean stay of 17 days, while the records with other disorders show a mean age of 49 years and 14 days of stay.

Table 3. Analysis of clinical data.

Variables	Total (N = 11,884 Admission Records)	
	n = 5968 Records Schizophrenia	n = 5916 Records Non-Schizophrenia
Gender (%)		
Male	71.0	40.6
Female	29.0	59.4
Age, mean (years)	43	49
<18 years	10	36
18–30 years	1048	737
31–45 years	2493	1756
46–60 years	1624	1844
>60 years	793	1543
Days of stay, mean (days)	17	14
Main diagnoses of the predictive variable		
Diag_Sec02_Code for records with schizophrenia		
Non-compliance with medical treatment	473	130
Tobacco abuse disorders	353	111
Family record of psychiatric disease	229	118
Abuse of continuous cannabis	200	70
Alcohol abuse	159	86
Main diagnoses of the predictive variable		
Diag_Sec02_Code for records without schizophrenia		
Dysthymic disorder	19	687
Personality disorder	75	265
Neom arterial hypertension	140	177
Personality histrionic disorder	4	167
Psychosis	40	162

The Diag_Sec02_Code is the second most relevant variable in the study according to the ranking of features in Table 1. The analysis of the main diagnoses of this variable shows different disorders in both groups of records. Records with schizophrenia show 473 records of non-compliance with medical treatment, 353 records of tobacco abuse disorders, 229 of a family record of psychiatric disease, 200 records of continuous cannabis abuse, and 145 records of alcohol abuse. The diagnoses of patient records without schizophre-

nia present 687 records with dysthymic disorder, 265 records with personality disorder, 177 records with neom arterial hypertension, 167 records with personality histrionic disorder, and 162 records with neom psychosis.

3.2. Model Evaluation

The authors have used the Random Forest, AdaBoost, Naïve Bayes, k-NN, Decision Tree, and SVM algorithms to predict hospitalized patients with schizophrenia. The evaluation of the predictive model was performed using a 10-fold stratified cross-validation. The performance results of the methods used are shown in Table 4. The results when identifying schizophrenia in hospitalized patients against each of these groups separately are presented in Supplementary Materials Tables S1 and S2.

Table 4. Performance metrics applying 10-fold stratified cross-validation.

Algorithms	AUC	Accuracy	Precision	F1-Score	Recall
Random Forest	0.796	0.727	0.728	0.727	0.727
AdaBoost	0.765	0.708	0.708	0.708	0.708
Decision Tree	0.682	0.682	0.682	0.681	0.681
k-NN	0.729	0.677	0.676	0.676	0.676
Naïve Bayes	0.729	0.670	0.671	0.669	0.670
SVM	0.641	0.657	0.657	0.657	0.657

Consequently, Random Forest presents the best accuracy with 72.7%, as well as the rest of the metrics. AdaBoost and Decision Tree showed 70.8% and 68.2% accuracy, respectively. Naïve Bayes reports 67.0% accuracy, k-NN with 67.7% accuracy, and SVM with the lowest value of accuracy at 65.7%. Naïve Bayes and k-NN algorithms present better results in terms of AUC compared to Decision Tree; however, Decision Tree improves in terms of accuracy, precision, F1-score, and recall.

The ROC curves (receiver operating characteristic curve) shown in Figures 3 and 4 show graphs of sensitivity vs. 1-specificity through different cut points.

The authors evaluated for target = 0 and target = 1 with False Positive (FP) = 500, False Negative (FN) = 500, and probability of target = 50.0%. In this way, it is possible to visually evaluate the general performance of each classification algorithm to be evaluated. AUC under the ROC curve provides a normalized mean performance of the classifier, considering the entire range of output decision thresholds in the plane of specificity-sensitivity.

Considering the presented ROC curves created for the 10-fold stratified cross-validation, the authors obtained an AUC value of 0.796, 0.765, 0.682, 0.729, 0.729, and 0.641 for Random Forest, AdaBoost, Decision Tree, k-NN, Naïve Bayes, and SVM, respectively. These values recommend the Random Forest algorithm as the best normalized average classifier performance, with an AUC value much higher than the other algorithms. Therefore, when comparing the ROC curves shown in Figures 3 and 4, the sensitivity and specificity values are balanced, and the algorithms discriminate between hospitalized patients with schizophrenia and without the disorder with approximately the same probability.

The applied method is compared with some studies available in the literature that use similar techniques (See Table 5).

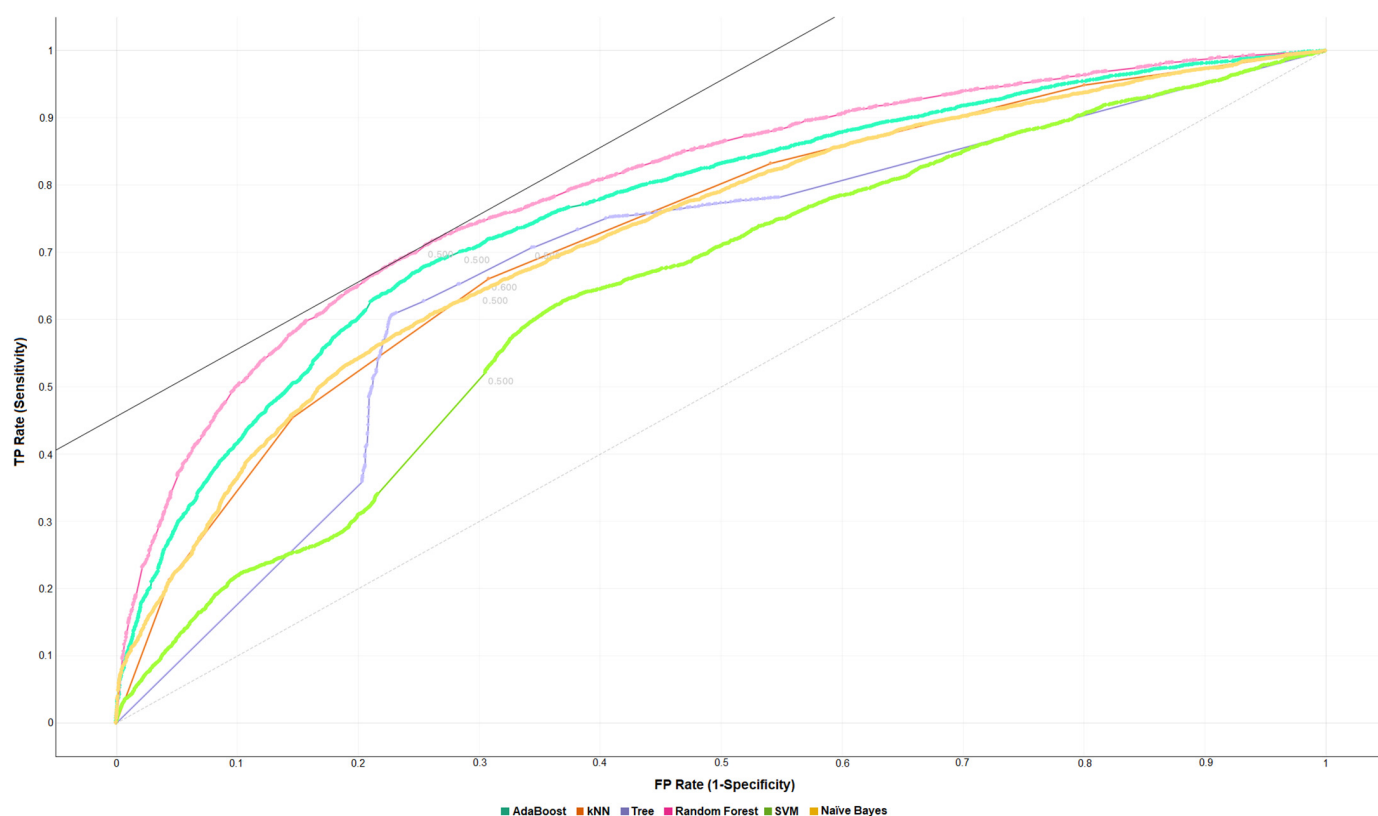


Figure 3. ROC curve for target = 0 with FP = 500, FN = 500, and target probability = 50.0%. The graph shows the ROC curves created by the false positive and false negative values with 10-folds stratified cross-validation. Each ROC curve is represented by a different color (See legend). The Random Forest algorithm shows the best value of AUC= 0.796 (See Supplementary Materials Table S1) for class 0 (records of non-schizophrenia).

Table 5. Comparison of results obtained with other studies.

Reference	Method	Validation	Dataset	AUC	Accuracy (%)
[35]	Random Forest	Cross-Validation k = 10	N = 345 patients	0.67	66.00
[36]	Random Forest	Cross-Validation k = 10	N = 86 patients	-	90.69
[36]	Naïve Bayes	Cross-Validation k = 10	N = 86 patients	-	90.69
[36]	SVM	Cross-Validation k = 10	N = 86 patients	-	90.69
[37]	SVM	Leave-One-Out Cross-Validation (LOOCV)	N = 68 patients	-	78.24
[38]	Random Forest	Cross-Validation k = 10	N = 72 patients	0.68	68.60
[39]	Random Forest	Cross-Validation	N = 466 patients	-	85.10
Our study	Random Forest	Cross-Validation k = 10	N = 6933 patients	0.79	72.74

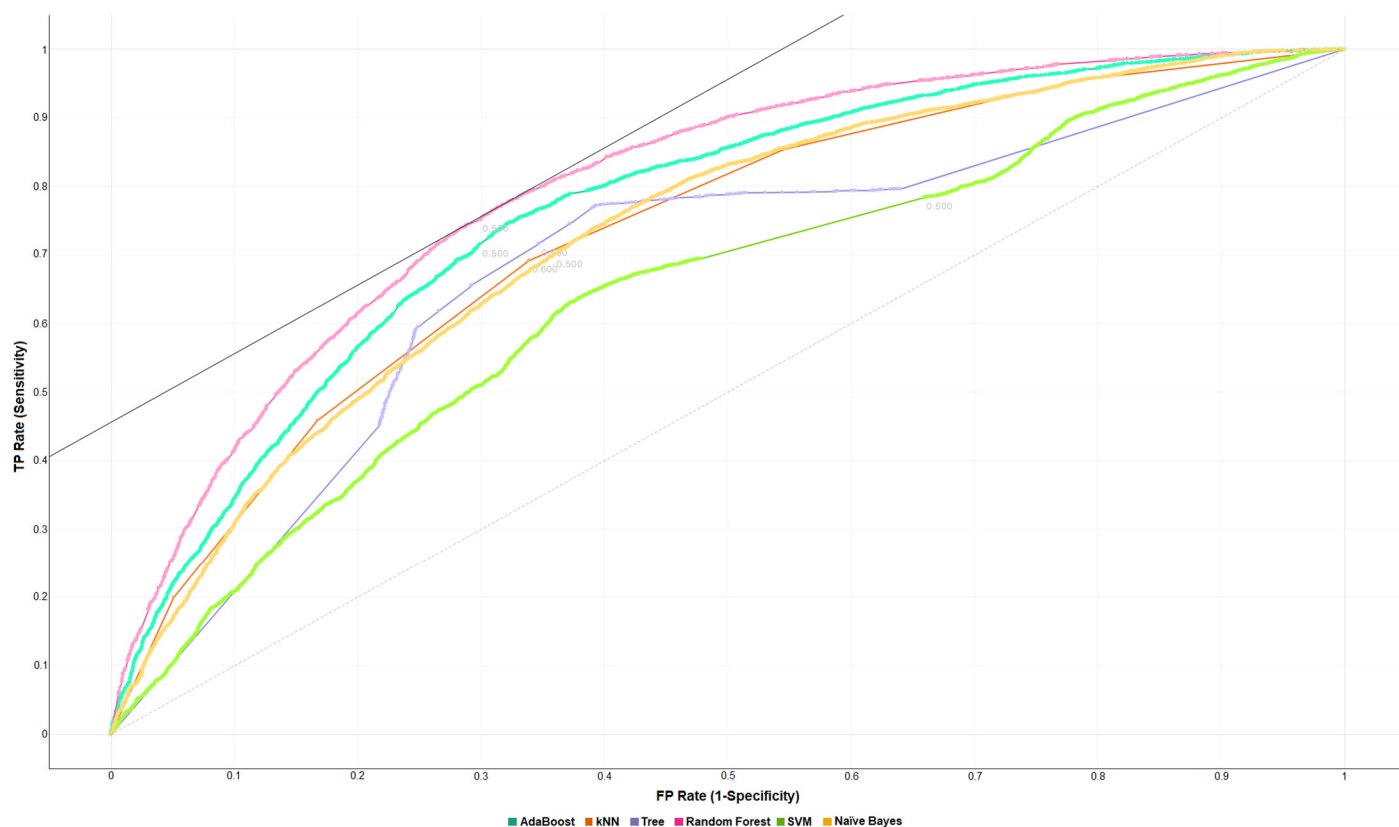


Figure 4. ROC curve for target = 1 with FP = 500, FN = 500, and target probability = 50.0%. The graph shows the ROC curves created by the false positive and false negative values with 10-folds stratified cross-validation. Each ROC curve is represented by a different color (See legend). The Random Forest algorithm shows the best value of AUC= 0.796 (See Supplementary Materials Table S2) for class 1 (records of schizophrenia).

4. Discussion

The diagnosis of mental disorders such as schizophrenia is the first step in a set of actions that are selected to save patients' lives or improve their health [59]. People with this disorder tend to be hospitalized frequently and have high rates of disability, imposing an economic cost on a general level. Therefore, from the viewpoint of our study, the implementation of predictive models in medical systems can be a helpful tool in the prevention of hospitalizations of patients with schizophrenia in the region of Castilla and León.

Multiple research studies use EEG and functional magnetic resonance imaging (fMRI) to identify patients with schizophrenia using machine learning techniques [19,41,60,61]. Other studies consider symptoms, cognitive functions, and non-verbal signals to explore the feasibility of automatic interview transcripts to classify schizophrenic patients [62,63].

In our study, the most relevant variables considered according to the ranking of features in Table 1 are Gender, Diag_Sec02_Code, and Age. Regarding the clinical aspect and the behavior of these variables, the gender of hospitalized patient records with schizophrenia is determinant. Men have a higher risk of suffering acute hospital admissions by the disease in the age range from 28 to 50 years compared to women. In the secondary diagnoses of these patients, they mainly present disorders of tobacco, alcohol, and cannabis abuse, determining substance abuse as a risk factor. The hospitalized patient records without schizophrenia do not show a significant difference regarding the patient's gender and age. The main physical diagnoses in this group are personality disorders and arterial hypertension, which are not shown as the main risk factors for hospitalized patients with schizophrenia in this study.

The classification methods used have a predictive accuracy of 65.7–72.7%, in terms of predicting hospitalized patients with schizophrenia (See Table 4). When comparing the performance metrics of the six machine learning methods, Random Forest is recommended as the best performance accuracy with 72.7%.

In the study proposed by the authors of [35], the classification of suicides in patients with schizophrenia presented an accuracy rate of 66.0% with an AUC = 0.67 and AUC = 0.70 for Random Forest and SVM, respectively, considering a sample of 345 patients. In [38], the authors diagnosed schizophrenia using four machine learning algorithms. The study presented an accuracy of 58.2–68.6% with an AUC of 0.63–0.68, with Random Forest obtaining the highest values, and SVM obtaining the lowest value.

The results presented in this research using machine learning algorithms show a reliable approach to predict hospitalized patients with schizophrenia in public hospitals in the Castilla and León region. However, the proposed study has several limitations. There is a relevant limitation regarding the significance features available in the data set used. From the 27 variables of the original data set, the authors have selected only 8, for a sample of 11,884 records. The selection of 8 features has the advantages of using less computational requirements in the development of the model and fast processing of the data collection. Furthermore, the age range of the participants in this study is broad. The authors aim to perform new experiments using patients with a strict age range to understand the impact of age in predicting schizophrenia.

Comparing the data set used in this study with other studies presented in the literature, the number of features ($N = 8$) used is inferior. In the study [37], the authors with a sample of 20 features obtained an average classification accuracy of the combined feature set of 78.2% using SVM. In [36] with a sample of 204/410 features, the authors obtained accuracy values of 86.04–90.69% using SVM, ANN, Random Forest, and Naïve Bayes algorithms. By reducing the number of features to 11, the values decrease to 82.55–83.72%. Therefore, the authors consider that it is necessary to increase the number of features to obtain metrics with higher values; however, this will require more computing resources and time for data collection. In general, in machine learning models, the sample size and predictor variables influence the training and effective validation of the predictions [49]. In our study, the classification of hospitalized patients with schizophrenia is based mainly on disease diagnostic features. If we compare our attributes with those used in studies such as [19,37,41], which focused on the prediction of schizophrenia using EEG and fMRI, we obtain performance metrics with a lower value. If we compare our attributes with the study [38], our model performance is higher. Therefore, the inclusion of new attributes related to neuroimaging can affect the performance of our models and allow improvement in patient-focused treatments. The database used in the study contains hospitalized patients with schizophrenia and other mental health diseases as the main diagnostic. Similar studies found in the literature show results between healthy and schizophrenia patients; this notable difference becomes a limitation of the study since predictive models cannot be developed using records of healthy patients. Furthermore, it is necessary to mention that in the study we only evaluated patients who are hospitalized, and people with schizophrenia from that region who are not hospitalized have not been included. Therefore, the results obtained are not generalizable to the entire population with schizophrenia in the Castilla and León region. Another limitation of our study is that we only applied six machine learning algorithms based on previous studies available in the literature [27,35,50]. However, it is necessary to emphasize that we presented the model parameters when most related studies did not provide this information, which was necessary to reproduce the results. In the future, the authors aim to use different machine learning techniques for improved accuracy.

The use of predictive models can be applied in different types of hospitals (private and/or public) to predict hospitalized patients with schizophrenia. These models can help in the hospital management of this type of disorder. Furthermore, the predictive model approach would be a useful tool for preventing hospitalizations of patients with schizophrenia in this region, reducing hospitalization costs.

5. Conclusions

This paper presents the development of prediction models of hospitalized patients with schizophrenia using k-NN, Decision Tree, SVM, Naïve Bayes, Random Forest, and AdaBoost machine learning algorithms and compares their system performance. The study identifies the AdaBoost (accuracy = 70.8% and AUC = 0.765) and Random Forest (accuracy = 72.7% and AUC = 0.796) ensemble algorithms as the best at classifying hospitalized patients with schizophrenia in this population. These algorithms present a higher value in their performance metrics than the SVM (accuracy = 65.7% and AUC = 0.641) algorithm. The results obtained suggest that the use of algorithms such as Random Forest can help in the hospital management of this type of disorder. The predictive modelling approach would be a useful tool in the prevention of hospitalizations of patients with schizophrenia in this region. Therefore, based on future lines, we will focus on the prediction of risk factors in hospitalized patients with schizophrenia and the prediction of readmissions in acute units of each health complex.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/s22072517/s1>, Table S1: Scores with target = 0; Table S2: Scores with target = 1.

Author Contributions: Conceptualization, S.G.A. and I.D.I.T.D.; methodology, S.G.A.; software, G.M.; validation, S.G.A. and D.A.; formal analysis, S.G.A. and G.M.; investigation, I.D.I.T.D.; resources I.D.I.T.D. and M.F.-M.; data curation, S.G.A.; writing—original draft preparation, S.G.A.; writing—review and editing, G.M.; visualization, D.A.; supervision, M.F.-M.; project administration, I.D.I.T.D. and M.F.-M.; funding acquisition, M.F.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Health Regional Service (GRS 1801/A/18).

Institutional Review Board Statement: All procedures performed in this study involving human participants were in accordance with the ethical standards of the Ethics Committee of the University of Valladolid (Ref. PI 20-1780).

Informed Consent Statement: Patient consent was waived because the personal data of the patients included in the study are anonymized.

Data Availability Statement: Third party data and restrictions apply to the availability of these data. The data were obtained from the Junta de Castilla y León and the Hospital of Zamora; therefore, they are available with the permission of both institutions.

Acknowledgments: Thanks to the Service of Psychiatry of the Provincial Hospital of Zamora, Spain, for the collaboration in this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pachange, S.; Joglekar, B.; Kulkarni, P. An ensemble classifier approach for disease diagnosis using Random Forest. In Proceedings of the 2015 Annual IEEE India Conference (INDICON), New Delhi, India, 17–20 December 2015; pp. 1–5. [\[CrossRef\]](#)
2. Zhao, Z.; Zhang, X.; Li, W.; Hu, X.; Qu, X.; Cao, X.; Liu, Y.; Lu, J. Applying Machine Learning to Identify Autism with Restricted Kinematic Features. *IEEE Access* **2019**, *7*, 157614–157622. [\[CrossRef\]](#)
3. Hou, C.; Zhong, X.; He, P.; Xu, B.; Diao, S.; Yi, F.; Zheng, H.; Li, J. Predicting Breast Cancer in Chinese Women Using Machine Learning Techniques: Algorithm Development. *JMIR Med. Inform.* **2020**, *8*, e17364. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Yoon, S.; Taha, B.; Bakken, S. Using a Data Mining Approach to Discover Behavior Correlates of Chronic Disease: A Case Study of Depression. *Stud. Health Technol. Inform.* **2014**, *201*, 71–78. [\[CrossRef\]](#)
5. Awad, A.; Bader-El-Den, M.; McNicholas, J.; Briggs, J. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *Int. J. Med. Inform.* **2017**, *108*, 185–195. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Tai, A.M.Y.; Albuquerque, A.; Carmona, N.E.; Subramaniepillai, M.; Cha, D.S.; Sheko, M.; Lee, Y.; Mansur, R.; McIntyre, R.S. Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artif. Intell. Med.* **2019**, *99*, 101704. [\[CrossRef\]](#) [\[PubMed\]](#)

7. Dhaka, P.; Johari, R. Big data application: Study and archival of mental health data, using MongoDB. In Proceedings of the 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 3–5 March 2016; pp. 3228–3232. [\[CrossRef\]](#)
8. Xu, Z.; Zhang, Q.; Li, W.; Li, M.; Yip, P.S.F. Individualized prediction of depressive disorder in the elderly: A multitask deep learning approach. *Int. J. Med. Inform.* **2019**, *132*, 103973. [\[CrossRef\]](#)
9. Charlson, F.J.; Ferrari, A.J.; Santomauro, D.F.; Diminic, S.; Stockings, E.; Scott, J.G.; McGrath, J.; Whiteford, H.A. Global Epidemiology and Burden of Schizophrenia: Findings from the Global Burden of Disease Study 2016. *Schizophr. Bull.* **2018**, *44*, 1195–1203. [\[CrossRef\]](#)
10. Orrico-Sánchez, A.; López-Lacort, M.; Muñoz-Quiles, C.; Sanfélix-Gimeno, G.; Díez-Domingo, J. Epidemiology of schizophrenia and its management over 8-years period using real-world data in Spain. *BMC Psychiatry* **2020**, *20*, 149. [\[CrossRef\]](#)
11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; Elsevier: Washington, DC, USA, 2013.
12. Kendler, K.S. Phenomenology of Schizophrenia and the Representativeness of Modern Diagnostic Criteria. *JAMA Psychiatry* **2016**, *73*, 1082–1092. [\[CrossRef\]](#)
13. GeethaRamani, R.; Sivaselvi, K. Data mining technique for identification of diagnostic biomarker to predict Schizophrenia disorder. In Proceedings of the 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India, 18–20 December 2014; pp. 1–8. [\[CrossRef\]](#)
14. Allende-Cid, H.; Zamora, J.; Alfaro-Faccio, P.; Alonso-Sanchez, M.F. A Machine Learning Approach for the Automatic Classification of Schizophrenic Discourse. *IEEE Access* **2019**, *7*, 45544–45553. [\[CrossRef\]](#)
15. Lurie, I.; Shoval, G.; Hoshen, M.; Balicer, R.; Weiser, M.; Weizman, A.; Krivoy, A. The association of medical resource utilization with physical morbidity and premature mortality among patients with schizophrenia: An historical prospective population cohort study. *Schizophr. Res.* **2021**, *237*, 62–68. [\[CrossRef\]](#) [\[PubMed\]](#)
16. De Pedro Cuesta, J.; Saiz Ruiz, J.; Roca, M.; Noguer, I. Mental health and public health in Spain: Epidemiological surveillance and prevention. *Psiquiatr. Biol.* **2016**, *23*, 67–73. [\[CrossRef\]](#)
17. McGrath, J.; Saha, S.; Chant, D.; Welham, J. Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. *Epidemiol. Rev.* **2008**, *30*, 67–76. [\[CrossRef\]](#)
18. Hor, K.; Taylor, M. Suicide and schizophrenia: A systematic review of rates and risk factors. *J. Psychopharmacol.* **2010**, *24* (Suppl. 4), 81–90. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Zhang, L. EEG Signals Classification Using Machine Learning for The Identification and Diagnosis of Schizophrenia. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 4521–4524. [\[CrossRef\]](#)
20. Chen, Z.H.; Yan, T.; Wang, E.L.; Jiang, H.; Tang, Y.Q.; Yu, X.; Zhang, J.; Liu, C. Detecting Abnormal Brain Regions in Schizophrenia Using Structural MRI via Machine Learning. *Comput. Intell. Neurosci.* **2020**, *2020*, 6405930. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Jin, H.; Mosweu, I. The Societal Cost of Schizophrenia: A Systematic Review. *PharmacoEconomics* **2017**, *35*, 25–42. [\[CrossRef\]](#)
22. Kovacs, G.; Almasi, T.; Millier, A.; Toumi, M.; Horváth, M.; Kóczyán, K.; Götze, A.; Kaló, Z.; Zemplenyi, A.T. Direct healthcare cost of schizophrenia—European overview. *Eur. Psychiatry* **2018**, *48*, 79–92. [\[CrossRef\]](#)
23. Tovar, D.; Cornejo, E.; Xanthopoulos, P.; Guarracino, M.R.; Pardalos, P.M. Data Mining in Psychiatric Research. *Psychiatr. Disord.* **2012**, *829*, 593–603. [\[CrossRef\]](#)
24. Bari Antor, M.; Jamil, A.H.M.; Mamtaz, M.; Monirujjaman Khan, M.; Aljahdali, S.; Kaur, M.; Singh, P.; Masud, M. A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer’s Disease. *J. Healthc. Eng.* **2021**, *2021*, 9917919. [\[CrossRef\]](#)
25. Bhagya Shree, S.R.; Sheshadri, H.S. An initial investigation in the diagnosis of Alzheimer’s disease using various classification techniques. In Proceedings of the 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India, 18–20 December 2014; pp. 1–5.
26. Sheshadri, H.S.; Shree, S.R.B.; Krishna, M. Diagnosis of Alzheimer’s Disease Employing Neuropsychological and Classification Techniques. In Proceedings of the 2015 5th International Conference on IT Convergence and Security (ICITCS), Kuala Lumpur, Malaysia, 24–27 August 2015; pp. 1–6. [\[CrossRef\]](#)
27. Tejeswinee, K.; Shomona, G.J.; Athilakshmi, R. Feature Selection Techniques for Prediction of Neuro-Degenerative Disorders: A Case-Study with Alzheimer’s and Parkinson’s Disease. *Procedia Comput. Sci.* **2017**, *115*, 188–194. [\[CrossRef\]](#)
28. Byeon, H. A Prediction Model for Mild Cognitive Impairment Using Random Forests. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 8–12. [\[CrossRef\]](#)
29. Cao, H.; Meyer-Lindenberg, A.; Schwarz, E. Comparative Evaluation of Machine Learning Strategies for Analyzing Big Data in Psychiatry. *Int. J. Mol. Sci.* **2018**, *19*, 3387. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Thabtah, F.; Kamalov, F.; Rajab, K. A new computational intelligence approach to detect autistic features for autism screening. *Int. J. Med. Inform.* **2018**, *117*, 112–124. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Bersimis, F.G.; Varlamis, I. Use of health-related indices and classification methods in medical data. In *Classification Techniques for Medical Image Analysis and Computer Aided Diagnosis*; Elsevier Inc.: San Diego, CA, USA, 2019. [\[CrossRef\]](#)
32. Alabi, R.O.; Elmusrati, M.; Sawazaki-Calone, I.; Kowalski, L.P.; Haglund, C.; Coletta, R.D.; Mäkitie, A.A.; Salo, T.; Almangush, A.; Leivo, I. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. *Int. J. Med. Inform.* **2020**, *136*, 104068. [\[CrossRef\]](#) [\[PubMed\]](#)

33. Kwakernaak, S.; van Mens, K.; Cahn, W.; Janssen, R. Using machine learning to predict mental healthcare consumption in non-affective psychosis. *Schizophr. Res.* **2020**, *218*, 166–172. [[CrossRef](#)]
34. Berardelli, I.; Rogante, E.; Sarubbi, S.; Erbuto, D.; Lester, D.; Pompili, M. The Importance of Suicide Risk Formulation in Schizophrenia. *Front. Psychiatry* **2021**, *12*, 779684. [[CrossRef](#)]
35. Hettige, N.C.; Nguyen, T.B.; Yuan, C.; Rajakulendran, T.; Baddour, J.; Bhagwat, N.; Bani-Fatemi, A.; Voineskos, A.N.; Chakravarty, M.M.; De Luca, V. Classification of suicide attempters in schizophrenia using sociocultural and clinical features: A machine learning approach. *Gen. Hosp. Psychiatry* **2017**, *47*, 20–28. [[CrossRef](#)]
36. Almutairi, M.M.; Alhamad, N.; Alyami, A.; Alshobbar, Z.; Alfayez, H.; Al-Akkas, N.; Alhiyafi, J.A.; Olatunji, S.O. Preemptive Diagnosis of Schizophrenia Disease Using Computational Intelligence Techniques. In Proceedings of the 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 1–3 May 2019; pp. 1–6. [[CrossRef](#)]
37. Shim, M.; Hwang, H.-J.; Kim, D.-W.; Lee, S.-H.; Im, C.-H. Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features. *Schizophr. Res.* **2016**, *176*, 314–319. [[CrossRef](#)]
38. Jo, Y.T.; Joo, S.W.; Shon, S.H.; Kim, H.; Kim, Y.; Lee, J. Diagnosing schizophrenia with network analysis and a machine learning method. *Int. J. Methods Psychiatr. Res.* **2020**, *29*, e1818. [[CrossRef](#)]
39. Khan, S.I.; Islam, A.; Hossen, A.; Zahangir, T.I.; Hoque, A.S.M.L. Supporting the Treatment of Mental Diseases using Data Mining. In Proceedings of the 2018 International Conference on Innovations in Science, Engineering and Technology (ICISSET), Chittagong, Bangladesh, 27–28 October 2018; pp. 339–344.
40. Deng, Y.; Hung, K.S.Y.; Lui, S.S.Y.; Chui, W.W.H.; Lee, J.C.W.; Wang, Y.; Li, Z.; Mak, H.K.F.; Sham, P.C.; Chan, R.C.K.; et al. Tractography-based classification in distinguishing patients with first-episode schizophrenia from healthy individuals. *Prog. Neuro Psychopharmacol. Biol. Psychiatry* **2019**, *88*, 66–73. [[CrossRef](#)]
41. Lee, J.; Chon, M.-W.; Kim, H.; Rathi, Y.; Bouix, S.; Shenton, M.E.; Kubicki, M. Diagnostic value of structural and diffusion imaging measures in schizophrenia. *NeuroImage Clin.* **2018**, *18*, 467–474. [[CrossRef](#)] [[PubMed](#)]
42. Zhu, L.; Wu, X.; Xu, B.; Zhao, Z.; Yang, J.; Long, J.; Su, L. The machine learning algorithm for the diagnosis of schizophrenia on the basis of gene expression in peripheral blood. *Neurosci. Lett.* **2021**, *745*, 135596. [[CrossRef](#)] [[PubMed](#)]
43. Jahmunah, V.; Oh, S.L.; Rajinikanth, V.; Ciaccio, E.J.; Cheong, K.H.; Arunkumar, N.; Acharya, U.R. Automated detection of schizophrenia using nonlinear signal processing methods. *Artif. Intell. Med.* **2019**, *100*, 101698. [[CrossRef](#)] [[PubMed](#)]
44. Johannesen, J.K.; Bi, J.; Jiang, R.; Kenney, J.G.; Chen, C.-M.A. Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults. *Neuropsychiatr. Electrophysiol.* **2016**, *2*, 3. [[CrossRef](#)]
45. Santos-Mayo, L.; San-José-Revuelta, L.M.; Arribas, J.I. A Computer-Aided Diagnosis System with EEG Based on the P3b Wave During an Auditory Odd-Ball Task in Schizophrenia. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 395–407. [[CrossRef](#)]
46. Lin, E.; Lin, C.-H.; Lane, H.-Y. A bagging ensemble machine learning framework to predict overall cognitive function of schizo-phrenia patients with cognitive domains and tests. *Asian J. Psychiatr.* **2022**, *69*, 103008. [[CrossRef](#)]
47. Bae, Y.J.; Shim, M.; Lee, W.H. Schizophrenia Detection Using Machine Learning Approach from Social Media Content. *Sensors* **2021**, *21*, 5924. [[CrossRef](#)]
48. Birnbaum, M.L.; Ernala, S.K.; Rizvi, A.F.; De Choudhury, M.; Kane, J.M. A Collaborative Approach to Identifying Social Media Markers of Schizophrenia by Employing Machine Learning and Clinical Appraisals. *J. Med. Internet Res.* **2017**, *19*, e289. [[CrossRef](#)]
49. Wang, K.Z.; Bani-Fatemi, A.; Adanty, C.; Harripaul, R.; Griffiths, J.; Kolla, N.; Gerretsen, P.; Graff, A.; De Luca, V. Prediction of physical violence in schizophrenia with machine learning algorithms. *Psychiatry Res.* **2020**, *289*, 112960. [[CrossRef](#)]
50. Alonso, S.G.; De La Torre-Díez, I.; Hamrioui, S.; López-Coronado, M.; Barreno, D.C.; Nozaleda, L.M.; Franco, M. Data Mining Algorithms and Techniques in Mental Health: A Systematic Review. *J. Med. Syst.* **2018**, *42*, 161. [[CrossRef](#)]
51. Alonso, S.G.; Sainz-De-Abajo, B.; De La Torre-Díez, I.; Franco-Martin, M. Health Care Management Models for the Evolution of Hospitalization in Acute Inpatient Psychiatry Units: Comparative Quantitative Study. *JMIR Ment. Health* **2020**, *7*, e15776. [[CrossRef](#)] [[PubMed](#)]
52. Commission on Professional and Hospital Activities. The International Classification of Diseases, 9th Revision, Clinical Modification. 2014. Available online: https://www.msssi.gob.es/estadEstudios/estadisticas/docs/CIE9MC_2014_def_accesible.pdf (accessed on 6 May 2021).
53. CRAN.R-Project. Dplyr Package. Available online: <https://cran.r-project.org/package=dplyr> (accessed on 10 September 2021).
54. CRAN.R-Project. TidyR Package. Available online: <https://cran.r-project.org/package=tidyR> (accessed on 10 September 2021).
55. Zamora Saiz, A.; Quesada González, C.; Hurtado Gil, L.; Mondéjar Ruiz, D. Data Analysis with R. In *An Introd to Data Anal R*; Springer: Cham, Switzerland, 2020; pp. 183–271. [[CrossRef](#)]
56. Singh, D.; Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **2020**, *97*, 105524. [[CrossRef](#)]
57. Breiman, L. Random Forest. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
58. Kaur, H.; Kumari, V. Predictive modelling and analytics for diabetes using a machine learning approach. *Appl. Comput. Inform.* **2022**, *18*, 90–100. [[CrossRef](#)]
59. Abou-Warda, H.; Belal, N.A.; El-Sonbaty, Y.; Darwish, S. A Random Forest Model for Mental Disorders Diagnostic Systems. In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016, Cairo, Egypt, 24–26 October 2016; Springer: Cham, Switzerland, 2017; pp. 670–680. [[CrossRef](#)]

60. Dvey-Aharon, Z.; Fogelson, N.; Peled, A.; Intrator, N. Schizophrenia Detection and Classification by Advanced Analysis of EEG Recordings Using a Single Electrode Approach. *PLoS ONE* **2015**, *10*, e0123033. [[CrossRef](#)]
61. Kalmady, S.V.; Greiner, R.; Agrawal, R.; Shivakumar, V.; Narayanaswamy, J.C.; Brown, M.R.G.; Greenshaw, A.J.; Dursun, S.M.; Venkatasubramanian, G. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. *NPJ Schizophr.* **2019**, *5*, 2. [[CrossRef](#)]
62. Xu, S.; Yang, Z.; Chakraborty, D.; Tahir, Y.; Maszczyk, T.; Chua, V.Y.H.; Dauwels, J.; Thalmann, D.; Thalmann, N.M.; Tan, B.-L.; et al. Automatic Verbal Analysis of Interviews with Schizophrenic Patients. In Proceedings of the 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), Shanghai, China, 19–21 November 2018; pp. 1–5. [[CrossRef](#)]
63. Walsh-Messinger, J.; Jiang, H.; Lee, H.; Rothman, K.; Ahn, H.; Malaspina, D. Relative importance of symptoms, cognition, and other multilevel variables for psychiatric disease classifications by machine learning. *Psychiatry Res.* **2019**, *278*, 27–34. [[CrossRef](#)]