Computers in Biology and Medicine 167 (2023) 107696

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Advancing prognostic precision in pulmonary embolism: A clinical and laboratory-based artificial intelligence approach for enhanced early mortality risk stratification☆

Seyed-Ali Sadegh-Zadeh [a,*], Hanie Sakha [a], Sobhan Movahedi [b], Aniseh Fasihi Harandi [b], Samad Ghaffari [c], Elnaz Javanshir [c], Syed Ahsan Ali [d], Zahra Hooshanginezhad [e], Reza Hajizadeh [f]

[a] Department of Computing, School of Digital, Technologies and Arts, Staffordshire University, Stoke-on-Trent, England, United Kingdom
[b] Azad University, Science and Research, Tehran, Iran
[c] Cardiovascular Research Centre, Tabriz University of Medical Sciences, Tabriz, Iran
[d] Health Education England West Midlands, Birmingham, England, United Kingdom
[e] Department of Cardiovascular Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
[f] Department of Cardiology, Urmia University of Medical Sciences, Urmia, Iran

ARTICLE INFO

Keywords:
Pulmonary embolism
Early mortality prediction
Machine learning algorithms
Oversampling techniques

ABSTRACT

Background: Acute pulmonary embolism (PE) is a critical medical emergency that necessitates prompt identification and intervention. Accurate prognostication of early mortality is vital for recognizing patients at elevated risk for unfavourable outcomes and administering suitable therapy. Machine learning (ML) algorithms hold promise for enhancing the precision of early mortality prediction in PE patients.

Objective: To devise an ML algorithm for early mortality prediction in PE patients by employing clinical and laboratory variables.

Methods: This study utilized diverse oversampling techniques to improve the performance of various machine learning models including ANN, SVM, DT, RF, and AdaBoost for early mortality prediction. Appropriate oversampling methods were chosen for each model based on algorithm characteristics and dataset properties. Predictor variables included four lab tests, eight physiological time series indicators, and two general descriptors. Evaluation used metrics like accuracy, F1_score, precision, recall, Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC) curves, providing a comprehensive view of models' predictive abilities.

Results: The findings indicated that the RF model with random oversampling exhibited superior performance among the five models assessed, achieving elevated accuracy and precision alongside high recall for predicting the death class. The oversampling approaches effectively equalized the sample distribution among the classes and enhanced the models' performance.

Conclusions: The suggested ML technique can efficiently prognosticate mortality in patients afflicted with acute PE. The RF model with random oversampling can aid healthcare professionals in making well-informed decisions regarding the treatment of patients with acute PE. The study underscores the significance of oversampling methods in managing imbalanced data and emphasizes the potential of ML algorithms in refining early mortality prediction for PE patients.

☆ We acknowledge that the transparency and disclosure of potential conflicts of interest are essential to maintaining the integrity and credibility of scientific research. By providing this declaration, we assure readers that our findings and interpretations in this paper are unbiased and impartial, and any potential conflicts of interest have been disclosed accurately and transparently.

* Corresponding author.
E-mail address: ali.sadegh-zadeh@staffs.ac.uk (S.-A. Sadegh-Zadeh).

https://doi.org/10.1016/j.compbiomed.2023.107696
Received 19 July 2023; Received in revised form 25 October 2023; Accepted 7 November 2023
Available online 11 November 2023

1. Introduction

An occlusion of a blood vessel in the main or branching arteries of the lung by clot is known as a pulmonary embolism (PE) [1]. The embolization usually origins from veins of lower and upper extremities results in pulmonary artery blockage [2]. Untreated PE is expected to have a mortality rate of 30% while treated PE is estimated to have a mortality rate of 8%. PE is a life-threatening condition [3]. Resources and staffing are frequently constrained in medical centers due to congestion. Additionally, early in the medical procedure, there is very little clinical data accessible [4]. When a patient is at risk of developing a pulmonary embolism or possibly death unexpectedly, it can be challenging for clinicians to appropriately monitor all patient conditions. As a result, precise triage methods that can spot high-risk individuals are being taken into consideration [5–7].

Based on Electronic Health Record (EHR) data, which primarily contains demographic data, physiological measurements, laboratory tests, clinical observations, and therapeutic options, the primary motivation for predicting mortality among this group of patients is to compare the effectiveness of medications, care recommendations, surgery, and other interventions [8,9]. As a result, a number of scoring techniques were suggested for the patient's illness severity assessment and outcome prediction, including the Emergency Severity Index [10], Acute Physiologic Assessment and Chronic Health Evaluation [11], Mortality Probability Models [12], and the Sequential Organ Failure Assessment [13]. Although widely used, they had a rather low accuracy and were typically employed for benchmarking [14,15].

Machine learning is effective at uncovering useful patterns from vast amounts of data. It may investigate the complex relationships between many components and forecast shifting patterns. Clinical EHR data in abundance and publicly accessible datasets like Mimic-iv [16] and HiRID [17] have aided in the development and implementation of machine learning in the study of medical data. The effectiveness and adaptability of machine learning in the critical care setting were demonstrated by the PhysioNet Computing in Cardiology Challenge 2012 [18] and the WiDS (Women in Data Science) Datathon 2020 [19], both of which sought to predict mortality of patients in intensive care units (ICUs).

Clinical decision support systems have become more prevalent in recent years thanks to the use of various machine learning algorithms [20,21]. The majority of research to date have shown that machine learning models perform better than clinical grading systems [22,23]. For instance, Klug et al. [24] used gradient boosting and logistic regression models for both short-term mortality prediction and early mortality prediction (up to 2 days after hospital registration) (2–30 days post hospital registration). In comparison to severity scores like the Shock Index, Modified Shock Index, and Aged Shock Index, higher AUC might be attained [25]. Deep learning models were taken into consideration for postoperative in-hospital mortality prediction by Lee et al. [26] and Hofer et al. [27].

Despite the growing interest in utilizing machine learning algorithms for predictive medical analysis, there remains a noticeable gap in the literature when it comes to early mortality prediction specifically for patients with pulmonary embolism. While various studies have explored mortality prediction in other medical contexts, the unique characteristics of pulmonary embolism, its rapid onset, and the complexity of physiological responses have not been adequately addressed using these advanced techniques. Moreover, existing methodologies often rely on simplified representations of time series data, neglecting the intricate temporal dynamics that could hold crucial predictive information. This research aims to bridge this gap by introducing a comprehensive approach that integrates a wide array of clinical and laboratory variables, enabling a more accurate assessment of early mortality risk. By leveraging the power of machine learning, this study endeavors to enhance the precision of prognostic outcomes for patients with pulmonary embolism, ultimately contributing to improved clinical decision-making and patient care.

In contrast to conventional approaches that often rely on traditional statistical methods and limited variables to predict mortality in cases of pulmonary embolism, our research introduces a novel and comprehensive perspective. We acknowledge that existing models frequently overlook the intricate temporal patterns inherent in time series-based variables, potentially leading to the loss of critical information. Furthermore, previous studies have tended to incorporate only a restricted set of easily quantifiable parameters, neglecting the broader spectrum of potential predictors. Remarkably, our study pioneers the application of machine learning techniques for early mortality prediction in patients with pulmonary embolism, a pioneering avenue that, to our knowledge, remains unexplored in the current literature. We leverage an extensive array of raw features sourced from diverse clinical contexts, allowing for a more holistic representation of the patient's condition. These novel predictors not only expand the scope of mortality prediction but also have the potential to complement existing methodologies. By extensively detailing the performance of applied models, our study provides invaluable insights into their capabilities and effectiveness, contributing to the advancement of prognostic precision in the realm of pulmonary embolism.

The remainder of the paper is set up as follows. Section 3 discusses the study's conceptual framework, which aims to develop a framework for predicting hospital mortality in patients with pulmonary embolism using and machine learning theory. The dataset description, as well as the tools and techniques used in this work, are described in Section 3. The experimental findings on a testing device are presented in Section 4, which is followed by a discussion and closing remarks in Section 5.

2. Literature review

Acute PE is a critical medical emergency with potentially severe consequences [28]. Early and accurate prediction of mortality risk in patients with PE is of paramount importance for timely intervention and optimal patient care. Traditional scoring systems and severity assessment models, while widely used, often lack the desired accuracy and fail to capture the intricate temporal dynamics and comprehend'sive range of predictors inherent in the condition [29].

The landscape of predictive modelling in medical contexts has been undergoing a transformation with the advent of ML techniques [30]. Machine learning approaches have demonstrated remarkable potential in extracting meaningful patterns and relationships from complex medical data. This potential has been showcased in various medical domains, including intensive care units (ICUs) and postoperative settings, where ML models have outperformed traditional scoring systems [31].

Despite the growing interest in applying machine learning to medical prognostication, there remains a noticeable gap in the literature when it comes to early mortality prediction specifically in the context of acute pulmonary embolism. While some studies have explored mortality prediction in other medical scenarios [32–36], the unique characteristics of PE, such as its rapid onset and intricate physiological responses, present challenges that have yet to be comprehensively addressed by advanced techniques.

Traditional approaches to mortality prediction often rely on simplified representations of time series data and a limited set of easily quantifiable variables [37]. This approach, though practical for clinical implementation, risks overlooking critical temporal patterns and broader predictive factors. Consequently, there is a need for a more comprehensive and accurate approach that leverages the full potential of machine learning algorithms [38].

In this context, the present study introduces a pioneering approach that integrates a wide array of clinical and laboratory variables for enhanced early mortality prediction in PE patients. By employing machine learning algorithms, this study aims to bridge the gap between traditional scoring systems and the untapped potential of machine

learning, enabling a more accurate assessment of mortality risk. The proposed study not only expands the range of predictors used for early mortality prediction but also acknowledges the challenges posed by imbalanced data distribution in the death class. Imbalanced data distribution can undermine model performance, which necessitates the application of oversampling techniques to achieve a more balanced representation of classes [39].

The current literature underscores the potential of machine learning algorithms in advancing prognostic precision for patients with acute pulmonary embolism. This study fills a crucial gap by introducing a comprehensive approach that leverages a diverse range of predictors and oversampling techniques to enhance early mortality risk stratification. The findings of this study could revolutionize clinical decision-making and patient care in the context of acute PE.

3. Framework of study

Fig. 1 illustrates the process of converting the clinical data to decision using machine learning and Fig. 2 demonstrates a paradigm for predicting hospital mortality that was suggested based on the principles of machine learning.

A heatmap in machine learning is a graphical representation of data where individual values contained in a matrix are represented as colors. It's an effective way to visualize complex data sets, enabling easy identification of patterns, correlations, and trends. Heatmaps are particularly useful in machine learning for exploring the relationship between features or understanding the distribution of data. In the context of the study you're referring to, the heatmap of the dataset in Fig. 3 would provide a visual representation of the data's structure and relationships. It could show how different variables interact with each other, highlight areas of high and low density, or illustrate other key insights drawn from the data. To get a detailed understanding of the specific insights from this heatmap, it's essential to view Fig. 3 directly, as it would contain the specific color-coded information pertinent to this study's dataset. To begin with, data cleaning was used to address issues with the patient's original data, including data format inconsistencies, missing data, outliers, and duplicate data. Second, the features were chosen. Third, once the dataset was randomly split into a training set (80%) and testing set (20%), 5 machine learning models were created. By randomly rearranging the data in each independent run, 5-fold cross validation was carried out ten times to produce a more reliable model evaluation.

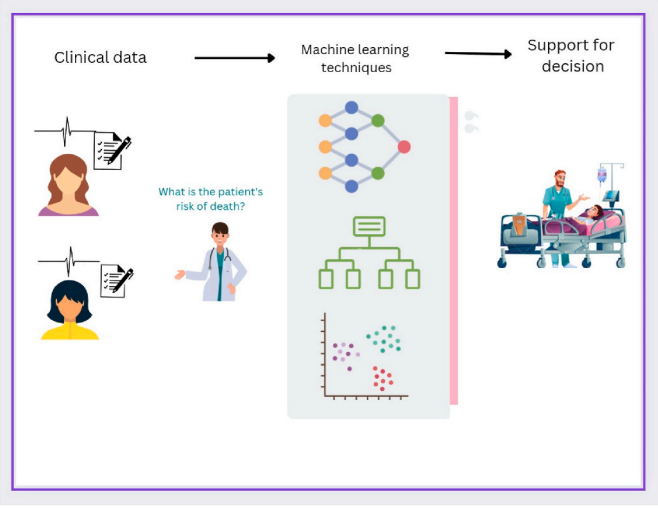

Fig. 1. Process of converting the clinical data to decision using machine learning.

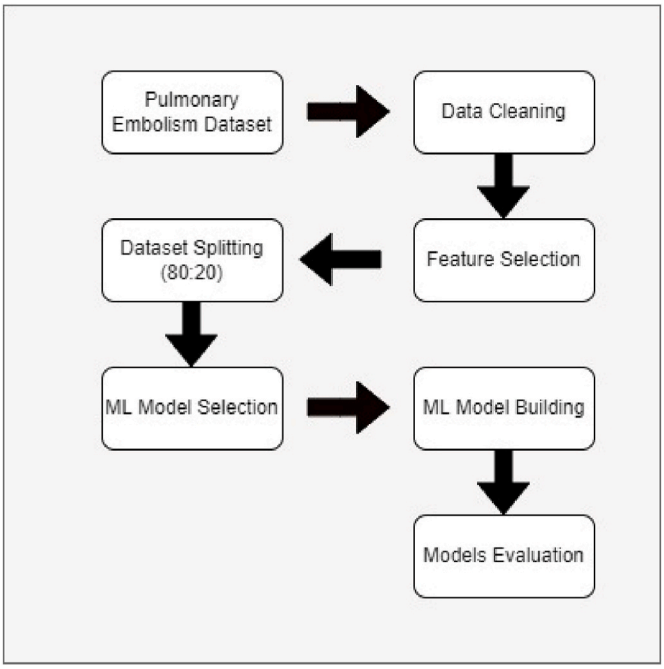

Fig. 2. The framework of early death mortality prediction.

4. Materials and methods

This section outlines the foundational components of the research process, detailing the materials and methodologies employed to address the research objectives. The successful execution of a scientific study depends heavily on the careful selection of materials and the meticulous application of appropriate methods.

4.1. Dataset description

Between July 2013 and December 2020 the list of all inpatients aging at least 18 years old, admitted with definite diagnosis of acute pulmonary embolism in two tertiary centers in northwest of Iran were abstracted. (Shahid Madani Heart Hospital and Ayatollah Taleghani hospital, two university-affiliated referral centers in Tabriz and Urmia respectively, in Iran). We used International Classification of Diseases (ICD) appropriate codes by a computer-assisted search. Patients with chronic pulmonary embolism or non-thrombotic pulmonary embolism were excluded from this study. Finally, 604 consecutive patients with acute PE, diagnosed by computerized tomography (CT) angiography (Siemens 32 and 64 slice computed tomography scanners) were included in this study. In-hospital mortality was defined as any death in patients admitted with pulmonary embolism during hospital course, after excluding mortality due to other causes such as bleeding or stroke. Heat-map correlation for dataset features is represented in Fig. 2.

4.2. Study features and outcome

The study included four laboratory tests, 12 physiological time series vital indicators, including heart rate, blood pressure, oxygen saturation, and a history of heart failure, chronic obstructive pulmonary embolism, syncope, right ventricular enlargement, and S1T3Q3 in the patient's electrocardiogram, taking fibrinolytic therapy, in hospital death, admission days and embolectomy treatment and finally home mortality, as well as 2 general descriptors (age and gender). When the patient was admitted to the hospital, demographic data was gathered. The monitoring equipment in the ICU took automatic measurements of the vital signs continuously, and the medical personnel checked the accuracy of the readings. Multiple observations were made during laboratory

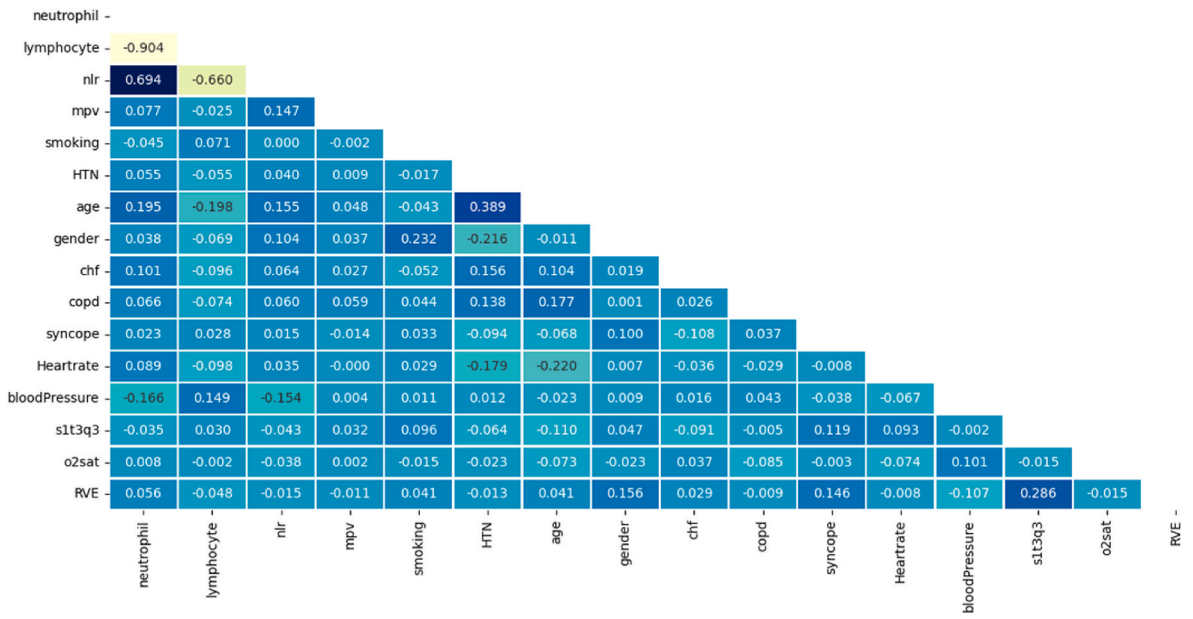

Fig. 3. Heat-map correlation for dataset features.

testing, and in some situations, these observations were made only once, more than once, or not at all. Each observation had a time-stamp attached to it that showed the time it was made in hours and minutes. The clinically significant factors in the dataset are displayed in Table 1. Univariate and multivariable logistic regression analysis for Mortality is represented in Table 2.

4.3. Data pre-processing

Data pre-processing has been shown to begin with data cleaning, which corrects inaccurate, noisy, or inconsistent initial records [40–42]. The data format was standardised, redundant or mismatched elements were deleted, and outright outliers were given null values. The volume of numerical data can be intuitively understood using visualisation approaches. The predictor variables were represented graphically using a boxplot approach, which also revealed whether any data still fell outside the upper and lower quartiles and, if so, replaced any such outliers with the maximum and minimum values of the normal range. Real data always has a high rate of missing data. While the data from the prior time was used to pad the current missing data in time series vital signs, the median was utilized to fill in the gaps in laboratory tests.

To transform category values into numerical types is the goal of data encoding, gender was encoded as a binary variable in this study (0 = female, 1 = male). The binary class (0 = patient died, 1 = patient survived) used to describe the model output's representation of patient mortality. The min-max normalising approach was used to rescale the range in for each feature.

Table 1
Variables of pulmonary embolism mortality prediction.

Variables		Total (n = 604)	No mortality (n = 542)	Mortality (n = 62)	p-value	Missing (%)
age		61.90 ± 17.18	65.56 ± 17.46	61.48 ± 17.11	0.076[a]	0(0%)
Lymphocyte		19.22 ± 8.96	19.45 ± 8.39	17.20 ± 12.83	0.182[a]	0(0%)
Neutrophil to lymphocyte ratio		5.92 ± 5.08	8.76 ± 8.09	5.59 ± 4.52	0.003[a]	0(0%)
Mean platelet volume		9.98 ± 0.91	10.30 ± 0.93	9.94 ± 0.90	0.003[a]	0(0%)
Oxygen saturation		85.93 ± 9.22	86.40 ± 8.63	81.81 ± 12.70	0.007[a]	0(0%)
Smoking		81 (13.4%)	72(13.3%)	9(14.5%)	0.787[b]	0(0%)
Neutrophil		73.94 ± 9.65	73.51 ± 9.39	77.68 ± 11.06	0.001[a]	0(0%)
Hypertension		244 (40.4%)	215(39.7%)	29(46.8%)	0.280[b]	0(0%)
Gender(F/M)		297/307	259/283	38/24	0.044[b]	0(0%)
Congestive heart failure		52(8.6%)	45(8.3%)	7(11.3%)	0.427[b]	0(0%)
Chronic obstructive pulmonary disease		55(9.1%)	51(9.4%)	4(6.5%)	0.640[c]	0(0%)
Syncope		66(10.9%)	62(11.4%)	4(6.5%)	0.287[c]	0(0%)
S1T3Q3		213(35.3%)	187(34.5%)	26(41.9%)	0.246[b]	0(0%)
Right ventricular enlargement		403(66.7%)	352(64.9%)	51(82.3%)	0.006[b]	0(0%)
Blood pressure	≤90	47(7.8%)	35(6.5%)	12(19.4%)	<0.001[b]	0(0%)
	≥100	520(86.1%)	481(88.7%)	39(62.9%)		
	90–100	37(6.1%)	26(4.8%)	11(17.7%)		
Heart rate	<100	313(51.8%)	280(51.7%)	33(53.2%)	0.815[b]	0(0%)
	≥100	291(48.2%)	262(48.3%)	29(46.8%)		
Fibrinolytic		124(20.5%)	105(19.4%)	19(30.6%)	0.037[b]	0(0%)
Embolectomy		29(4.8%)	21(3.9%)	8(12.9%)	0.002[b]	0(0%)
Admission days		9.75 ± 6.52	8.31 ± 13.79	9.91 ± 5.07	0.366[a]	0(0%)

[a] Independent Samples Test.
[b] Chi-Square Tests.
[c] Fisher's Exact Test.

Table 2
Univariate and multivariable logistic regression analysis for Mortality.

Variable	Univariate			multivariate		
	Unadjusted OR	95% CI	P-value[a]	Adjusted OR	95% CI	P-value[a]
Neutrophil count	1.052	1.020–1.085	0.001	1.002	0.958–1.047	0.944
Lymphocyte count	0.968	0.937–1.001	0.060			
Neutrophil to lymphocyte ratio	1.083	1.037–1.131	<0.001	1.081	1.006–1.161	0.034
Mean platelet volume	1.492	1.141–1.950	0.003	1.509	1.123–2.027	0.006
Smoking	1.108	0.524–2.34	0.787			
Hypertension	1.337	0.789–2.266	0.281			
Age	1.015	0.998–1.032	0.078			
Gender(female)	1.730	1.010–2.963	0.046	2.610	1.429–4.766	0.002
Congestive heart failure	1.406	0.605–3.268	0.429			
Chronic obstructive pulmonary disease	0.664	0.232–1.904	0.446			
Syncope	0.534	0.187–1.521	0.240			
Heart rate	1.065	0.629–1.803	0.815			
bloodPressure (≥100)	0.192	0.088–0.417	<0.001	0.233	0.101–0.539	0.001
ST3Q3	1.371	0.803–2.340	0.247			
Oxygen saturation	.961	0.940–0.983	<0.001	0.964	0.941–0.987	0.002
Right ventricular enlargement	2.503	1.274–4.915	0.008	3.287	1.520–7.107	0.002

+Chi-Square Tests.
#Fisher's Exact Test.
[a] Logistic Regression.

4.4. Machine learning algorithms

Artificial Neural Networks, Support Vector Machines, Decision Trees, Random Forests, and Adaptive Boosting were five machine learning methods that we used in this study to predict the mortality of pulmonary embolism. These algorithms were chosen due to their effective implementation and simple performance [43,44].

Cross validation has been used to evaluate and find the best hyperparameters, and the most data (80% of the total data) was divided into four parts and finally the test data was measured with the best hyperparameters. A simple MLP was used for neural networks and it had a hidden layer. MLPs may consist of several layers and each layer has a certain number of perceptron's. In each perceptron, the values of the previous columns are multiplied by their weights and finally added together, and the final value may be included in an activation function. Scikit-learn and TensorFlow were used for implementation.

The SVM algorithm looks for the best decision boundary with the largest margin. Margin is the distance of the closest data points to the decision boundary. One of the most important parameters of this algorithm is Kernel. In this study, it is only from RBF or Radial basis function, Polynomial and Linear kernels. Kernel is a mathematical method to increase the dimensions of the dataset so that calcification can be done with a decision boundary. Decision trees are made by nodes and leaves. During the training of the model, a feature is measured in each node, so that the node is divided into two parts based on that, so that the data points go to a certain path. Finally, each path reaches a leaf where classification is done [45].

Random forest is called ensemble algorithm and belongs to the category of bagging algorithms. In simple words, random forest includes a wide network of decision trees, which explains the name of this algorithm. To train each of these decision trees, parts of the data are selected randomly. Finally, for the prediction of a data point, the result of each decision tree is aggregated. Adaboost stands for Adaptive boosting, and this algorithm, like random forest, is an ensemble algorithm, and the only difference with Random forest algorithm is that when training the model, the data is not randomly separated, but the entire dataset is given to each model. Each model is called a weak learner and the consensus of these models includes the Adaboost algorithm. In each iteration when the dataset is used to train weak learners, the weights of the data are updated [46].

Also, due to the lack of data balance, data augmentation is felt. In this study, three methods were used for this purpose: oversampling, undersampling, and oversampling and under-sampling combined. For oversampling, the ADASYN algorithm was used, which obtained the best result. Random under-sampling was used for under-sampling. For combined oversampling and under-sampling, SMOTEENN algorithm is used, where SMOTE is used for oversampling and ENN is used for undersampling. In the following, these algorithms are described.

ADASYN and SMOTE algorithms are similar in that they add synthetic data to the minority class, but their functions are different. SMOTE considers a new data point between two data points and determines the class of that new data point based on the neighbouring data points. The difference with SMOTE is that ADASYN considers the dispersion of minority class data points to perform oversampling. Because of this, ADASYN is suitable for datasets that have a lot of outliers, and our dataset has a small amount of this problem. The random under sampler algorithm randomly removes data points of the majority class. The ENN or Edited nearest neighbour algorithm removes data points close to the decision boundary in the majority class.

4.5. Model evaluation

The most logical statistic for assessing classification model performance is predictive accuracy [47]. To assess the overall model performance, we also employed AUC, recall, F1-score, and precision. Considering that medical research usually has imbalanced datasets and hospital death is important for us, we have to use AUC, recall, f1_score, and precision measurement techniques. Precision is a measure of hospital deaths that are correctly predicted.

$$precision = \frac{Number\ of\ hospital\ death\ correctly\ predicted}{Number\ of\ people\ predicted\ to\ die}$$

Recall is a measure of hospital deaths that are correctly predicted on the total number of people who actually died.

$$recall = \frac{Number\ of\ hospital\ death\ correctly\ predicted}{Number\ of\ hospital\ death\ at\ the\ dataset}$$

F1_score is a measure that tries to maintain a middle ground between precision and recall.

$$f1\_score = 2 * \frac{recall * precision}{recall + precision}$$

Area under the ROC Curve is referred to as AUC. roc is the graph of the ratio of FPR (false positive rate) to TPR (true positive rate), where positive is hospital death. TPR is the rate of Recall and FPR is the complementary measure of specificity. specificity is the measure of

people who are correctly predicted to be alive to the total number of people who are predicted to be alive. auc tells us how likely the model is to correctly distinguish hospital death from survival.

5. Results

The study aimed to predict hospital deaths using machine learning models, including decision trees, random forests, adaptive boosting, support vector machines (SVM), and neural networks using multilayer perceptron (MLP). The dataset used had a high imbalance of hospital death class, and different oversampling techniques were explored. The random-over-sampler, ADASYN, and SVM-Smote techniques were chosen, as they showed the best results in overcoming the dataset's class imbalance. The hyperparameters for each model were identified using grid search.

The decision tree model was trained with hyperparameters, including max_depth = 5, max_leaf_nodes = 13, and min_samples_leaf = 9. The SVM model's hyperparameters included C = 1, gamma = 1, and kernel = "poly." The neural network model using MLP had 20 layers, with eight neurons per layer, solver Adam, and a maximum of 1000 iterations. The adaptive boosting approach used n_estimates = 50. Finally, the random forest model had hyperparameters max_features = 14 and n_estimaters = 14, which yielded the best accuracy in predicting death.

The modelling outcomes were displayed in Fig. 4, showing the performance of each machine learning model used. The results indicated that the random forest model had the best accuracy in predicting hospital deaths.

The ROC curve was used to evaluate the performance of each machine learning model, as shown in Fig. 5. The AUC values for the decision-tree model, MLP, adaptive boosting, and SVM were 0.771, 0.794, 0.872, and 0.846, respectively. The adaptive boosting model showed the highest AUC value of 0.872, indicating the best predictive performance in comparison to other models.

The experiments found that the random forest model with hyperparameters max_features = 14 and n_estimaters = 14 provided the best accuracy in predicting hospital deaths. The study also identified the usefulness of oversampling techniques such as random-over-sampler, ADASYN, and SVM-Smote in handling the high class imbalance in the dataset. These findings can be valuable for healthcare professionals in improving patient care and management by providing early predictions of hospital mortality.

Fig. 6 presents the results of these experiments performed without implementing oversampling, enabling readers to perform a comparative analysis with the findings shown in Fig. 4, which leveraged the advantages of oversampling. Importantly, the outcomes achieved through oversampling demonstrate a significant improvement, highlighting its substantial contribution to enhancing our results.

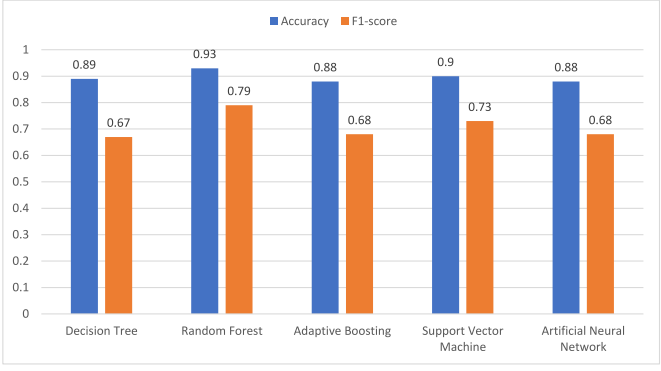

Fig. 4. The accuracy value and the f1-score value for the output of the models using oversampling.

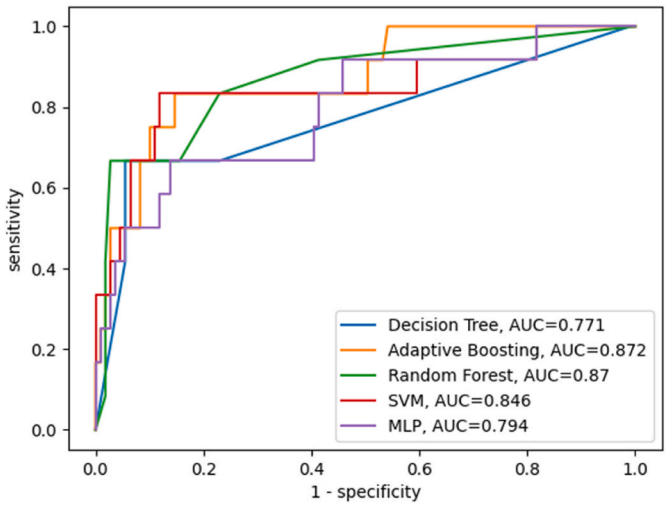

Fig. 5. ROC curve for five machine learning models: decision tree, random forest, adaptive boosting, svm and neural network.

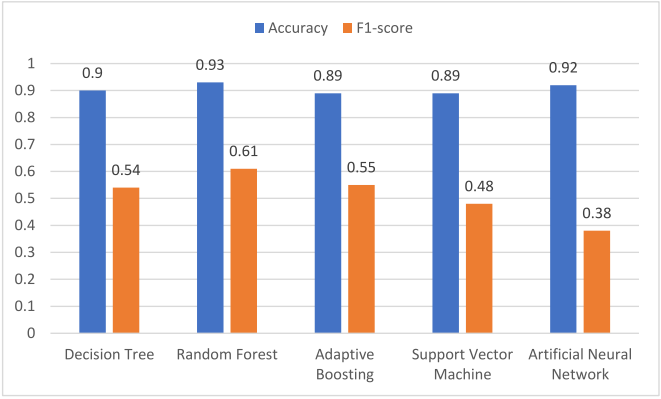

Fig. 6. The accuracy value and the f1-score value for the output of the models without oversampling.

6. Discussion

In this study, multiple oversampling techniques were applied to different machine learning models to improve the performance of the models in the task of early mortality prediction in pulmonary embolism. The choice of oversampling technique for each model was based on the characteristics of the algorithm and the nature of the data. For instance, the decision tree model was trained using SVM-SMOTE, a hybrid method that combines the Support Vector Machine (SVM) algorithm with Synthetic Minority Over-sampling Technique (SMOTE), while random oversampling was used for neural networks. The evaluation results showed that each oversampling technique had a different impact on the performance of the models, highlighting the importance of selecting the appropriate technique for each algorithm.

One of the challenges in the early mortality prediction task is the imbalance of the death class, which can affect the performance of the models. To address this issue, the oversampling technique was applied to balance the distribution of samples among the classes. The results showed that the random forest model with random oversampling had the best performance among the five models evaluated. The evaluation metrics, such as accuracy, f1_score, precision, and recall, indicated that the model with oversampling was highly reliable for predicting the death class. The high accuracy and precision, combined with high recall, suggest that the model can predict the death class with a high degree of accuracy, while minimizing the false positives and false negatives.

The observed disparity in accuracy values between Fig. 4, which represents models with oversampling, and Fig. 6, which portrays models without oversampling, raises important insights into the impact of oversampling techniques in the context of machine learning. Firstly, it's essential to acknowledge that oversampling is primarily employed to address class imbalance issues within datasets, ensuring that the minority class has a sufficient number of instances for the model to learn from. However, it's evident from the results that oversampling hasn't universally improved the models' predictive accuracy. In the case of the decision tree, adaptive boosting, and artificial neural network, accuracy values have deteriorated with oversampling. This counterintuitive outcome suggests that oversampling may have introduced noise or increased the complexity of the datasets, thereby hampering the models' ability to generalize effectively. It also emphasizes that addressing class imbalance is not a one-size-fits-all solution and should be applied judiciously, considering the specific characteristics of the data and the learning algorithms.

Secondly, this discrepancy between oversampling and non-oversampling results highlights the importance of evaluating machine learning models in a comprehensive and context-specific manner. Accuracy alone may not be the sole performance metric to rely on. It's imperative to consider other evaluation metrics such as precision, recall, F1-score, or area under the ROC curve, as well as conducting a thorough analysis of false positives and false negatives. In some cases, even if accuracy declines with oversampling, the model's ability to correctly classify the minority class, as reflected in precision and recall, might significantly improve. Therefore, the decision to employ oversampling should be guided by a more nuanced understanding of the dataset and the specific objectives of the machine learning task, rather than a simple reliance on accuracy as the ultimate performance measure.

To best of our knowledge, this is the first study comparing different models to find best method to predict mortality among patients with acute pulmonary embolism. We used different machine learning models to find which model could predict an adverse outcome according to data gathered from real population. In current predictive models, used in emergency rooms, because doctors have limited time and using many items for a scoring system can be complex and sophisticated, it interferes with on-time decision making, and could affect medical care adversely. Hence, only limited items are used for outcome prediction, so that physicians could use it simply and correctly. By development of machine learning based on artificial intelligence, we can use all items, affecting patients' health, without any mistake and time wasting. In this situation we can have good control on all aspects of patient's health, the main disease and comorbidities could be aggravated during hospital course. On the other hand, medical data has its own characteristics which could be totally different from data gathered in physics and mathematics. So, it is important to find best models in each situation.

In this study, by using different machine learning models, we used four laboratory tests, eleven physiological time series vital indicators, past medical history, and treatment data; including heart rate, blood pressure, oxygen saturation, and a history of heart failure, chronic obstructive pulmonary embolism, syncope, right ventricular enlargement, and S1T3Q3 in the patient's electrocardiogram, taking fibrinolytic therapy, admission days and embolectomy treatment, as well as 2 general descriptors (age and gender), totally 17 items to predict in-hospital mortality.

Artificial Neural Networks, Support Vector Machines, Decision Trees, Random Forests, and Adaptive Boosting were five machine learning methods that we used in this study to predict the mortality of pulmonary embolism. Our results showed that Random Forests model was the best model in predicting mortality of patient with acute pulmonary embolism. Rigatti showed that the Random Forest technique as a regression tree technique could achieve an excellent predictive accuracy. By using colon cancer data from 66,807 patients, a Cox model and a random forest models were developed and were compared to find their predictive power. This study showed that both models performance was the same [48]. Li et al. also used a random forest model. By using a large colorectal cancer datasets from the US and China, this study showed that proposed random forest model had excellent prediction power in discrimination and calibration of patients using multicenter clinical data [49]. Velazquez et al. used nine clinical features including demographic data, information about brain volume, and cognitive testing. They also used oversampling to balance the initially imbalanced classes then started training the model with 1000 estimators. They showed that a random forest model had a 93.6% accuracy in predicting progression of early mild cognitive impairment to Alzheimer's Disease [50]. Mbonyinshuti et al. used a random forest model to predict the pattern of medicine consumption for treating non-communicable diseases management by using historical consumption data. This study showed that this model had seventy-eight percent accuracy rate for the training set and a 71% accuracy rate for the testing set [51].

The proposed machine learning approach can effectively predict the mortality of patients with acute pulmonary embolism. The authors achieved promising results with Random Forests model and demonstrated the importance of oversampling techniques in handling imbalanced data. The research findings have practical implications in the medical field and can assist clinicians in making informed decisions about the treatment of patients with acute pulmonary embolism.

7. Conclusion

In conclusion, this study has advanced our understanding of predicting early mortality in pulmonary embolism patients through the application of oversampling techniques and diverse machine learning models. The findings underscore the significance of tailoring oversampling methods to the specific characteristics of each algorithm and dataset. Importantly, this research highlights the potential of machine learning in enhancing the accuracy of mortality predictions for patients with acute pulmonary embolism, providing valuable insights for clinical decision-making. Among the models explored, the Random Forests model emerged as a standout performer, demonstrating its ability to effectively predict mortality by achieving high accuracy, precision, and recall rates. This success not only emphasizes the utility of machine learning techniques in prognostication but also holds promise for refining existing predictive models used in emergency medical settings. By leveraging the power of machine learning, healthcare professionals can harness a broader range of data to make more informed and accurate decisions regarding the management of patients with acute pulmonary embolism. Looking ahead, future research endeavours could delve into the exploration of alternative machine learning algorithms and novel oversampling techniques, with the aim of further enhancing the precision of mortality predictions in this patient population. Moreover, the insights garnered from this study underscore the potential for machine learning to revolutionize healthcare decision support systems, paving the way for more personalized and effective patient care strategies.

Declaration of competing interest

We, the authors of the paper entitled "Advancing Prognostic Precision in Pulmonary Embolism: A Clinical and Laboratory-Based Artificial Intelligence Approach for Enhanced Early Mortality Risk Stratification," hereby declare our interests in relation to the subject matter presented in the manuscript.

References

[1] M. Ghoreishi, C. Pasrija, Z. Kon, Venoarterial extracorporeal membrane oxygenation for massive pulmonary embolism: when is the time to wean? Ann. Thorac. Surg. 114 (1) (2022) 355.

[2] L. Ryan, et al., Predicting pulmonary embolism among hospitalized patients with machine learning algorithms, Pulm. Circ. 12 (1) (2022), e12013.

[3] U.M. Nagamalesh, V.S. Prakash, K.C.K. Naidu, S. Sarthak, A.V. Hegde, T. Abhinay, Acute pulmonary thromboembolism: epidemiology, predictors, and long-term outcome–A single center experience, Indian Heart J. 69 (2) (2017) 160–164.

[4] S. Lee, S. Hong, W.C. Cha, K. Kim, Predicting adverse outcomes for febrile patients in the emergency department using sparse laboratory data: development of a time adaptive model, JMIR Med. informatics 8 (3) (2020), e16117.

[5] S. Hong, S. Lee, J. Lee, W.C. Cha, K. Kim, Prediction of cardiac arrest in the emergency department based on machine learning and sequential characteristics: model development and retrospective clinical validation study, JMIR Med. informatics 8 (8) (2020), e15932.

[6] J. Kwon, Y. Lee, Y. Lee, S. Lee, H. Park, J. Park, Validation of deep-learning-based triage and acuity score using a large national dataset, PLoS One 13 (10) (2018), e0205836.

[7] O.C. Redfern, et al., Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: development and validation of a multivariable model, Resuscitation 133 (2018) 75–81.

[8] M. Fernandes, S.M. Vieira, F. Leite, C. Palos, S. Finkelstein, J.M.C. Sousa, Clinical decision support systems for triage in the emergency department using intelligent systems: a review, Artif. Intell. Med. 102 (2020), 101762.

[9] M. Ghassemi, L.A. Celi, D.J. Stone, State of the art review: the data revolution in critical care, Crit. Care 19 (1) (2015) 1–9.

[10] N. Gilboy, P. Tanabe, D.A. Travers, A.M. Rosenau, D.R. Eitel, Emergency Severity Index, Version 4: Implementation Handbook 2012 Edition, Rockv. Agency Healthc. Res. Qual., 2011.

[11] W.A. Knaus, Apache 1978-2001: the development of a quality assurance system based on prognosis: milestones and personal reflections, Arch. Surg. 137 (1) (2002) 37–41.

[12] S. Lemeshow, D. Teres, J. Klar, J.S. Avrunin, S.H. Gehlbach, J. Rapoport, Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients, JAMA 270 (20) (1993) 2478–2486.

[13] T. Toma, A. Abu-Hanna, R.-J. Bosman, Discovery and inclusion of SOFA score episodes in mortality prediction, J. Biomed. Inf. 40 (6) (2007) 649–660.

[14] S. Levin, et al., Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index, Ann. Emerg. Med. 71 (5) (2018) 565–574.

[15] M.T. Keegan, O. Gajic, B. Afessa, Severity of illness scoring systems in the intensive care unit, Crit. Care Med. 39 (1) (2011) 163–169.

[16] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L.A. Celi, R. Mark, Mimic-iv, 2020, https://doi.org/10.13026/a3wn-hq05. *PhysioNet*, version 0.4.

[17] S.L. Hyland, et al., Early prediction of circulatory failure in the intensive care unit using machine learning, Nat. Med. 26 (3) (2020) 364–373.

[18] I. Silva, G. Moody, D.J. Scott, L.A. Celi, R.G. Mark, Predicting in-hospital mortality of icu patients: the physionet/computing in cardiology challenge 2012, in: 2012 Computing in Cardiology, 2012, pp. 245–248.

[19] M. Lee, et al., WiDS (Women in Data Science) Datathon 2020: ICU Mortality Prediction, PhysioNet, 2020, version 1.0. 0.

[20] A. Artetxe, A. Beristain, M. Grana, Predictive models for hospital readmission risk: a systematic review of methods, Comput. Methods Progr. Biomed. 164 (2018) 49–64.

[21] A. Artetxe, M. Graña, A. Beristain, S. Ríos, Balanced training of a hybrid ensemble method for imbalanced datasets: a case of emergency department readmission prediction, Neural Comput. Appl. 32 (10) (2020) 5735–5744.

[22] P.-Y. Tseng, et al., Prediction of the development of acute kidney injury following cardiac surgery by machine learning, Crit. Care 24 (1) (2020) 1–13.

[23] Y. Raita, T. Goto, M.K. Faridi, D.F.M. Brown, C.A. Camargo, K. Hasegawa, Emergency department triage prediction of clinical outcomes using machine learning models, Crit. Care 23 (1) (2019) 1–13.

[24] M. Klug, et al., A gradient boosting machine learning model for predicting early mortality in the emergency department triage: devising a nine-point triage score, J. Gen. Intern. Med. 35 (1) (2020) 220–227.

[25] M. Torabi, A. Mirafzal, A. Rastegari, N. Sadeghkhani, Association of triage time shock index, modified shock index, and age shock index with mortality in emergency severity index level 2 patients, Am. J. Emerg. Med. 34 (1) (2016) 63–68.

[26] C.K. Lee, I. Hofer, E. Gabel, P. Baldi, M. Cannesson, Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality, Anesthesiology 129 (4) (2018) 649–662.

[27] I.S. Hofer, C. Lee, E. Gabel, P. Baldi, M. Cannesson, Development and validation of a deep neural network model to predict postoperative mortality, acute kidney injury, and reintubation using a single feature set, NPJ Digit. Med. 3 (1) (2020) 1–10.

[28] A.U. Zaidi, K.K. Hutchins, M. Rajpurkar, Pulmonary embolism in children, Front. Pediatr. 5 (2017) 170.

[29] S.F. Thieme, et al., Severity assessment of pulmonary embolism using dual energy CT–correlation of a pulmonary perfusion defect score with clinical and morphological parameters of blood oxygenation and right ventricular failure, Eur. Radiol. 22 (2012) 269–278.

[30] G. Cammarota, et al., Gut microbiome, big data and machine learning to promote precision medicine for cancer, Nat. Rev. Gastroenterol. Hepatol. 17 (10) (2020) 635–648.

[31] S. Zhao, G. Tang, P. Liu, Q. Wang, G. Li, Z. Ding, Improving mortality risk prediction with routine clinical data: a practical machine learning model based on eICU patients, Int. J. Gen. Med. (2023) 3151–3161.

[32] G. Halasz, et al., A machine learning approach for mortality prediction in COVID-19 pneumonia: development and evaluation of the Piacenza score, J. Med. Internet Res. 23 (5) (2021), e29058.

[33] R.A. Taylor, et al., Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach, Acad. Emerg. Med. 23 (3) (2016) 269–278.

[34] X. Li, H. Liu, J. Yang, G. Xie, M. Xu, Y. Yang, Using machine learning models to predict in-hospital mortality for ST-elevation myocardial infarction patients, in: MEDINFO 2017: Precision Healthcare through Informatics, IOS Press, 2017, pp. 476–480.

[35] M.A. Bin Hameed, Z. Alamgir, Improving mortality prediction in Acute Pancreatitis by machine learning and data augmentation, Comput. Biol. Med. 150 (2022), 106077.

[36] B. Nistal-Nuño, Machine learning applied to a cardiac surgery recovery unit and to a coronary care unit for mortality prediction, J. Clin. Monit. Comput. 36 (3) (2022) 751–763.

[37] A.E.W. Johnson, M.M. Ghassemi, S. Nemati, K.E. Niehaus, D.A. Clifton, G. D. Clifford, Machine learning and decision support in critical care, Proc. IEEE 104 (2) (2016) 444–466.

[38] F. Xie, et al., Deep learning for temporal data representation in electronic health records: a systematic review of challenges and methodologies, J. Biomed. Inf. 126 (2022), 103980.

[39] X. Gao, et al., An ensemble imbalanced classification method based on model dynamic selection driven by data partition hybrid sampling, Expert Syst. Appl. 160 (2020), 113660.

[40] K. Maharana, S. Mondal, B. Nemade, A review: data pre-processing and data augmentation techniques, Glob. Transitions Proc. (2022).

[41] R. Verma, S. Maheshwari, A. Shukla, Feature engineering combined with 1-D convolutional neural network for improved mortality prediction, Bio. Algorithm Med. Syst. 16 (4) (2021).

[42] S.-A. Sadegh-Zadeh, M. Bahrami, A. Najafi, M. Asgari-Ahi, R. Campion, A. M. Hajiyavand, Evaluation of COVID-19 pandemic on components of social and mental health using machine learning, analysing United States data in 2020, Front. Psychiatr. 13 (2022).

[43] I. Ibrahim, A. Abdulazeez, The role of machine learning algorithms for diagnosing diseases, J. Appl. Sci. Technol. Trends 2 (2021) 10–19, 01.

[44] S.-A. Sadegh-Zadeh, A. Rahmani Qeranqayeh, E. Benkhalifa, D. Dyke, L. Taylor, M. Bagheri, Dental caries risk assessment in children 5 Years old and under via machine learning, Dent. J. 10 (9) (2022) 164.

[45] S.-A. Sadegh-Zadeh, et al., Machine learning modelling for compressive strength prediction of superplasticizer-based concrete, Infrastructures 8 (2) (2023) 21.

[46] Z. Jourahmad, et al., Machine Learning Techniques for Predicting the Short-Term Outcome of Resective Surgery in Lesional-Drug Resistance Epilepsy, 2023 *arXiv Prepr. arXiv2302.10901*.

[47] S.-A. Sadegh-Zadeh, et al., An approach toward artificial intelligence alzheimer's disease diagnosis using brain signals, Diagnostics 13 (3) (2023) 477.

[48] S.J. Rigatti, Random forest, J. Insur. Med. 47 (1) (2017) 31–39.

[49] J. Li, et al., A multicenter random forest model for effective prognosis prediction in collaborative clinical research network, Artif. Intell. Med. 103 (2020), 101814.

[50] M. Velazquez, Y. Lee, A.D.N. Initiative, Random forest model for feature-based Alzheimer's disease conversion prediction from early mild cognitive impairment subjects, PLoS One 16 (4) (2021), e0244773.

[51] F. Mbonyinshuti, J. Nkurunziza, J. Niyobuhungiro, E. Kayitare, Application of random forest model to predict the demand of essential medicines for non-communicable diseases management in public health facilities, Pan Afr. Med. J. 42 (2022).