Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2021;8(1): 8-23.

Original Article

Using Stacking methods based Genetic Algorithm to predict the time between symptom onset and hospital arrival in stroke patients and its related factors

Firouz Amani[1], Jafar Abdollahi[2*], Alireza Mohammadnia [3], Paniz Amani[4], Ghasem Fattahzadeh-Ardalani[5]

[1]Department of Biostatistics and Community Medicine, School of Medicine, Ardabil University of Medical Science, Ardabil, Iran.
[2]Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran.
[3]Department of Basic Science, School of Medicine, Ardabil University of Medical Science, Ardabil, Iran.
[4]Student in Electronic Engineering, Tabriz University, Tabriz, Iran.
[5]Department of Neurology, School of Medicine, Ardabil University of Medical Science, Ardabil, Iran.

ARTICLE INFO

Received 11.06.2021
Revised 17.08.2021
Accepted 29.11.2021
Published 15.03.2022

Key words:
Stroke;
Machine learning;
Classification;
Random Forest;
Hospital.

ABSTRACT

Introduction: The early arrival of patients with acute ischemic stroke to start of treatment by recombinant tissue plasminogen activator (rt-PA) within 4.5 hours after onset of stroke and its modeling by data mining methods is an important issue in care of stroke patients. In this paper, the aim was to provide methods to predict the time between symptom onset and hospital arrival in stroke patients and related factors, in addition to improve classification in minority class data, also to maintain the ability of classifying majority class data at an acceptable level.
Methods: We included 676 patients with ischemic stroke who referred to hospital of Ardabil city in the northwest of Iran in 2018. A new method using a combination of machine learning algorithms and genetic algorithms has been proposed to solve this problem. The performances were evaluated with accuracy, sensitivity, specificity, positive predictive value, and negative predictive value.
Results: In this study, the stacking technique provides a better result (accuracy 99.51%, sensitivity 100%, and specificity 99.40%) among all other techniques.
Conclusion: Results of this study showed that this model can be used as a valuable tool for clinical decision making.

Introduction

Early initiation of treatment in patients with acute ischemic stroke with recombinant tissue plasminogen activator (rt-PA) within 4.5 hours after stroke would be important. Because intravenous tissue plasminogen activator (tPA) is a time-sensitive treatment with proven benefits in stroke patients. Early treatment after the onset of stroke symptoms is more effective in preventing disability and stroke mortality.[1, 2] Machine learning plays a key role in medical

*.Corresponding Author: J.abdollahi77@gmail.com

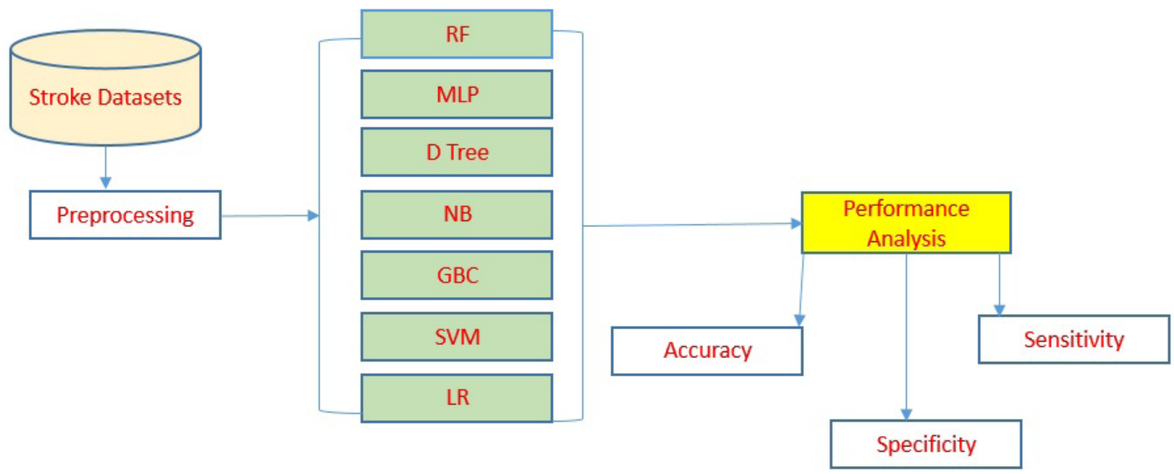

Figure .1 The proposed model structure

decision making and has special skills in integrating multiple risk factors into one predictive tool.[3] The rt-PA has been used in many countries for the first 4.5 hours after the onset of symptoms, while only a small proportion of stroke cases could receive the drug on time.[4-6] Median time from onset of stroke symptoms to admission to the hospital in more studies ranged from 4 hour to more than 24 hours.[7-9]

Increasing awareness of patients and their family about the importance of early arrival to hospital after onset of stroke symptoms for early treatment can be very important. Early neurological attention to stroke cases could be associated with better functional outcome of patients and less hospitalization time. Several studies reported that, certain factors such as age, gender, having underlying diseases, also transport options had main role in late arrival of patients with stroke to the hospitals.[10-12]

Analyzing the factors associated with delay in arrival of stroke patients to hospital are more important for health system and also health system providers. The aim of this study was to develop a machine learning method to predict the time between symptom onset and hospital arrival in stroke patients and its related factors.

Materials and Methods

Stacking is an ensemble learning technique that uses predictions from multiple models (for example decision tree, KNN or SVM) to build a new model. This model is used for making predictions on the test set. Figure 1 shows a flow chart of the algorithms we applied in this study. The stacked teaching methodology consists of two consecutive stages:

- Training stage
- Test stage

In the training method of the stacked method, 6 prediction models are generated using training samples. Within the next step, to predict the category or position of the experimental sample, the stacked method calculates the output of every prediction model and aggregates it with another one.

We know the collective model will perform better than the single model when each of its constituent models is independent of the others but in this paper, we used different learning algorithms for different learning models to create independent learners.

For a stack model to perform better than other models, each of its components must be independent of each other. In addition, its accuracy should be more than 0.5. Therefore, to extend the accuracy of the proposed model and also the effective performance of group algorithms and particularly to scale back its error rate in diagnosing the disease, the additional tree classifier algorithm has been used as a meta-algorithm for the primary time.

Data collection method and dataset

Data were collected from stroke registry project in Alavi hospital of Ardabil. Database included the information of 676 patients with ischemic stroke who referred to Ardabil central hospital in northwest of Iran in 2018 (population: 1,320,000). Patients under study were in range of 25-98 years old. All patients with acute stroke in the form of neurologic focal symptoms that longed more than 24 hours and existence changes in Brain CT which confirmed by neurologist, were entered in the study. Verbal informed consent was obtained from all subjects and relatives.

Data collected by a checklist including demographic and clinical data of patients such as age, sex, residence place, marital status, history of DM, HTN, hyperlipidemia, heart disease, smoking consumption, type of risk factor, mode of transportation to the hospital, time of arrival to the hospital emergency and time of onset diagnostic and therapeutically actions.

The detailed clarification about this dataset is given in Table 1.

Machine learning strategies

The aim of the present study was to detect related factors on arrival time of patients with stroke by using seven classification machine learning algorithms such as DT, NB, ANN, SVM, RF, GB and LR.

- **Decision Tree (DT):** A decision tree is a map of the possible results of a series of related choices or options so that it allows an individual or organization to weigh possible actions in terms of costs, opportunities, and benefits.[13]
- **Naive Bayes (NB)** is a group of simple classifiers based on probabilities created assuming the independence of random variables and based on Bayes theorem.[14]
- **Artificial Neural Network (ANN):** The artificial neural network creates a structure similar to the biological structure of the human brain and neural network to be able to learn to generalize and make the decision.[15,16]
- **Support Vector Machine (SVM)** is classified as a pattern recognition algorithm. The SVM algorithm can be used wherever there is a need to identify patterns or classify objects in specific classes.[16]
- **Random Forest (RF)** is a combined learning method for regression classification, which works on the training time and class output (classification) or for predictions of each tree separately, based on a structure consisting of

Table 1. Description of the stroke datasets.

Dataset	Sample Size	Feature size including class label	Classes	Presence of missing attribute	Presence of noisy attributes
Stroke	676	14	2	NO	NO

a large number of decision trees.[16, 17]

- **Gradient Boosting (GB)** is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees.
- **Logistic Regression (LR)** is a machine learning technique for regression and classification problems which to assign observations to a discrete set of classes.[16]

In this paper, the stochastic gradient boosting algorithm is used to its successful performance in most scenarios and studies, and the Multilayer Perceptron algorithm is used as the neural network.

Recursive Feature Elimination (RFE) for Feature Selection

Feature selection refers to techniques that it selects a subset of the foremost relevant features (columns) for a dataset. Fewer features can allow machine learning algorithms to run more efficiently (less space or time complexity) and be simpler. Some machine learning algorithms are often misled by irrelevant input features, leading to worse predictive performance.

Recursive Feature Elimination (RFE), maybe a popular feature selection algorithm. RFE is popular because it is easy to configure and use. Also in a large training data set, RFE is very useful in selecting those attributes (columns) that have the greatest impact on predicting the target variable.

Genetic Algorithm for Feature selection

In machine learning, the Feature Selection (FS) step seeks to pinpoint the foremost informative variables from data to make robust classification models. This becomes crucial within the Omics data era because the combination of high-dimensional data with information from various sources (clinical and environmental) enables researchers to check complex diseases like cancer or disorder thoroughly. Given the quantity and class of knowledge, accurate prediction, for instance, of the character of the disease and/or the result of patients is difficult, but the planning of high-performance classification models through the appliance of machine learning is strongly required.[18]

Feature selection could be a crucial step in machine learning analysis. Currently, many feature selection approaches don't ensure satisfying results, in terms of accuracy and computational time, when the number of information is large, like in 'Omics' datasets.

Here, we propose an innovative implementation of a genetic algorithm, for fast and accurate identification of informative features in multi-class and high-dimensional datasets.[18]

There are two important configuration options when using RFE: the choice in the number of features to select and the choice of the algorithm used to help choose features. Both of these hyper parameters can be explored, although the performance of the method is not strongly dependent on these hyper parameters being configured well.

Data preprocessing

Data preprocessing step for all the algorithms used consists of three steps:

Phase 1 - Data Exploration

To understand a number of the potential features and to think about whether data cleaning is required.

Phase 2 - Data Clearance

- Several factors must be considered within

Table 2. Parameters used in algorithms

Algorithms	Best Hyper Parameters
RF	n_estimators=100, max_features=outo, criterion=gini, bootstrap=true
ANN(MLP)	hidden_layer_sizes=(100,), activation==’relu’, solver=’sgd’, learning_rate=’constant’, learning_rate_init=0.001
GB(Stochastic)	max_depth= 4, 'n_estimators': 200, 'max_depth': 4, 'min_samples_split': 5, 'learning_rate': 0.01, 'loss': 'ls'
DT	Criterion=’gini’, min_impurity_split=1e-07, min_samples_leaf=1, splitter=’best’
SVM	kernel=’rbf’
LR	Penalty=’l1’, solver=’lbfgs’

the data cleansing process.

- Duplicate or irrelevant observations.
- Improper labeling of information, a batch occurs several times.
- Data points are missing or empty.
- Unexpected throws.

Phase 3 - Modeling with machine learning algorithms

Data preprocessing are necessary to prepare the stroke data in a manner that a machine learning model can accept. Separating the training and testing datasets ensure that the model learns only from the training data and tests its performance with the testing data. The dataset was divided into training and test data. The training data contain 70% of the total dataset, and the test and validation data each contain 15%. At first, all of the data were shuffled.[15, 16] All of necessary parameters for ML algorithms are listed in Table 2.

Variable selection

Features Selection for classification model attempts to pick minimally sized subset in keeping with following criteria: (1) The classification accuracy should need to increase; (2) The values for the chosen features should close as possible to the initial class distribution. All features are listed in the Table 3.

Model development and validation

For the study, Anaconda Platform and Jupyter notebook was used for implementation and Python 3.7 programming language was used for coding. Figure 2 shows a flow chart of the algorithms we applied in this study.

How to deal with unbalanced data has always been a challenging issue in data mining, so that the number of errors in identifying smaller category data is always fraught with many errors. This leads to a very poor prediction of minority class examples, because minority class education is not done properly. Therefore, one of the important issues in the field of data mining is the issue of classifying unbalanced data sets. In this paper, the aim was to provide methods for classification that, in addition to improving classification in minority class data, also maintain the ability to classify majority class data at an acceptable level. Since in previous works, artificial data has generally been used for categorization work, so in present article, real data related to stroke have been collected. A new method using a combination of machine learning algorithms and genetic algorithms has been proposed to solve this problem.

Our review indicated that application of a ML algorithm alone in detection and prediction of diseases has not been precise and successful.[14] Next, we used 7 standardized and widely used algorithms to classify Stroke on our samples for comparison. In order to boost algorithms

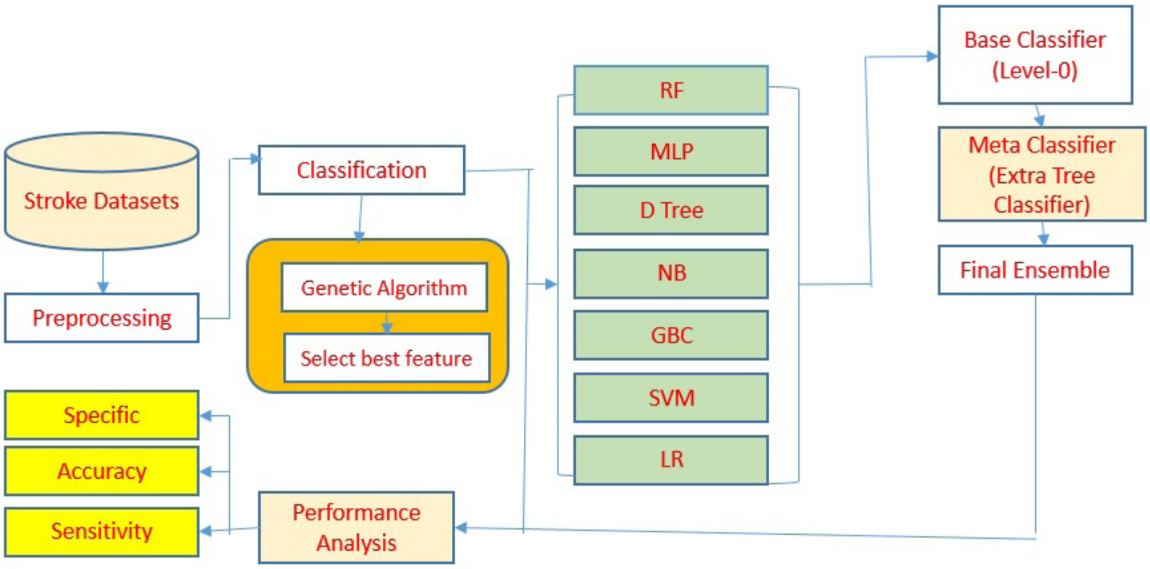

Figure 2. Proposed of research

Table 3. Features of stroke type dataset

Feature	Class	Type
Gender	2 class [0 , 1]	Integer
Age	25-98	Integer
Residence Place	2 class [0 , 1]	Integer
Marital Status	2 class [0 , 1]	Integer
HTN	2 class [0 , 1]	Integer
DM	2 class [0 , 1]	Integer
HLP	2 class [0 , 1]	Integer
Alcohol	2 class [0 , 1]	Integer
Ciggaret	2 class [0 , 1]	Integer
CVA	2 class [0 , 1]	Integer
Valvular	2 class [0 , 1]	Integer
AF	2 class [0 , 1]	Integer
Coronary	2 class [0 , 1]	Integer
Type of Used drugs	2 class [1, 2, 3]	Integer
C (Target)= Arrival time	2 class [1,2]	Integer

used to this paper, we also applied genetic algorithms. We adopted two different strategies called Cross-Validation and Holdout to our comparison studies.

Model assessment

The confusion matrix has been used to determine the relationship between the actual values and predicted values. Table 4 shows the structure of confusion matrix. We compared the classification performance using Accuracy, Sensitivity, Specificity, and Area under Curve (AUC), Positive predictive value (PPV), Negative predictive value (NPV), F1 score, and Youden Index. FP and FN mean the number of false-positive or false-negative samples.[13-17]

Table 4. Structure of confusion matrix

		Actual class	
		Negative	Positives
Predicted class	Negative	TP	FP
	Positives	FN	TN

TP and TN represent the amount of true positive or true negative samples. Specificity measures the ratio of negatives that are correctly discriminated against. Sensitivity measures the ratio of positives that are correctly discriminated against. NPV was accustomed evaluate the algorithm for screening. PPV was the probability of getting a disease when the diagnostic index is positive. AUC is an index to live the performance of the classifier. F1 score was a measure of the accuracy of a binary model. Additionally, the performance was evaluated with F-measure (F1) to check the similarity and variety of performance. Following quality parameters were wont to evaluate the results.[13-17]

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{\mathrm{TN}}{\mathrm{TN+FP}} \quad (3)$$

Results

Patient's characteristics

Of all patients, 382 (56.5%) were female and the mean age of patients was 69.3±13.2 (25-98). The main stroke risk factors were as follows: high blood pressure (69.2%), heart disease (43.8%), diabetes (31%), smoking (10.8%), alcohol (0.4%) and CVA (29%). It was determined that 171 patients (25.3%) arrived at the hospitals within 4.5 hours and 505 patients (74.7%) arrived at the hospitals after 4.5 hours from the onset of stoke symptoms (Table 5).

Recursive Feature Elimination (RFE) for Feature Selection

RFE can be a wrapper-type feature selection algorithm which can be used to features selection by using machine learning algorithms. This can be in contrast to filter-based feature selections that score each feature and choose those features with the most important (or smallest) score.

Performance of machine learning algorithms

In this ML model, we predicted the whole dataset using 10-fold cross-validation and evaluated the performance on stroke arrival time dataset by AUC, accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. Table 5 shows the performance of the predictive model using different data mining algorithm techniques. As shown in Table 5, the suggested method (Stacking) with 100% and DT with 97% had the highest sensitivity than other models by 10-fold cross-validation technique.

Table 5. socio-Demographic characteristics of the stroke patients (n=676)

Variables	Groups	n	%
Age, years (69.3±13.2, Range: 25-98)			
	<40	14	2.1
	40-60	155	22.6
	>60	507	75
Gender			
	Female	294	43.5
	Male	382	56.5
Marital Status			
	Single	16	2.4
	Married	660	97.6
Place of Residence			
	Rural	195	28.8
	Urban	481	71.2
Smoking status			
	Yes	73	10.8
	No	603	89.2
Alcohol consumption			
	Yes	3	0.4
	No	673	99.6
Arrival time to hospital			
	Early (≤4.5)	171	25.3
	Late (>4.5)	505	74.7

As shown in Table (5), the outcome of comparing various algorithms using Holdout approach showed that the algorithm (Stacking) with 99% and DT with 98% had the highest accuracy rate than others in the diagnosing and predicting of arrival time in stroke patients.

Technically, RFE could be a wrapper-style feature selection algorithm that also uses filter-based feature selection internally. All of necessary Features in all various machine learning algorithms based on their importance are listed in Figure 3 to Figure 9.

In the ensemble feature selection method, if the burden adjustment is performed on each feature subset used, the ensemble effect is significantly different; therefore, the way to find the optimized weight vector may be a key and challenging problem. Aiming at this optimization problem, this paper proposes an ensemble feature selection approach supported by a genetic algorithm. The employment of a genetic algorithm for a feature selection is presented, reducing the first size of the dataset used, a crucial aspect when working with the limited resources of a mobile device. For the evaluation of this process, seven different classification methods are applied. Finally, a comparison of the models obtained supported the accuracy, is performed, so as to spot the classification method that presents the simplest performance within the development of a model. Table 6 examines the various methods of feature selection with genetic algorithms supported by machine learning algorithms.

Interestingly, comparison of these algorithms via Holdout strategy reproduced almost the same data as cross validation did, indicating the Stacking algorithm carries the highest accuracy amongst all algorithms used in our study (Table 7).

According to the previous researches, the success of group learning techniques depends on the variety of different classifiers, the purpose of using the decision tree algorithm

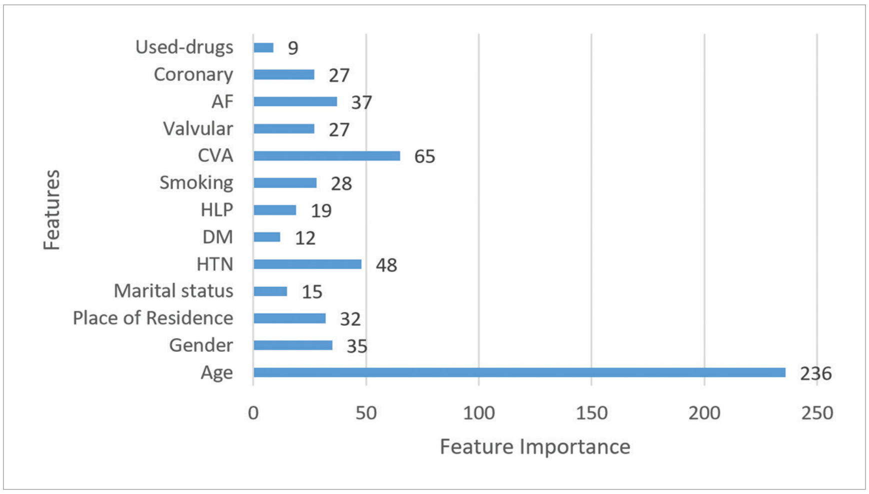

Figure 3. Feature importance based XGB Classifier

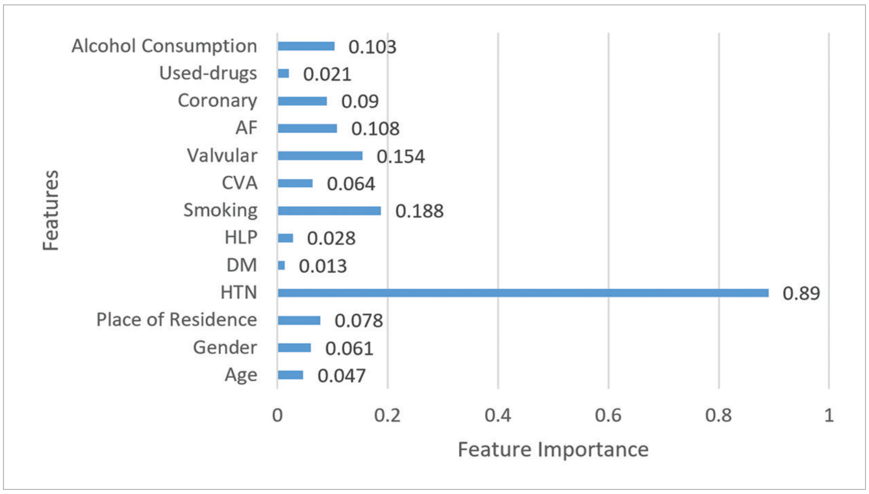

Figure 4. Feature selection based on Extra Trees Classifier and RFECV

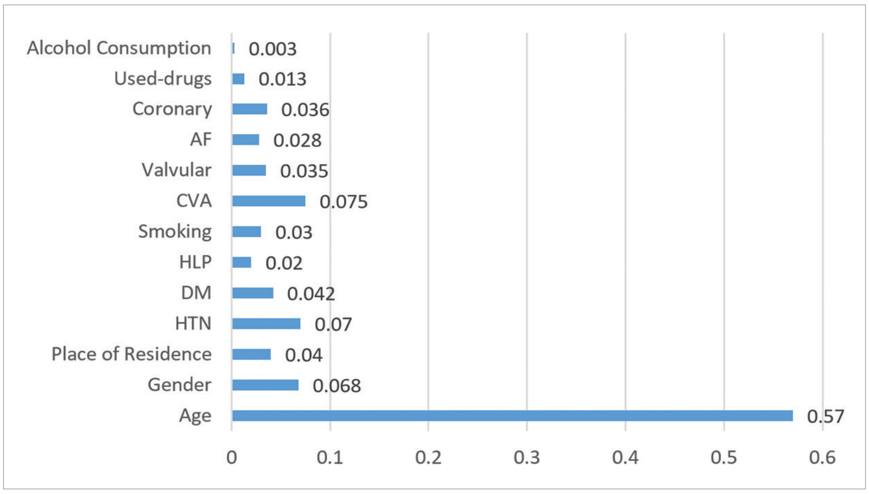

Figure 5. Feature selection based on Random Forest Classifier and RFECV

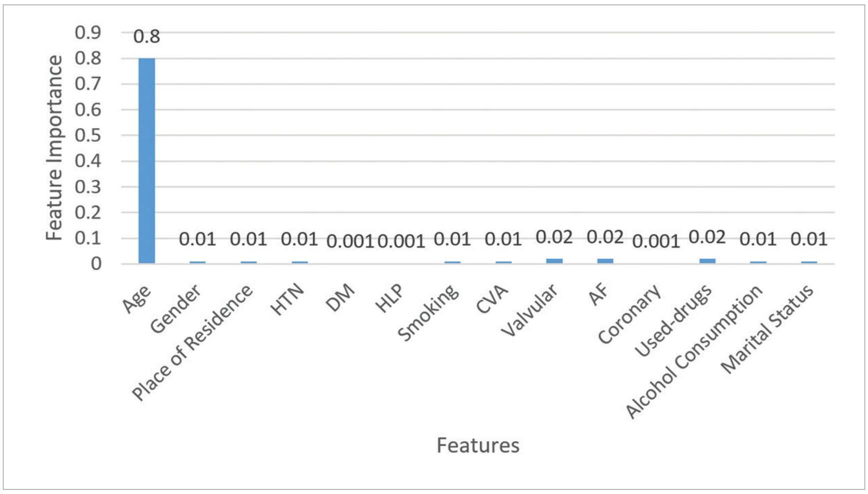

Figure 6. Feature selection based on AdaBoost Classifier and RFECV

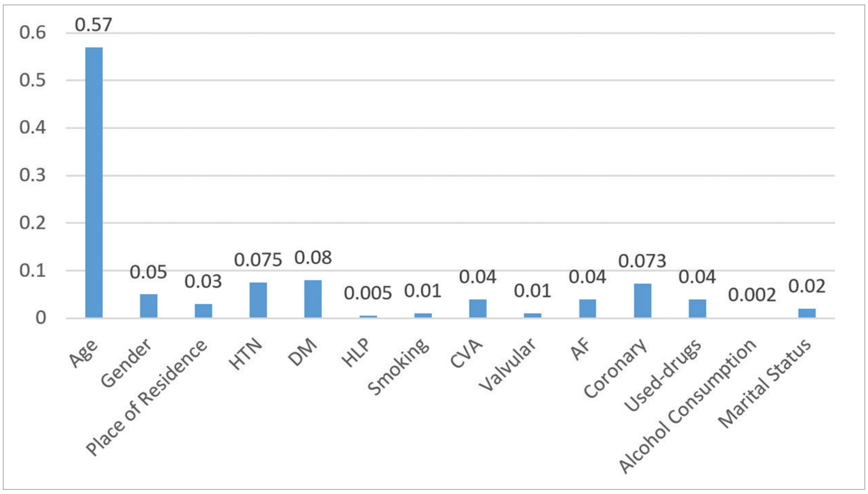

Figure 7. Feature selection based on Decision Tree Classifier and RFECV

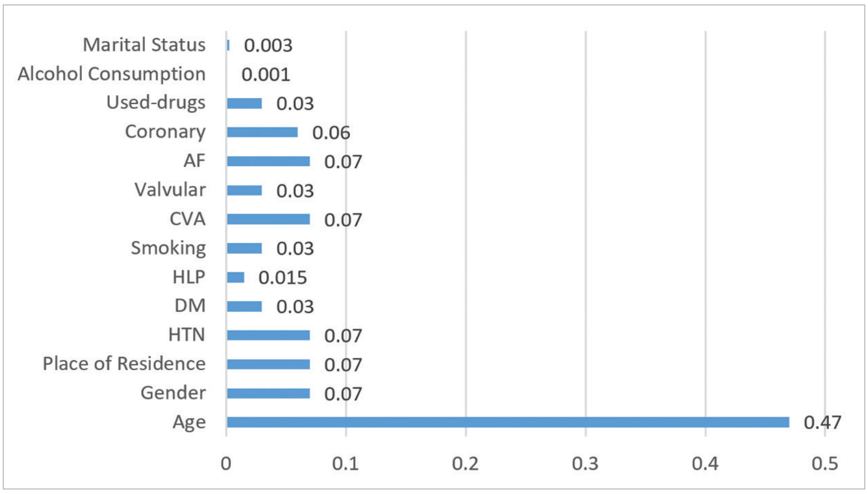

Figure 8. Feature selection based on Gradient Boosting Classifier and RFECV

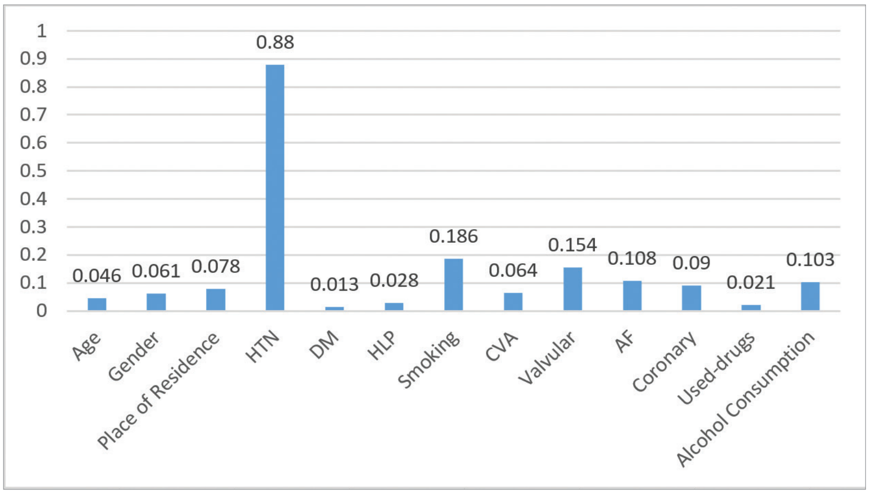

Figure 9. Feature selection based on Logistic Regression and RFECV

Table 6. Genetic Based feature selection

	RF	NB	GB	ANN (MLP)	DT	SVM	ETC	LR
	Age	Age	Age	Age	Age	DM	Gender	Age
	Gender	Gender	DM	Gender	marital	Alcohol	Place of Residence	Place of Residence
	marital	HTN	Alcohol	Place of Resi-dence	Alcohol	Cigarette	Alcohol	DM
	DM	DM	Coronary	marital	Cigarette	Valvular	Cigarette	HLP
	HLP	HLP	USED-drug	HTN	AF		Valvular	Cigarette
Feature	Alcohol	Alcohol		DM	USED-drug		AF	CVA
	Cigarette	AF		CVA			Coronary	Valvular
	CVA	Coronary		Valvular			USED-drug	AF
	AF	USED-drug		AF				Coronary
				Coronary				

Table 7. Comparison performance of different machine learning algorithms on detection effective factors on arrival time of patients with stroke to hospital with Holdout approach

	RF	NB	ANN	ANN	DT	SVM	LR	Stacking
Accuracy	98	72	73	73	98	73	74	99
Sensitivity	94	61	70	70	97	72	72	100
Specific	96	70	71	71	94	64	66	99

is to achieve a level of confidence and lower P-value and to combine the proposed model with weak and strong categories. Because the use of amplification algorithms such as RF and GB will definitely reach the proposed model with high accuracy. And this may lead us to the conclusion whether the Stacking model works better by combining weak and strong categories, or should only strong and reinforced categories be used as the basic algorithms in this method (Stacking) (Table 8).

Diagnosis of detection effective factors on arrival time of patients with stroke to hospital was compared in rates among various algorithms with Holdout approach, 3,5 and 10-fold cross validation are listed in Table 9.

Table 8. Feature importance based XGB Classifier

Feature	Percentage of importance
Age	0.054036893
Gender	0.08409524
Place of Residence	0.09044497
Marital status	0.020976264
HTN	0.06232663
DM	0.01751493
HLP	0.086411856
Alcohol	0.0
Smoking	0.06443382
CVA	0.094390385
Valvular	0.112008244
AF	0.1454139
Coronary	0.08051951
USED-drugs	0.08742731

Table 9. Comparison of accuracy rates within models by Holdout, 3, 5 and 10 fold cross-validation

Models	K-fold=3	K-fold=5	K-fold=10	Holdout
RF	0.71	0.91	0.98	0.71
NB	0.69	0.74	0.72	0.57
ANN	0.68	0.73	0.73	0.69
GB	0.49	0.56	0.79	0.73
DT	0.49	0.51	0.98	0.60
SVM	0.73	0.74	0.73	0.73
LR	0.72	0.74	0.74	0.71
Stacking	0.97	0.98	0.99	0.97

RF, Random Forest; SVM, Support Vector Machine; ANN, Artificial Neural Network; LR, Logistic Regression; GB, Gradient Boosting Classifier; DT, Decision Tree Classifier; NB, Naïve Bayes; Accuracy, Sensitivity, Specificity, PPV, Positive Predictive Value; NPV, Negative Predictive Value; TPR, True Positive Rate; FPR, False Positive Rate

One of the highlights of this text is managing unbalanced data. To resolve the possible problem of over fit or under fit, a replacement method employing a combination of machine learning algorithms and genetic algorithms has been proposed to resolve this problem. In learning unbalanced data, they'll be over fit or under fit. We used the Stacked Generalization method, which additionally to increasing the classification accuracy, reduces the error rate. After combining the proposed model and evaluation on the test data set, the proposed model is in a position to spot test samples with high accuracy, the results of which are shown within the figure 10.

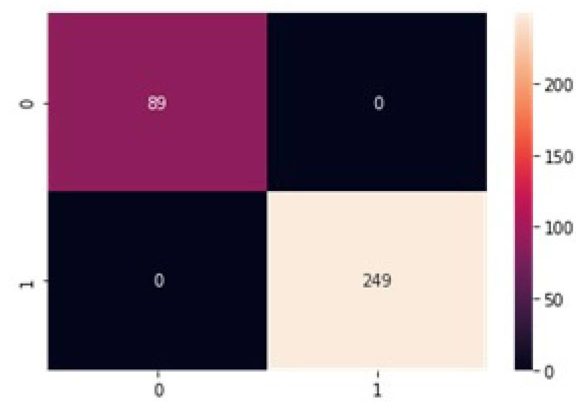

Figure 10. Confusion matrix

Discussion

In this study, we used various Machine Learning (ML) algorithms to improve prediction of effective factors on arrival time of patients with stroke to hospital after onset of symptoms that provided significant insights compared with traditional statistical models. Among ML models, Stacking and DT models showed higher performance than others. Similar to this study, Wu et al in a study entitled “Prediction of fatty liver disease using machine learning algorithms” showed that among four algorithms, the random forest model showed higher performance than other classification models which was in line with our study results.[19]

To our knowledge, this is the first study attempted to predict effective factors on arrival time of patients with stroke to hospital by using various Machine Learning algorithms. There are many kinds of machine learning algorithms which have been developed along with the most popular Bayesian algorithm and logistic regression, it is hard to make a proper algorithm for clinical decision making and clinical practices.[11]

Therefore, the performances of various algorithms are the foremost important consideration, together with the straightforward to use and therefore the interpretation of the models. However, our model could effectively detect effective factors on arrival time of patients with stroke to hospital for anyone by initial screening without using advanced methods. In addition, the model can provide an easy, fast, low cost, and noninvasive method to accurately detection of effective factors on arrival time of patients with stroke to hospital after onset of symptoms.[12]

As healthcare data is increasing every day, machine learning techniques allow huge volumes of data to be analyzed quickly.[20] Therefore, it is the opportunity to apply machine learning algorithms to the care of individual patients in medical practice. By using various machine learning prediction models, physicians can be able to extract the minimum data necessary to make a therapeutic decision.[21] Our model has the potential to early arrival time detection that will help to improve precise and appropriate treatment pattern.

It is very important for physicians and clinicians to know most of predictive variables which affects the time from onset of symptom to hospital admission in patients with stroke for have the best treatment outcome. Patient's baseline characteristics might be the strongest predictors of effective factors on arrival time of patients with stroke to hospital. Therefore, we carefully adopted a feature selection strategy and used 10-fold cross-validation to repeatedly screen potential variables. Data were included from a medical center of stroke without additional clinical assessments, and our high-performance prediction model can be easily integrated into these dataset to identify effective factors on arrival time of patients with stroke to hospital after onset of symptoms. Our model can help to identify stroke patients and may be had significantly impact on treatment pattern. Early prediction using this model might bring benefits from treatment reduction, and decreasing medical cost in future.

Limitation

It was better we run the preprocessing of data for each algorithm separately but unfortunately, we did not have this and it is one of our study limitations. Because we did preprocess all dataset in one time and prepared the preprocessed dataset to training and test steps in necessary algorithms. We know that stacking can deal with imbalance but it is due to chance by adding several base learners. In this study we exposure with imbalance dataset in using stacking method and we can solve this problem with a proper resampling process in one of the base learners without the need to go through the computational burden of stacking and this is other limitation of this study.

Also, in about the third limitation, we know that stacking method never does worse than its base learners but to make sure that it is better that predictions of its base learners must be as uncorrelated as possible unless going through too much computational burden would be for nothing. And the stacked model might come up the same as one of its base learners which could be achieved without doing the stacking at all and this problem can be solved by including the correlation table of base learner predictions but we do not this.

Conclusion

In recent times, researchers showed huge interest in Machine Learning approaches that try and develop information representations via computational modules.[22] Therefore, it becomes essential to assist healthcare professionals to diagnose correctly and facilitate treatment planning.[23] In this study, seven machine learning techniques were used to detection effective factors on arrival time of patients with stroke to hospital. All the algorithms worked with a reasonable accuracy and speed. However, the Stacking algorithm showed maximum precision and minimum errors, and showed better performance than other ML classification techniques. This prediction outcome has the potential to help clinicians make more precise and meaningful decisions about detection effective factors on arrival time of patients with stroke to hospital after onset of their symptoms. We need to pay more attention to the explanations given in the study limitations and apply them in the future similar studies.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

1. Saver JL, Fonarow GC, Smith EE, et al. Time to Treatment with Intravenous Tissue Plasminogen Activator and Outcome From Acute Ischemic Stroke. JAMA. 2013;309(23):2480–2488. doi:10.1001/jama.2013.6959.

2. Weintraub MI. Thrombolysis (tissue plasminogen activator) in stroke a medicollegal quagmire. Stroke. 2006; 37: 1917-22.

3. P. Groves, B. Kayyali, D. Knott, S.V. Kuiken, The big data revolution in healthcare: Accelerating value and innovation, 2016.

4. Marler JR, Tilley BC, Lu M, et al. Early stroke treatment associated with better outcome: the NINDS rt-PA study. Neurology. 2000; 55: 1649–55.

5. Stolz E, Hamann GF, Kaps M, et al. Regional differences in acute stroke admission and thrombolysis rates in the German federal State of Hesse. Dtsch Arztebl Int. 2011; 108(36): 607–9.

6. M Al Khathaami A, O.Mohammad Y, S.Alibrahim F and A.Jradi H. Factors associated with late arrival of acute stroke patients to emergency department in Saudi Arabia. SAGE Open Medicine. 2018;6:1-7.

7. Ashraf VV, Maneesh M, Praveenkumar R, Saifudheen K, Girija AS. Factors delaying hospital arrival of patients with acute stroke. Ann Indian Acad Neurol. 2015; 18(2): 162–6.

8. Chen CH, Huang P, Yang YH, Liu CK, Lin TJ, Lin RT. Pre-hospital and in-hospital delays after onset of acute ischemic stroke: a hospital-based study in southern Taiwan.

Kaohsiung J Med Sci. 2007; 23(11): 552-9.

9. Hong ES, Kim SH, Kim WY, Ahn R, Hong JS. Factors associated with prehospital delay in acute stroke. Emerg Med J. 2011; 28(9): 790–3.

10. Moser DK, Kimble LP, Alberts MJ, Alonzo A, Croft JB, Dracup K, Evenson KR, Go AS, Hand MM, Kothari RU, Mensah GA, Morris DL, Pancioli AM, Riegel B, Zerwic JJ: Reducing delay in seeking treatment by patients with acute coronary syndrome and stroke: A scientific statement from the American Heart Association Council on cardiovascular nursing and stroke council. Circulation. 2006; 114: 168-182.

11. Kleindorfer DO, Lindsell CJ, Broderick JP, Flaherty ML, Woo D, Ewing I, Schmit P, Moomaw C, Alwell K, Pancioli A, Jauch E, Khoury J, Miller R, Schnider A, Kissela BM: Community socioeconomic status and prehospital times in acute stroke and transient ischemic attack. Do poorer patients have longer delays from 911 call to the emergency department?. Stroke. 2006; 37: 1508-1513.

12. Rajajee V, Saver J: Prehospital care of the acute stroke patient. Tech Vasc Interv Radiol. 2005; 8: 74-80.

13. Abdollahi, J., Moghaddam, B. N., & Parvar, M. E. (2019). Improving diabetes diagnosis in smart health using genetic-based Ensemble learning algorithm. Approach to IoT Infrastructure. Future Gen Distrib Systems J, 1, 23-30.

14. Abdollahi, J., Keshandehghan, A., Gardaneh, M., Panahi, Y., & Gardaneh, M. (2020). Accurate detection of breast cancer metastasis using a hybrid model of artificial intelligence algorithm. Archives of Breast Cancer, 18-24.

15. Abdollahi, J., & Nouri-Moghaddam, B. (2021). Hybrid stacked ensemble combined with genetic algorithms for Prediction of Diabetes. arXiv preprint arXiv:2103.08186.

16. Abdollahi, J., Nouri-Moghaddam, B., & Ghazanfari, M. (2021). Deep Neural Network Based Ensemble learning Algorithms for the healthcare system (diagnosis of chronic diseases). arXiv preprint arXiv:2103.08182.

17. Abdollahi, J., & Nouri-Moghaddam, B. (2021). Feature selection for medical diagnosis: Evaluation for using a hybrid Stacked-Genetic approach in the diagnosis of heart disease. arXiv preprint arXiv:2103.08175.

18. Chiesa, M., Maioli, G., Colombo, G. I., & Piacentini, L. (2020). GARS: Genetic Algorithm for the identification of a Robust Subset of features in high-dimensional datasets. BMC bioinformatics, 21(1), 54.

19. Wu C.C, Yeh WC, Hsu WD, Islam M, Nguyen PA, Poly TA, ET AL. Prediction of fatty liver disease using machine learning algorithms. Computer Methods and Programs in Biomedicine 2019;170: 23–9.

20. T. Condie, P. Mineiro, N. Polyzotis, M. Weimer, Machine learning on big data. Data Engineering (ICDE), 2013 IEEE 29th International Conference on. IEEE (2013), 1242-1244.

21. T.B. Murdoch, A.S. Detsky, The inevitable application of big data to health care, Jama 309 (2013), 1351- 1352.

22. Sharma, A. K., Nandal, A., Dhaka, A., & Dixit, R. (2020). A survey on machine learning based brain retrieval algorithms in medical image analysis. Health and Technology, 1-15.

23. Açıcı, K., Sümer, E., & Beyaz, S. (2021). Comparison of different machine learning approaches to detect femoral neck fractures in x-ray images. Health and Technology, 1-11.