ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research

Vol. 17, Issue, 02, pp.31964-31967, February, 2025
DOI: https://doi.org/10.24941/ijcr.48320.02.2025

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

SIMILARITY OF SVM AND RF REGRESSION IN PREDICTING BRAIN STROKE FOR IMPROVED ACCURACY IN HEALTHCARE

Yuvaraj[1] and *Dr. Raja, S.R.[2]

[1]MCA, PG Student, Centre for Open and Digital Education Hindustan Institute of Technology and Science, India
[2]Associtae Professor, M.C.A, M.Phil, Ph. D, Associate Professor, Centre for Open and Digital Education Hindustan Institute of Technology and Science, India

ARTICLE INFO

Article History:
Received 30th September, 2024
Received in revised form
15th November, 2024
Accepted 26th December, 2024
Published online 27th February, 2025

Key words:
Machine Learning, SVM, Novel RF, Brain Stroke, Prediction, Accuracy, Health.

***Corresponding author:**
Dr. Raja, S.R.

ABSTRACT

This work focuses on comparing the performance of a new support vector machine (SVM) and a random forest (RF), both of which have been developed as models for stroke prediction. By using precise prediction techniques, this research has the potential to avoid brain stroke and eventually lower death rates.A dataset consisting of 10 clinical and demographic variables was obtained from reliable sources. It contains 10 sample sizes in each group and totally 20 sample sizes for both the groups. After preprocessing the data, both models were trained and tested. The results showed that novel SVM outperformed RF in terms of predictive performance. This study suggests that machine learning algorithms can accurately predict brain stroke. The p-value of 0.001 in this research is statistically significant (i.e., smaller than the criterion of 0.05). Therefore, the two sets are distinguishable from one another statistically. An improvement of 99.01% was obtained by the innovative RF approach over the conventional SVM (83%) in this study, which shows the value of machine learning methods for predicting cerebral vascular disease.

Copyright©2024, Yuvaraj and Raja. 2025. *This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

Citation: Yuvaraj and Dr. Raja, S.R. 2025. *"*Similarity of SVM and RF Regression in Predicting Brain Stroke for Improved Accuracy in Healthcare*".* *International Journal of Current Research,* 17, (02), 31964-31967.

INTRODUCTION

When brain tissue suddenly loses its blood flow, this condition is known as a stroke. The lack of blood supply causes the brain cells to progressively die off, and the extent of disability varies based on which part of the brain is affected (Dritsas and Trigka 7 2022). Every year, stroke impacts approximately 16 million people globally and results in significant societal expenses (Sirsat, Fermé, and Câmara 10 2020).Timely identification of stroke is an essential prerequisite for effective treatment. Machine Learning (ML) technology has the potential to assist healthcare providers in making clinical predictions and decisions (Sirsat, Fermé, and Câmara 10 2020). In recent times, machine learning has become increasingly prevalent in addressing intricate challenges across various health and scientific fields, particularly in medical diagnosis or prognostic prediction (Lin et al. 7 2020). Brain stroke prediction using machine learning has been a subject of significant interest in the past few years, with a plethora of studies published in IEEE Explore and Science Direct. A total of 56 results were identified in IEEE Xplore, while Science Direct yielded 2763 results for the last 5 years.C. Heng Lin et al. work is a widely referenced study on the use of machine learning to the prediction of brain strokes. It utilized RF, SVM, and ANN algorithms to predict brain strokes among various subjects. (Lin et al., 2020, July 7)Other important research in this area includes A. Sudha, P. Gayathri, and N. Jaisankar, who discovered that the RF algorithm achieved a classification accuracy of 94.44% (Sudha, Gayathri, and Jaisankar, 2012) and V. Bandi, D. Agrawal, and S. According to research by Bandi, Bhattacharyya, and D. MidhunChakkaravarthy (12.2020), early diagnosis and treatment are crucial.Also, E. Dritsas and M. Trigka reported that the RF algorithm was the most accurate at predicting stroke risk, with an accuracy of 89.1%.However, the greatest publication in this area a paper written by M. Sanjay Sirsat, E. Fermé, and J. Câmara. This paper discusses the different machine learning algorithms and how they might be used to predict strokes. The article compares the accuracy of different algorithms and highlights the need for further research to develop more efficient and reliable predictive models.(Sirsat, Fermé, and Câmara 10 2020). In previous studies (Dritsas and Trigka 7 2022) SVM has not performed well in terms of accuracy of prediction in Brain stroke so this study aims to compare and identify a superior regression model that can facilitate early detection of stroke, thereby preventing loss of life.

MAT ERIALS AND METHODOS

Research on stroke prediction was conducted at the Machine Learning contrasting the performance of SVM or RF, two machine learning techniques. This research does not need ethical approval. The study utilized the "Stroke Prediction Dataset" by FEDESORIANO, obtained from Kaggle, which includes information on over 5000 patients and 12 attributes. The research compared two groups: Group 1 used SVM, while Group 2 used RF Regression. The sample size of 5110 patient records was the same for both groups and was determined using pre-test power analysis with an 80% power and an alpha value of 0.05 on clincalc.com (Tao et al. 2022). There are no NaN values in the dataset, but outliers exist in all continuous features. There are more women than men within our patient population. The correlation matrix reveals that age, hypertension, heart disease, and average glucose level have a significant correlation with the output. Elderly patients, particularly those over 60 years old, have a higher chance of stroke. Patients with hypertension and heart disease are also at a higher risk of stroke. A higher average glucose level increases the chances of stroke. Non-smokers and patients who have never consumed alcohol have a lower chance of stroke. Patients who are physically active and have a BMI between 18.5 to 24.9 have a lower chance of stroke. Stroke is more common in patients who reside in metropolitan regions than in those who do not. Finally, patients with a history of stroke in their family have a higher risk of stroke. The dataset underwent cleaning and preprocessing, including handling missing values, feature scaling, one-hot encoding, and outlier handling. It was then split into training and testing sets for SVM and RF regression models. Various dataset sizes were used to optimize accuracy. Evaluation metrics, such as accuracy, were used to assess the models. The models were tested on Google Colab, with necessary package imports and dataset uploads. Precision, recall, and the F1 score were only few of the indicators used to evaluate the model's efficacy. Planned Structure, as Depicted in Fig1.

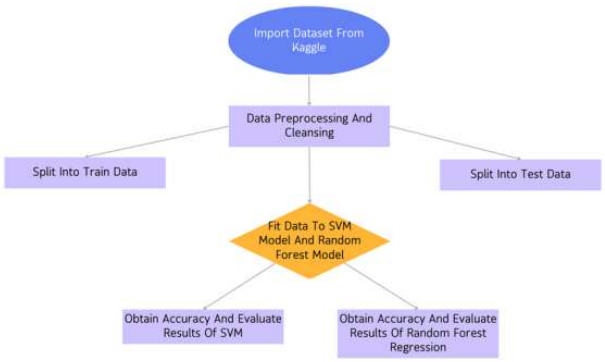

Fig. 1. Methodology of the research shown as a flowchart

Algorithm 1. Pseudo Code of the proposed method

Set the number of data sets to 10
Bring in the sklearn.metrics required for evaluation (roc_auc_score, accuracy_score, f1_score, recall_score, precision_score).
Build an evaluation grid that features SVC with probability=True &RandomForestClassifier with random_state=2022.
Make a list of several types of models, such as "SVM" and "RF."
Make the following tables without any data in them to hold the results of the evaluation metrics: roc_auc_scores, accuracy_scores, f1_scores, recall_scores, and precision_scores.
The number of data sets to iterate through is equal to range(num_data_sets).

a. By running train_test_split with the parameters X_sample, y_sample, test_size=0.3, or stratify=y_sample, random_state=i, we can create a training set and a testing set from the available data.
b. Iterate through (ml_model_names, clf) in enumerate(zip(model_names), models) for each model idx:
i. Use X_train and y_train as training data for a clf model fit.
ii. Predict labels for X_test using the fitted model clf, and save the results in y_pred.
iii. Get the testing data X_test's predicted probabilities from the fitted model clf and save them in y_pred_proba.
iv. You should add the roc_auc_scores for y_test and y_pred_proba to the ones you already have.
v. Accuracy scores from y_test and y_pred should be included to accuracy_scores.
vi. Include f1_scores for both y_test and y_pred.
vii. The recall scores from y_test and y_pred should be added to recall_scores.
viii. Accuracy scores from y_test and y_pred should be included to precision_scores.
ix. Results for the present model and data set's evaluation metrics should be printed.
a. "Model: {}, Data Set: {}".format(ml_model_names, i)
b. "Accuracy: {}".format(accuracy_scores[-1])
c. "ROC AUC Score: {}".format(roc_auc_scores[-1])
d. "F1 Score: {}".format(f1_scores[-1])
e. "Recall Score: {}".format(recall_scores[-1])
f. "Precision Score: {}".format(precision_scores[-1])
g. "\n"

To record the outcomes of the assessment metrics, establish a dictionary with the following keys and values: Model, Dataset, ROC, AUC, Accuracy, F1 score, recall, Precision. Convert the dictionary of score values into a pandas DataFrame and save it in the models_scores_df variable. The groupby function applied to models_scores_df, with the 'Model' and 'Accuracy Score' columns selected, yields a mean accuracy score for each model. The groupby function applied to models_scores_df, with the 'Model' and 'F1 Score' columns selected, yields a mean F1 score for each model. The groupby function applied to models_scores_df, with the 'Model' and 'Recall Score' columns selected, yields a mean recall score for each model.
The groupby function applied to models_scores_df, with the 'Model' and 'Precision Score' columns selected, yields the mean precision score for each model.

Statistical Analysis: In this study, SPSS software version 29 was utilized to perform statistical analysis on two groups, SVM and Logistic Regression. To evaluate the accuracy of the two algorithms, we calculated their means, standard deviations, standard errors, and subjected them to t-tests. In this investigation, stroke was the main endpoint. In order to determine the differences between the groups, a t-test was performed on independent samples, with a 95% confidence interval.

RESULTS

In Table 1, Independent data was used to evaluate a unique SVM against the RF technique for predicting brain stroke. We found that the SVM was only 83.42% accurate on average,

Table 1. Predicting brain stroke data value

	group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	SVM	10	83.4294	.55126	.17432
	RF	10	99.0158	.20957	.06627

Table 2. Comparing the predictive accuracy of a new SVM and RF

	Levene's Equality of variances				T-test for EOM				Confidence interval of the difference	
					Significance					
	F	Sig	t	df	1 p	2 p	MD	SED	L	U
EVA	7.227	.015	-83.575	18	<.001	<.001	-15.586	.186	-15.978	-15.194
EVNA			-83.575	11.548	<.001	<.001	-15.586	.186	-15.994	-15.178

Table 3. Calculates measures like precision, recall, accuracy, and F1 score.

Algorithm	Accuracy	F1 Score	Recall Score	Precision Score
SVM	83	86.46	93.96	80.07
RF	99.01	99	99.01	98.04

whereas the novel RF approach was 99.01% accurate. The SVM had a variation of 0.55126 for brain stroke prediction, whereas the novel RF method had a variance of 0.20957. Table .2, displays the results of sample-based experiments comparing the predictive accuracy of a new SVM to that of the RF technique for cerebral infarction. 95% CI = 0.015 because P>0.05 denotes statistical significance. These results suggest that neither group is significantly different from the other. Table 3:Calculates measures like precision, recall, accuracy, and F1 score to demonstrate the report's classification results using SVM and RF models.

EVA: Equal Variance Assumed; EVNA: Equal Variance Not Assumed; MD: Mean Difference; SED: Standard Error Difference; L: Lower; U: Upper.

Figure 2 compares the accuracy of RF to SVM algorithms. Results showed that when compared to the SVM approach, the mean accuracy of new RFs is 99.01%. The standard deviation of RF was also less than that of the novel SVM. The mean accuracy was displayed on the Y axis, and the results from both the conventional RF and the cutting-edge SVM algorithms were displayed on the X axis. The margin of error was demonstrated using the confidence interval and a standard deviation of two.

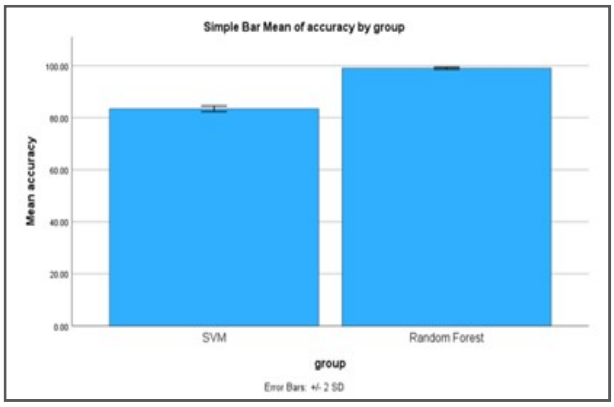

Figure 2. Contrast between RF's mean accuracy

DISCUSSION

In this research, we used SVM & RFto create models for predicting brain strokes based on preprocessed data and pertinent parameters. To evaluate the models, we calculated their accuracy, precision, recall, or F1 score on a separate test dataset. The results of stroke prediction using the proposed RF model improved over those using the prior SVM method. RF's average accuracy is 99.01%, whereas SVM only managed 98.09%. In comparison to the SVM model's 87.16% accuracy, the RF model achieves 90.07%. The F1 score for the RF model (99.03%) was also higher than the SVM model (85.77%). The findings show that the suggested model of RF regression outperforms the current SVM model in its ability to detect brain strokes. High accuracy may be achieved with a minimal amount of input features using the robust approach known as RF regression. When compared to other algorithms, such decision trees and neural networks, it is more robust and accurate in predicting strokes in the brain. Regarding the findings of the article, it appears that they have reported RF to have better accuracy than SVM in stroke prediction process. This is in contrast to our findings. However, depending on the specifics of the dataset and the characteristics being utilized, the accuracy of different machine learning algorithms may change. Therefore, it is not unusual to find different results in different studies.

CONCLUSION

Overall, the conclusion is that the proposed model of RF with 99% is a promising and effective approach for predicting brain strokes, and it may be worth exploring further in future research.

REFERENCES

Bandi, Vamsi, Debnath Bhattacharyya, and Divya Midhunchakkravarthy. 12 2020. "Prediction of Brain Stroke Severity Using Machine Learning." *Revue d'Intelligence Artificielle* 34: 753–61.

Dritsas, Elias, and Maria Trigka. 7 2022. "Stroke Risk Prediction with Machine Learning Techniques." *Sensors* 22. https://doi.org/10.3390/s22134670.

Islam, M. S., Hussain, I., Rahman, M. M., Park, S. J., & Hossain, M. A. (2022). Explainable artificial intelligence model for stroke prediction using EEG signal. Sensors, 22(24), 9859.

Saxena, Mrs Neha, Mr Arvind Choudhary, Mr Deep, Singh Bhamra, and Mr Preet Maru. 2022. "Issue 4 Www.jetir.org (ISSN-2349-5162)." Vol. 9. www.jetir.org.

Sirsat, Manisha Sanjay, Eduardo Fermé, and Joana Câmara. 10 2020. "Machine Learning for Brain Stroke: A Review." Vol. 29. W.B. Saunders. https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105162.

Sudha, A., P. Gayathri, and N. Jaisankar. 2012. "Effective Analysis and Predictive Model of Stroke Disease Using Classification Methods." *International Journal of Computers & Applications* 43 (14): 26–31.

Yu, J., Park, S., Kwon, S. H., Ho, C. M. B., Pyo, C. S., & Lee, H. (2020). AI-based stroke disease prediction system using real-time electromyography signals. Applied Sciences, 10(19),6791.

Lian, H., Xu, X., Shen, X., Chen, J., Mao, D., Zhao, Y., & Yao, M. (2020). Early prediction of cerebral-cardiac syndrome after ischemic stroke: the PANSCAN scale. BMC neurology, 20(1), 1-8.

****