

A Survey on Stroke Disease Classification and Prediction using Machine Learning Algorithms

Mrs. Veena Potdar¹,

¹Associate Professor,

Department of Computer Science and Engineering,
Dr. Ambedkar Institute of Technology, Bengaluru, India

Mrs. Lavanya Santhosh²,

²Assistant Professor,

Department of Computer Science and Engineering,
Dr. Ambedkar Institute of Technology, Bengaluru, India

Yashu Raj Gowda CY³

³MTech Student,

Department of Computer Science and Engineering,
Dr. Ambedkar Institute of Technology, Bengaluru, India

Abstract - Machine learning (ML) is a part of artificial intelligence (AI) that makes software applications to gain the exact accuracy to predict the end results not having to be directly involved to get the work done. This review aims to identify and analyze the Machine Learning approaches used for Stroke Prediction. We have considered the previously published works to review the Machine learning techniques used for Stroke Predictions. It's been found that the majority of the research work was done on mortality rate and functional outcome as the predicted outcomes. The most commonly used techniques were random forest, support vector machines, decision trees and neural networks. However, a few predictors and classifiers did primitive reporting standards for medical sector tools and none of which proved to be of any practical use.

Key Words: *Stroke prediction, Machine learning approaches, Sensitivity and Specificity, Comparison Analysis.*

1. INTRODUCTION

Many people fall victim to stroke and the numbers are increasing more in the developing countries. Several risk factors play a role in determining various types of stroke. Predictive algorithms establish a relationship between the risk factors and the types of strokes. Machine learning algorithms helps in early diagnosis and prevention of these stroke cases.

It is very difficult to predict the stroke symptoms and outbreaks taking note on the risk factors, since stroke is a complicated medical condition. This has enhanced the interests of people in technology sector to apply machine learning techniques to diagnose the stroke effectively by routinely collecting the datasets and delivering the accurate results for diagnosis. Furthermore, many papers have been published frequently which explains machine learning techniques to address the issue.

The agenda of this survey paper is to identify the better machine learning techniques used to predict stroke, which will also help to understand and resolve the problem in more effective ways.

2. RELATED MACHINE LEARNING APPROACHES

In this section, analysis and review is being done on the previously published papers related to work on prediction of stroke types using different machine learning approaches. At least, papers from the past decade have been considered for the review. They are explained below:

In 2014, Hamed Asadi, Richard Dowling, Bernard Yan, Peter Mitchell [1], conducted a look back study on a potential database of acute ischemic stroke. They did a comparison study on various machine learning techniques which may help in determining the end result of cardiovascular intervention in acute anterior circulation stroke of ischemic type. They appended 107 acute anterior ischemic stroke patients who were medicated by the cardiovascular method. The model included all the information of the patient, operation and medical factors. They have used algorithms such as MATLAB, SPSS, artificial neural networks & support vector to outline a supervised machine which is able to classify the above predictors. Regardless of using a small dataset, it showed an optimistic accuracy closer to 70% of predictive output using supervised machine learning. They also came up with a popular machine learning system that can predict the selection method for endovascular versus medical treatment in the running of acute stroke.

In 2018, a team of National Institute of Engineering, Karnataka [2], regulated a survey on AI applications in stroke and aimed to predict the accurate results of occurrence of stroke. They also used optimistic algorithms and frameworks that include patients' aspects like gender, age, height, BMI, etc., they built a data model using decision tree algorithm to analyze these parameters. The outcome was analyzed by using confusion matrix and the accuracy was 95%. To achieve this, they built the training model that helped to compare the newly fed data with the survey data. And the report was generated on the basis of this comparison.

In 2017, a team of medical university in Taiwan developed a model to automate the early detection of Ischemic stroke. They collected CT images of the brains to analyze the possibility of stroke. The system preprocessed

the CT images to rule out the unlikely areas for development of stroke. Later, by data augmentation method they selected out the patch images to enhance the quantity of these images. Using 256 patch images, they trained and tested the CNN module which was able to recognize ischemic stroke. It was observed that the proposed showed more than 90% result [3].

In 2012, Sudha. A under the guidance of her professors N. Jaisankar & P Gayathra, proposed a stroke predictive Model using classification (PCA) techniques. They used classification algorithms- decision Tree, Naïve Bayes & Neural Networks for predicting the stroke types with related attributes. They utilized principle component analysis algorithm (PCA) for dimension reduction. They studied & used sensitivity 7 accuracy indicators for evaluation. Decision tree achieved 95.29% of sensitivity & 98.01% of accuracy. Bayesian classifier achieved 87.10% & 91.30% respectively. They compared these techniques and chose the decision tree as the best classification Method. The proposed Model takes the patient detail, & checks with reduced attributes. The accuracy measured

Sl. No	Paper Title	Method Used	Result
1	An automatic detection of ischemic stroke using CNN Deep learning algorithm	Image pre-processing computer aided detection, Data augmentation, Convolutional Neural Network	It showed more than 90% accuracy
2	Effective Analysis and Predictive Model of Stroke Disease using Classification Methods	Decision Tree, Bayesian Classifier, Neural Networks	The accuracy of C4.5decision tree algorithm and KNN were 95.42% and 94.18% respectively, in stroke prediction

based on sensitivity & specificity. It was observed that Neural networks performance had more exactness when compared with other two classification techniques [4].

In 2010, Adithya Khosla and his team, proposed a new automatic feature selection algorithm that sorts out the popular attributes based on proposed rule; conservative average. They merged it with support vector machines, to aim a significant area under the ROC curve (AUC). Further-more, they proposed a margin-based regression algorithm that merged with margin-based classifiers to attain a finer similarity index more than the cox model. This model can be implemented for medical prediction of other diseases, where missing information are familiar & the consequences are not well understood. However, they concluded that this feature selection algorithm may not work well in other datasets with highly matched features as it evaluated the performance of each feature singly. To overcome this issue, they used an L1 regularized feature selection algorithm, to trim the features before applying conservative mean feature selection for fine-tuning [5].

In 2019, Department of Computer Architecture and Automation team of Universidad complutense de Madrid,

Spain researched about the testing the hypothesis that state of art machine learning based methods. They studied that confined monitoring technologies could help in the treatment of stroke. These techniques can even be employed for predicting future chance like the ultimate death of the patient. The collected dataset consisted of 119 patients with 7 predictors & 2 target variables which are used for prediction of stroke type & prognosis of death. They used 7 different machine learning algorithms, which are decision tree, KNN, logistic regression, naïve bayes, neural network, random forest & SVM & evaluation was done over all the 6 different metrics. Furthermore, they used 10-fold cross validation re-sampling method, for guaranteed validation set from training one & the validation of the trained classifier against on hidden sample. The model evaluation metrics used to differentiate the algorithms used were; sensitivity, accuracy, F measure, specificity and locale under the ROC as well as PRC. Among all of them, Random forest models yielded the best performance in the treatment of stroke types & prediction of deaths with values of 0.93±0.03 and 0.97±0.01 respectively [6].

In 2015, Balar Khalid & Naji Abdelwahab, proposed a model for predicting Ischemic stroke using Data mining Techniques which were classification, logistic regression. They studied the risk factors of ischemic stroke. Then they used data software WEKA 3.6 & C4.5 algorithm & logistic regression for preprocessing, cleaning & analyzing the data. It was observed that the model of logistic regression in their case study allowed then to analyze the correlation between the occurrence of ischemic stroke & its risk factors. The XLSTAT software showed a very good sensitivity of 77.58% & specificity of 83% respectively. The ROC Curve is sensitivity according to specificity. However, they concluded that prediction model achieved 19.7% error rate [7].

Table -1: Analysis of Methods and results

3	Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy using Machine Learning	SPSS, MATLAB, Rapid Miner, ANN, Support Vector algorithm	It showed a promising accuracy up to 70% of predicting outcome.
4	Comparison of different machine learning approaches to model stroke subtype classification and risk prediction	Decision tree, KNN, SVM, Neural Network, Logistic Regression, Random Forest, Naïve Bayes	Random Forest Model showed best performance with mean values of 0.93±0.03 and 0.9±0.01 respectively.
5	A model for prediction of end results in acute stroke using machine learning	Random Forest, Logistic Regression, Deep Neural Network	Deep neural network showed the highest accuracy.
6	Stroke prediction using an integrated machine learning approach	Conservative mean feature selection, L1 regularized logistic regression novel prediction algorithm	Overall, this approach outperformed the current state-of-the art in both metrics of AUC and Concordance index
7	Prediction of Ischemic Stroke using different approaches of data mining	SVM, penalized logistic regression (PLR) and Stochastic Gradient Boosting (SGB)	The AUC values with 95% CI were 0.9783 for SVM, 0.9757 for SGB and 0.853 for PLR respectively.

3. RESULTS

The evaluation on each paper was done by carefully studying and analyzing them. Henceforth, the remarks with respect to each paper are mentioned below.

- Hamed Asadi et al. [1] proposed a model for acute stroke. They have taken a large database and used many algorithms to design a supervised machine. But, the accuracy showed only 70% of the predictive result.
- Aishwarya Roy et al. [2] proposed a model that can assist the doctors in clinical trials. In this paper, they haven't mentioned clearly about the dataset they had used for prediction model and also about the method they have used.
- Chiun-Li-Chin et al. [3] used very less patch images (about only 256) to train the model, which decreases the efficiency of the system. But this proposed model can be effectively used by doctors to diagnose the diseases.
- A. Sudha et al. [4] observed that the effect of data dimension reduction on classification accuracy and other algorithm performance criteria must be examined. Most investigations are presented theoretically. So this field is unfamiliar to medical specialists.
- Aditya Khosla et al. [5] showed that cox model can be implemented for medical prediction of other types of diseases, but missing data here is usual and the consequences are not clearly understood.
- Luis Garcfa-Terriza et al. [6] have used different algorithms to predict stroke type (hemorrhagic v/s Ischemic) and to predict further complications of diseases. It helps the doctors to use preventive remedies to avoid unpleasant situation.
- Balar Khalid et al. [7] built a model with the aim of stroke disease prediction. But to achieve supreme quality medical data, all the necessary steps ought to be taken to build the medical information system that provides the accurate knowledge on the patient's medical history rather than their billing invoices.

4. CONCLUSIONS

Many machine learning techniques have contributed to predict stroke in several different scenarios. Deciding to use a specific machine learning technique should be based on considerations of scenarios, datasets, parameters and other analysis. We cannot conclude on the best technique to use for stroke prediction. Each technique has its own advantages and disadvantages. It is wise to choose one among them based on the necessity of the individual problem statement. One must perform statistical analysis and initialization to decide on the specific technique or model to use. However, random forest is considered as one of the most robust and approved technique for assessing a quantity for the sample data as it shows promising results.

REFERENCES

- [1] Hamed Asadi, Richard Dowling and Bernard Yan, "Machine Learning for outcome prediction of acute ischemic stroke", PLOS ONE Vol.9 Issue2, Feb2014.
- [2] Aishwarya Roy, Anwesh Kumar, Navin Kumar Singh and Shashank D, "Stroke Prediction using Decision Trees in Artificial Intelligence", IJARIT, Vol.4, Issue2, 2018, pp:1636-1642.
- [3] Chiun-Li-Chin, Guei-Ru Wu, Bing-Jhang Lin, Tzu-ChiehWeng, Cheng-Shiun Yang, Rui-CihSu and Yu-Jen Pan, "An Automated Early Ischemic Stroke Detection System using CNN Deep Learning Algorithm", IEEE 8th International Conference on Awareness Science and Technology, 2017.
- [4] A. Sudha, P. Gayathri, "Effective analysis & predictive model of stroke disease using classification methods", IJCA(0975-8887), Vol.43-No.14, April 2012.
- [5] Aditya Khosla, Yu cao, Honglak Lee & Associates, "An integrated machine learning approach to stroke prediction", 25-28 July 2010, Washington, DC, USA.
- [6] Luis Garcfa-Terriza, Risco Martin, Ayala and Gemma Reig Rosello, "Comparison of different Machine Learning approaches to model stroke subtype classification and risk prediction", Society for Modeling & Simulation International(SCS), 2019 April 29-May 2.
- [7] Balar Khalid and Naji Abdelwahab, "A model for predicting Ischemic stroke using Data Mining algorithms", IJSET, Vol. 2 Issue 11, Nov 2015, ISSN: 2348-7968.