AN ANALYSIS OF BRAIN STROKE PREDICTION USING MACHINE LEARNING

MURALI PONAGANTI

Associate Professor & HOD, Department Of MCA, Sree Chaitanya College of Engineering, Karimnagar

ABSTRACT:
A stroke, also known as a cerebrovascular accident or CVA, is when part of the brain loses its blood supply and the part of the body that the blood-deprived brain cells control stops working. This loss of blood supply can be ischemic because of lack of blood flow, or haemorrhagic because of bleeding into brain tissue. A stroke is a medical emergency because strokes can lead to death or permanent disability. There are opportunities to treat ischemic strokes, but that treatment needs to be started in the first few hours after the signs of a stroke begin. The patient, family, or bystanders should activate emergency medical services immediately should a stroke be suspected. A transient ischemic attack (TIA or mini stroke) describes an ischemic stroke that is short-lived where the symptoms resolve spontaneously. This situation also requires emergency assessment to try to minimize the risk of a future stroke. By definition, a stroke would be classified as TIA if all symptoms resolved within 24 hours. According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible to approximately 11% of total deaths. For survival prediction, our ML model uses dataset to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Unlike most of the datasets, our dataset focuses on attributes that would have a major risk factor of a Brain Stroke.

I. Introduction

Machine Learning (ML) delivers an accurate and quick prediction outcome and it has become a powerful tool in health settings, offering personalized clinical care for stroke patients. An application of ML and Deep Learning in health care is growing however, some research areas do not catch enough attention for scientific investigation though there is real need of research. Therefore, the aim of this work is to use ML algorithms like Logistic regression, SVM, KNN, Decision Tress and Random Forest to determine and predict the risk of Brain Strokes. A total of 39 studies were identified from the results of ScienceDirect web scientific database on ML for brain stroke from the year 2007 to 2019[2]. Support Vector Machine (SVM) is obtained as optimal models in 10 studies for stroke problems. Besides, maximum studies are found in stroke diagnosis although number for stroke treatment is least thus, it identifies a research gap for further investigation. Similarly, CT images are a frequently used dataset in stroke. Finally, SVM and Random Forests are efficient techniques used under each category [2]. The present study showcases the contribution of various ML approaches applied to brain stroke.

II. Literature Survey

In the research conducted by Manisha Sirsat, Eduardo Ferme, Joana Camara, the main aim of the research was to classify state-of-arts on ML techniques for brain stroke into 4 categories based on their functionalities or similarity, and then review studies of each category systematically. The study further discusses the outcomes and accuracies obtained by using different Machine Learning models using text and image-based datasets.

In this study, the authors discussed many stroke related problems from the state-of-art. The reviewed studies were grouped in several categories based on their similarities. The study notes that it is difficult to compare studies as they employed different performance metrics for different tasks, considering different datasets, techniques, and tuning parameters. Hence, it only mentions the research areas which were targeted in more than one study and the studies which report highest classification accuracy in each section.

Harish Kamal, Victor Lopez, Sunil A. Sheth, in their study discus how Machine Learning (ML) through pattern recognition algorithms is currently becoming an essential aid for the diagnosis, treatment, and prediction of complications and patient outcomes in several neurological diseases. The evaluation and treatment of Acute Ischemic Stroke (AIS) have experienced a significant advancement over the past few years, increasingly requiring the use of neuroimaging for decisionmaking. This study offers an insight into the recent developments and applications of ML in neuroimaging focusing on acute ischemic stroke. The implementations of machine learning are numerous, from

early identification of imaging diagnostic findings, estimating time of onset, lesion segmentation, and fate of salvageable tissue, to the analysis of cerebral edema, and predicting complications and patient outcomes after treatment.

The paper finally concludes by discussing how Machine learning applications are expanding in the medical field for diagnostic and therapeutic purposes, and the rapidly expanding and increasingly neuro-imaging reliant field of AIS is proving to be fertile ground. There is a particular need for ML solutions in this field, which is faced with the challenge of increasingly complex data, with limited human expert resources. Future directions in ML for AIS may require collaborative approaches across multiple institutions to build a robust dataset for efficient training of ML networks [2].

In the research conducted by Chuloh Kim, Vivienne Zhu, Jihad Obeid and Leslie Lenert, they have assessed performance of natural language processing (NLP) and machine learning (ML) algorithms for classification of brain MRI radiology reports into acute ischemic stroke (AIS) and non-AIS phenotypes. The method followed included All brain MRI reports from a single academic institution over a two-year period were randomly divided into 2 groups for ML: training (70%) and testing (30%). Using "quanteda" NLP package, all text data were parsed into tokens to create the data frequency matrix. Ten-fold crossvalidation was applied for bias correction of the training set. Labelling for AIS was performed manually, identifying clinical notes. They applied binary logistic regression, naïve Bayesian classification, single decision tree, and support vector machine for the binary classifiers, and we assessed performance of the algorithms by F1-measure. They also assessed how n-grams or term frequency-inverse document frequency weighting affected the performance of the algorithms.

The paper concluded with the understanding how supervised ML based NLP algorithms are useful for automatic classification of brain MRI reports for identification of AIS patients. Single decision tree was the best classifier to identify brain MRI reports with AIS [3].

In the research conducted by R. Punitha Lakshmi et al [4], they put forward their work on SVM Classifier Based On Otsu Thresholding For Ischemic Stroke Detection. The dataset used in order to train the algorithms/models were a set of 32 different types of brain MRI images which were in JPEG format. Both the classifiers i.e. the Random Forest Classifier and SVM Classifier were trained with the help of these images but with different procedure. All the MRI Images were first transformed using the wavelet transformation and the segmentation of those images were carried out by Otsu Thresholding.

The images are obtained from Open Access Series of Imaging Studies (OASIS) which makes the MRI data sets of the brain which is available for the research purpose. Noise reduction of these images were done in pre-processing so as to get accurate results. After that, the data was then fed as input to the SVM Classifier. Thus, the maximum accuracy was given by SVM Classifier being 88% and Random Forest Classifier being at 81%.

The paper concluded with the understanding of how maximum accurate segmentation of brain and brain lesions is achieved with the help of SVM Classifier based on Otsu Thresholding and the dataset with scattering lesion tissues can also help to improve further accuracy rates of this Classification [4].

In the research conducted by Jaehak Yu et al, an implementation of system for semantic analysis of early detection of stroke and also the recurrence of stroke in Koreans over the age of 65 years based on the National Institute of Health (NIH) Stroke Scale was done by the researchers. The research was made possible with the help of data which was collected from the emergency medical center of the Chungnam National University Hospital consisting of 287 stroke patients out of which 16 patients, which had no stroke symptoms were excluded. Final NIHSS Data consisted of 227 patients, excluding the 60 patients whose data included missing values or outlier values among the NIHSS questionnaires. Patient subjects were the elderly over 65 years old, and consisted of 117 men and 110 women.

The Machine Learning Algorithm which was used was C4.5 Decision Tree Algorithm. The researchers found out that it is the most advanced algorithm and its function of classification and prediction is already proven. The proposed system in this experiment classifies and predicts stroke severity score into four classes using representative classification and prediction models of machine learning and data mining methodology. To measure the experiment

accuracy of the proposed system, the recall and precision are used as the measurements. The experiment resulted in faster and more accurate predictions of stroke severity and efficient system operation with the help of various Machine Learning algorithm used and C4.5 decision tree and Random Forest classified and predicted the performance with high accuracy.

The paper concluded with the understanding of how efficient use of Machine Learning Methodologies and a proper dataset to build a model to predict Brain Stroke and also assess the severity of symptoms to predict results with high precision can be implemented to build a system providing an alarm service to visit a medical centre or hospital in real-time [5].

Gangavarapu Sailasya, Gorli L. Aruna Kumari, in their study discuss how Brain Stroke, which is the fouth leading cause of death in India, can be predicted with the help of trained Machine Learning Models so as to minimize risk of death due to Brain.

III. System Analysis

System Architecture:

Existing System:

Very few systems use the available clinical data for prediction purposes and even if they do ,they are restricted by the large number of association rules that apply.Diagnosis of the condition soley depends upon the Doctor's intuition and patient's records.The decision support system and will prove to be an aid for the physicians with the diagnosis.The algorithm,Fuzzy c means uses clustering and makes use of clusters and data points to predict the relativity of an attribute .Each data point is associated with multiple clusters depending upon the membership degrees.

Disadavantages:

- Detection is not possible at an earlier stage
- Parctical use of various collected data is time consuming

Proposed System:

The proposed system acts as a prediction support machine and will prove as an aid for the user with diagnosis. The algorithms used to predict the output have potential in obtaining a much better accuracy then the existing system. In proposed system, the practical use of various collected data has turned out to be less time consuming. .we calculate accuracy of machine learning algorithms for predicting heart disease, for this algorithms are k-nearest neighbor, decision tree, linear regression and support vector machine(SVM) by using UCI repository dataset for training and testing.

Adavantages:

- High performance and accurancy rate
- Machine Learning Algorithms is very flexible and is widely used in various domains with high rates of success.
- Data and information collected for prediction is easily available to the users. System provides users with precautions that can be taken to reduce risk factors.

System Study

Feasibility Study

The feasibility of the project is analyzed in this phase and the business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be

carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are,
- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

Economical Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

IV. Implementation

Modules Description:

- Users
- Data Collection
- Attribute Selection
- Preprocessing of data.

Users:

Users add the data to the database and view the data to the view data and predict the Brain stroke disease using ml.

Data Collection:

First step for predication system is data collection and deciding about the training and testing dataset. In this project we have used 73% training dataset and 37% dataset used as testing dataset the system.

Attribute Selection:
Attributes of dataset are property of dataset which are used for system and for Brain stroke many attributes are like heart bit rate of person, gender of the person, age of the person and many more predication system.

Preprocessing of data:

Preprocessing is needed for achieving prestigious results from the machine learning algorithms. For example, Random Forest algorithm does not support null values dataset and for this we have to manage null values from original raw data. For our project we have to convert some categorized value by dummy value means in the form of "0"and "1" by using following code.

Admin:
Admin will give authority to Users. In order to activate the users. the admin can Prediction Brain stroke Disease.

V. Conclusion:

After the literature survey, we came to know various pros and cons of different research papers and thus, proposed a system that helps to predict brain strokes in a cost effective and efficient way by taking few inputs from the user side and predicting accurate results with the help of trained Machine Learning algorithms. Thus, the Brain Stroke Prediction system has been implemented using the given 5 Machine Learning algorithm given a highest accuracy of 98.56%. The system is therefore designed to provide simple yet efficient User Interface design with an empathetic approach towards their users and patients. The system has potential for future scope which could lead to better results a better user experience. This will help the user to save their valuable time and will help them to take appropriate measures based on the results provided.

The future scope for the implemented system can be:
1. Increasing the accuracy of the model.
2. Additional information about brain stroke can be explained. 3. Allowing users to visualize their results based on their inputs.

References
[1] Manisha Sirsat, Eduardo Ferme, Joana Camara, "Machine Learning for Brain Stroke: A Review," Journal of stroke and cerebrovascular diseases: the official journal of National Stroke Association (JSTROKECEREBROVASDIS), 2020.
[2] Harish Kamal, Victor Lopez, Sunil A. Sheth, "Machine Learning in Acute Ischemic Stroke Neuroimaging," Frontiers in Neurology (FNEUR), 2018.
[3] Chuloh Kim, Vivienne Zhu, Jihad Obeid and Leslie Lenert, "Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke," Public Library of Science One (PONE), 2019.
[4] R. P. Lakshmi, M. S. Babu and V. Vijayalakshmi, "Voxel based lesion segmentation through SVM classifier for effective brain stroke detection," International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017.
[5] J. Yu et al., "Semantic Analysis of NIH Stroke Scale using Machine Learning Techniques," International Conference on Platform Technology and Service (PlatCon), 2019,
[6] Gangavarapu Sailasya and Gorli L Aruna Kumari, "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms," International Journal of Advanced Computer Science and Applications (IJACSA), 2021.
[7] "Stroke Prediction Dataset". Kaggle.Com, 2021, https://www.kaggle.com/fedesoriano/stroke-predictiondataset. Accessed 6 Oct 2021.