BRAIN STROKE PREDICTOR

E.R. Shyama

06-09-2023

Problem Statement

A stroke happens when a section of the brain experiences a reduction in blood flow, which results in the dysfunction of the body part that those brain cells control. This decrease in blood supply may be ischemic or haemorrhagic due to inadequate blood flow or bleeding into the brain tissue. A stroke is a medical emergency due to the potential for death or permanent disability.

Stroke significantly burdens both people and national healthcare systems. Stroke risk factors that may be altered include hypertension, diabetes, and cardiovascular disease, atrial fibrillation, abnormal glucose metabolism, and lifestyle risk factors. Therefore, the goal of our study is to correctly predict the stroke utilising large existing data sets based on potentially modifiable risk components using machine learning approaches.

Numerous apps currently in use implement brain stroke prediction utilising naive bayes and decision tree algorithms. It becomes challenging for the decision tree to decide on the proper threshold to divide the data points into distinct nodes in the case of continuous characteristics. Naive Bayes bases its usefulness in real-world use situations on the supposition that all features are independent. This approach encounters the "zero-frequency problem," where it gives a categorical variable with zero probability if its category was not present in the training dataset but was present in the test data set. Different machine learning models, like Logistic Regression, K-Nearest Neighbours, AdaBoost Classifier, XGBoost Classifier, and Random Forest Classifier, are used in the proposed solution. These models can be used to predict stroke and may be utilised by doctors to do so in the real world.

Market/Customer Need Assessment

With 400–800 strokes per 100,000, 15 million new acute strokes annually, 28,500,000 disability adjusted life years, and 28–30-day case fatality rates ranging from 17% to 35%, stroke is the second greatest cause of death and adult disability globally. With the number of fatalities from heart disease and stroke expected to rise to five million in 2020 from three million in 1998, the burden of stroke is anticipated to get worse. This will happen as a result of the ongoing demographic and health changes that will lead to an increase in the risk factors for vascular disease and the senior population. 85% of all stroke deaths worldwide occur in developing nations. The effects of stroke on society and the economy are significant. According to estimates, In the United States of America (USA), the cost of stroke was anticipated to be as high as $49 billion in 2002, whilst expenditures upon discharge were estimated to total of around 3 billion Euros in France.

Target Specification and Characterization

The proposed system can be used by clinics and doctors in more analysis on a patient so that he/she can be given more medical attention. The suggested system operates as a machine that aids in diagnosis and supports forecasts. This software is designed to be used by hospitals and other healthcare organisations so that they can give their patients preventive care by anticipating the possibility of brain stroke.

External Search

- About Stroke in Wikipedia - https://en.wikipedia.org/wiki/Stroke
- "Mayo Clinic" Stroke - Symptoms and causes - https://www.mayoclinic.org/diseases-conditions/stroke/symptoms-causes/syc-20350113
- https://www.jetir.org/papers/JETIR2204518.pdf
- https://towardsdatascience.com
- https://www.foreseemed.com/blog/machine-learning-in-healthcare

Business Model

Stroke cannot be easily predicted by evaluating doctor's diagnostic data. By predicting the likelihood of a stroke occurring in patients, an AI-based tool enables neurologists and cardiologists to give better treatment to their patients. Patients who sought treatment for a stroke as soon as possible had a good chance of continuing to be healthy for years to come.

A smartphone app powered by AI might be released on the market to help medical professionals. This opens up the possibility of marketing a service-based business model where customers subscribe to apps by paying monthly fees.

Concept Generation

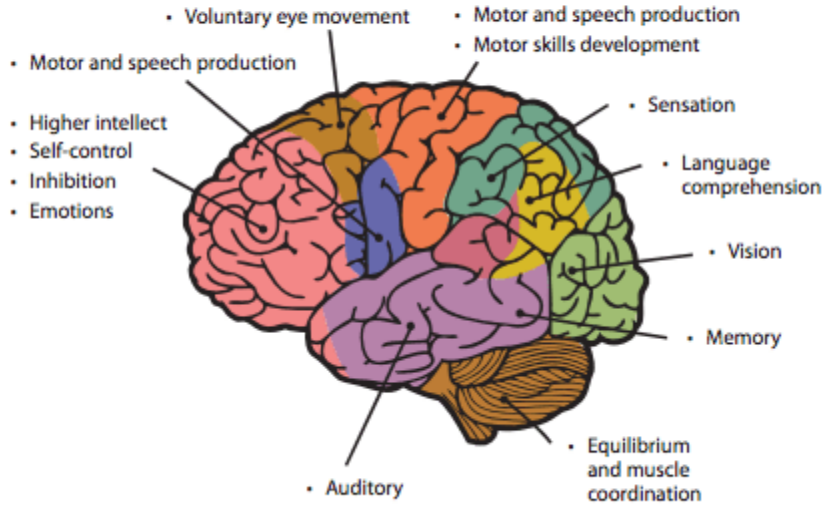
When the brain's blood supply is cut off or there is unexpected brain haemorrhage, a stroke may result. There are two distinct stroke kinds. An ischemic stroke is one when there is a blockage of blood flow to the brain. Blood cannot provide the brain with nutrition and oxygen. Brain cells start to die within minutes of being deprived of oxygen and nutrition. Hemorrhagic strokes are caused by abrupt bleeding in the brain and are the most common type of stroke. Blood leakage causes pressure on brain cells, which harms them.

Strokes that are ischemic (blood arteries are blocked) account for little under 90% of cases, whereas haemorrhagic strokes account for the remaining 10%. Based on the location of the blockage or bleeding in the brain, strokes are further categorised.

A stroke is an urgent medical matter. A stroke may result in permanent brain damage, chronic disability, or even fatality. Mild weakness, paralysis, or numbness on one side of the body or face can all be symptoms of a stroke. Other symptoms might be a sudden, strong headache, abrupt weakness, difficulty seeing, difficulty speaking or comprehending speech, and difficulty looking.

Functional areas of the brain

This illustration shows the brain's functional areas. After a stroke, deficits in function depend on which cerebral artery is affected.



We must thus utilise machine learning to create the model based on the aforementioned challenge. The study of computer algorithms that develop automatically via usage and learning from data is known as machine learning (ML). It is considered to be a component of artificial intelligence. Without being expressly taught to do so, machine learning algorithms create a model using sample data, sometimes referred to as training data, in order to make predictions or judgements.

Concept Development

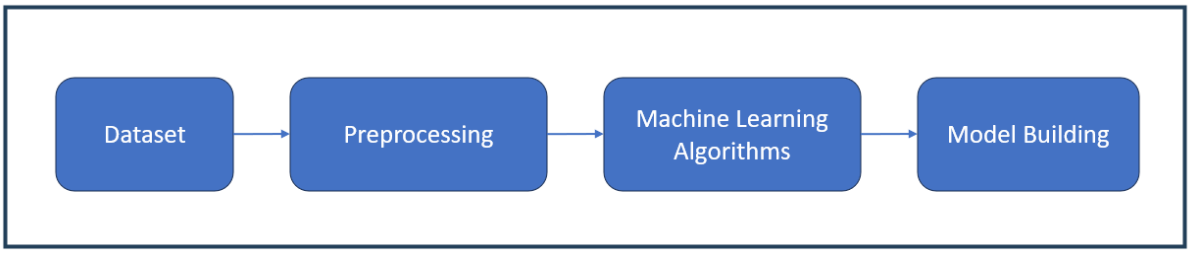
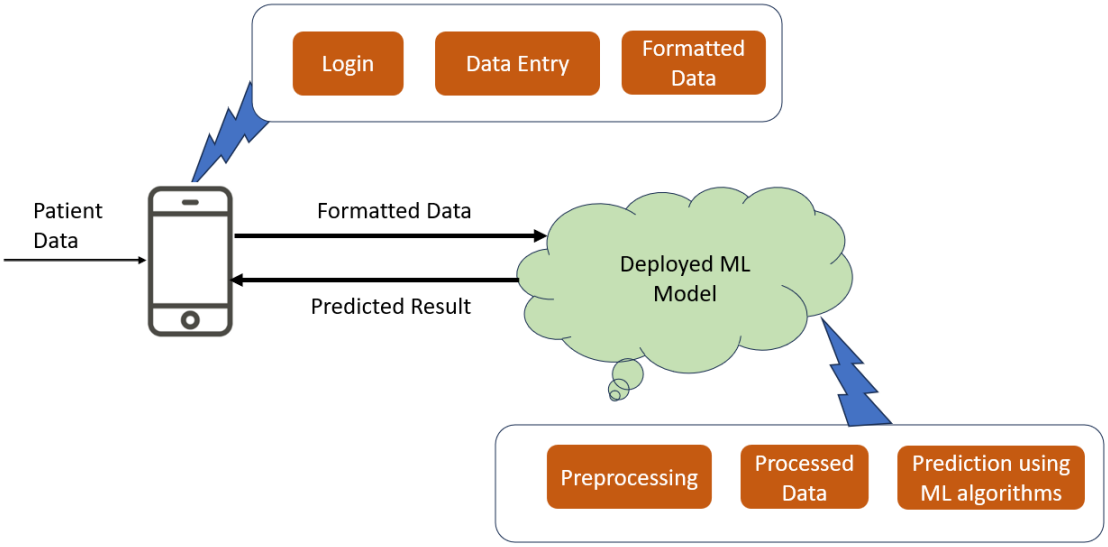


Fig : - Block diagram for the proposed system

Python is the programming language used by the Jupyter, which is where machine learning algorithms are learned and performed. A stroke dataset is first prepared. This dataset contains details of patients about hypertension, heart disease, type of employment and domicile, body mass index, average blood sugar level, smoker status, family history of stroke, blood pressure, and cholesterol levels. The dataset is made up of information gathered from individuals with a range of medical issues, from people of all ages, and from both genders.

The Jupyter environment imports this dataset. On the dataset, preprocessing is carried out. Boxplot diagrams are used to identify outliers and treat them. Statistical techniques like mean, median, and mode are used to identify and deal with missing values. The process of transforming categorical information into numerical representation is known as label encoding or one hot encoding. The dataset is then divided into training and test data. The Python has a number of libraries for training machine learning algorithms, including seaborn, scikit learn, and pandas. The model is then trained and tested using a variety of classification techniques, including Logistic Regression, K-Nearest Neighbours, AdaBoost Classifier, XGBoost Classifier, and Random Forest Classifier. Extensive testing is done on the chosen model and the algorithm that offers the best accuracy is chosen.

Final Product Prototype



The hospital uses an android application as part of the system. The patient's information is gathered by the hospital and translated to a JSON file format. All of the fundamental elements needed by the model will be included in JSON format, together with the metadata (patient name, hospital name, phone number). After then, it gets moved to AWS Cloud. The pre-processing takes place here. Following preprocessing, the processed data is fed to the ML models in order to make the prediction. The application receives and displays the projected outcome.

Conclusion

Researchers are creating models that are more and more accurate as datasets grow and improve in quality. ML will likely replace our neighborhood doctors in the future decades, which is fairly exciting even if we might not witness AI doing the role of a doctor today. The majority of ML models still lack sufficient data and are biased, so there is still a long way to go. Machine learning can be trained just as well as doctor prognosis, and prognosis is not an additional cost. The actual implementation is to be done to have clear understanding on the performance of the solution.