

ARTIFICIAL INTELLIGENCE FOR PREDICTION OF BRAIN STROKE

Name: Mr. Shubham Shekhar Gujar

Date: 2nd September 2023

Abstract:

An application of Artificial Intelligence in health care is growing however, some research areas do not catch enough attention for scientific investigation though there is real need of research. With the help of artificial intelligence, machine learning, and data science techniques, significant advancements have been made in the realm of clinical and medical services. Human brain is complex in structure. Stroke is one of the most serious diseases worldwide, directly or indirectly responsible for a significant number of deaths. Stroke is one of the most serious diseases worldwide, directly or indirectly responsible for a significant number of deaths. This situation also requires emergency assessment to try to minimize the risk of a future stroke. Early diagnosis of brain stroke is very important so that immediate treatment get started. So to overcome this condition we use Artificial Intelligence & Machine Learning techniques for the prediction of brain stroke.

1. Problem Statement:

Stroke is the second leading cause of death and leading cause of adult disability worldwide with 400-800 strokes per 100,000, 15 million new acute strokes every year. people who have had strokes generally have abnormal biosignals and they take that symptoms casually so that this situation will be occurs. Neurologists can diagnose the brain stroke by taking test like MRI as well as CT scan to predict what happen in our brain so that it will take time to diagnose this. Recently, AI have shown great promise in accurately predicting stroke risk by using machine learning. So we

can use machine learning algorithms which is subset of Artificial Intelligence for early prediction of Brain Stroke.

2.Introduction

Stroke is a ailment that impacts vessels that supply blood to the thoughts. mind stroke takes region which list blood glide to the mind is each reduced or interrupted. whilst this occurs, the mind no longer gets sufficient oxygen or other crucial components, and the brain cells start to die. A stroke effects important lengthy-time period incapacity or demise. mind stroke is one of the leading causes of death all around the world. There are 3 kinds of brain strokes: ischemic strokes, hemorrhagic strokes, and transient ischemic assault (TIA), which is also referred to as a caution or mini- stroke . Ischemic strokes arise due to loss of blood supply, and hemorrhagic strokes occur because of ruptured blood vessels.

The most typical kind IS ISCHEMIC STROKE stroke is this one. It occurs when the blood arteries in the brain narrow or block, significantly reducing the amount of blood flow (ischemia). Fat deposits that accumulate in blood vessels or blood clots or other debris that move through the bloodstream, typically from the heart, and lodge in the blood vessels in the brain cause blocked or restricted blood arteries.

Brain bleeding results in a hemorrhagic stroke. This may occur when a brain blood artery rupture or when bleeding occurs in the brain tissue. Pressure brought on by bleeding, edema, or a lack of blood flow can all contribute to hemorrhagic stroke damage. An ischemic stroke, which is a stroke brought on by a stopped blood supply, can result in bleeding in the brain tissue. As a result, the brain's tissue is harmed.

Transient ischemic attack, or TIA for short, is a dangerous repercussion. A TIA causes a temporary interruption in the blood supply to a portion of the brain. Another name for it is a "mini stroke," but don't be deceived by the diminutive. A TIA may be a precursor to a full blown stroke.The most common cause of TIAs is a blood clot that becomes stuck in an artery that carries blood to the brain. Your brain is oxygen-starved and unable to function normally if there isn't regular blood flow.

3. Customer Needs Assessment:

The World Stroke Organisation estimates that 13 million people worldwide experience a stroke each year, leading to 5.5 million fatalities. People who have had strokes generally have abnormal biosignals and they take that symptoms casually so that this situation will

be occurs. Neurologists can diagnose the brain stroke by taking test like MRI as well as CT scan to predict what happens in our brain so that it will take time to diagnose this. Neurologists can diagnose stroke accurately which is a nice part. main problem is that the process for diagnosis is very time consuming so that due to this patients can fall under brain hemorrhage or there will be blood clot in brain due to this our body gets affected and one by one body part gets damaged.

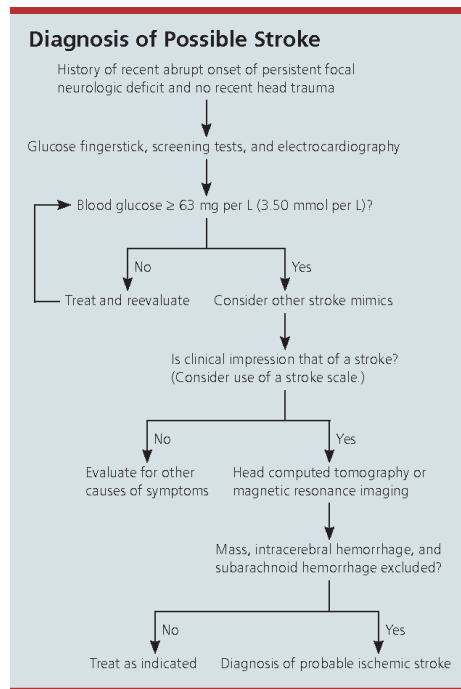


Figure 1. Algorithm for the diagnosis of acute stroke.

It is clear that people like fast service. Some small hospitals don't have facilities like CT & MRI, due to this patients are transferred to a centre where CT Scan & MRI is available. This is a very time-consuming process. With the advent of internet technology, there is vast data which we don't know. Machines can do anything. They can work faster than human so that we get good results very quickly.

In this article, I use an Artificial Intelligence for the prediction of brain stroke in very less time. For this, we use a machine learning algorithm, which is a branch of AI. Machine Learning plays a very important role in this article. In this system, we can take data, find the algorithms, patterns, and output by using AI.

4. Target Specifications and Characterization:

4.1 Reducing frustration and death of patients due to delay in the diagnosis process.

- 4.2 Predetermined brain stroke prediction and MRI of Stroke patients are used for the prediction.
- 4.3 To get a accurate results in fast and accurate process.

5. External Search:

5.1 Stroke Prediction using Artificial Intelligence

The stroke deprives person's brain of oxygen and nutrients, which can cause brain cells to die. Numerous works have been carried out for predicting various diseases by comparing the performance of predictive data mining technologies. In this work, we compare different methods with our approach for stroke prediction on the Cardiovascular Health Study (CHS) dataset. Here, decision tree algorithm is used for feature selection process, principle component analysis algorithm is used for reducing the dimension and adopted back propagation neural network classification algorithm, to construct a classification model. The proposed method use Decision Tree algorithm for feature selection method, PCA for dimension reduction and ANN for the classification. The experimental results show that the proposed method has higher performance than other related well-known methods.

5.2 Burden of Stroke in the World

Stroke is the second leading cause of death and leading cause of adult disability worldwide with 400-800 strokes per 100,000, 15 million new acute strokes every year, 28,500,000 disability adjusted life-years and 28-30-day case fatality ranging from 17% to 35%. The burden of stroke will likely worsen with stroke and heart disease related deaths projected to increase to five million in 2020, compared to three million in 1998. This will be a result of continuing health and demographic transition resulting in increase in vascular disease risk factors and population of the elderly. Developing countries account for 85% of the global deaths from stroke. The social and economic consequences of stroke are substantial. The cost of stroke for the year 2002 was estimated to be as high as \$49.4 billion in the United States of America (USA), while costs after discharge were estimated to amount to 2.9 billion Euros in France.

5.3 Stroke Prediction using Machine Learning:

Recently, machine learning algorithms have shown great promise in accurately predicting stroke risk based on various The associate editor coordinating the review of this manuscript and approving it for publication was Mounim A. El Yacoubi . clinical risk

factors. By leveraging these algorithms, clinicians can identify high-risk patients and intervene early, potentially reducing the number of stroke-related complications and improving patient outcomes. Additionally, there is a growing need for transparency and explainability in machine learning models in healthcare. The use of an interpretable machine learning model can provide clinicians with valuable insights into the factors that contribute to a patient's stroke risk, thereby aiding in treatment decisions.

5.4 Dataset

Important Columns in Dataset are

- A. Id of individual gender.
- B. Age (in years)
- C. Hypertension
- D. Heart diseases
- E. Ever Married
- F. Work Type
- G. Residence Type
- H. Average Glucose Level
- I. BMI
- J. Smoking Status
- K. Stroke

With the help of these attributes we can perform a stroke prediction analysis using AI.

6. Benchmarking of alternative product

7. Applicable Regulations (Government and Environmental)

- 7.1 Machine Learning Algorithms.
- 7.2 Law related to data for privacy concern.
- 7.3 Creating a service where patients get their report by whatsapp or through other source.
- 7.4 Protection.
- 7.5 Use of this can be feasible.

8. Applicable Constraints

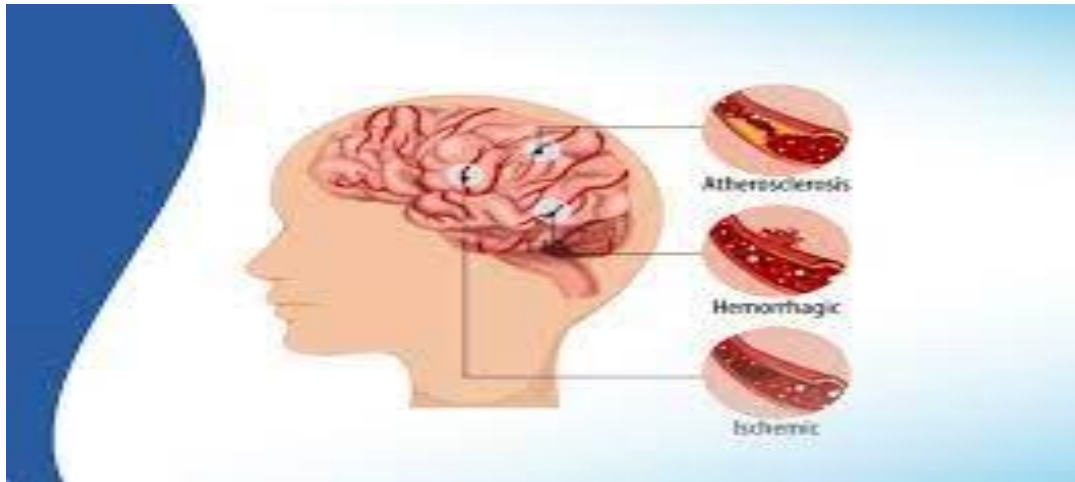
- 8.1 Require a lot of research on a brain stroke datasets for performing machine learning techniques.
- 8.2 Required a detail study of classification algorithms and Deep Learning techniques.
- 8.3 Giving result of this prediction through message system.
- 8.4 Confidential Health Data.
- 8.5 By Studying a dataset perform a machine learning algorithms for prediction of brain stroke in less time.

9. Business Opportunity

Doctors can predict a brain stroke by analyzing a image based reports like MRI and CT Scan. Due to complex structure of brain it is difficult to diagnose brain stroke immediately. Sometimes hospitals don't have facilities like MRI & CT, so patients want to go other lab for doing CT & MRI so it takes a time. In brain diseases problems it is very important to diagnose stroke immediately and give a treatment to patients. So to overcome this hazardous situation we use machine learning algorithms for the prediction of brain stroke.

10. Concept Generation

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 than the existing system. In proposed system, the practical use of various collected data has turned out to be less time consuming.

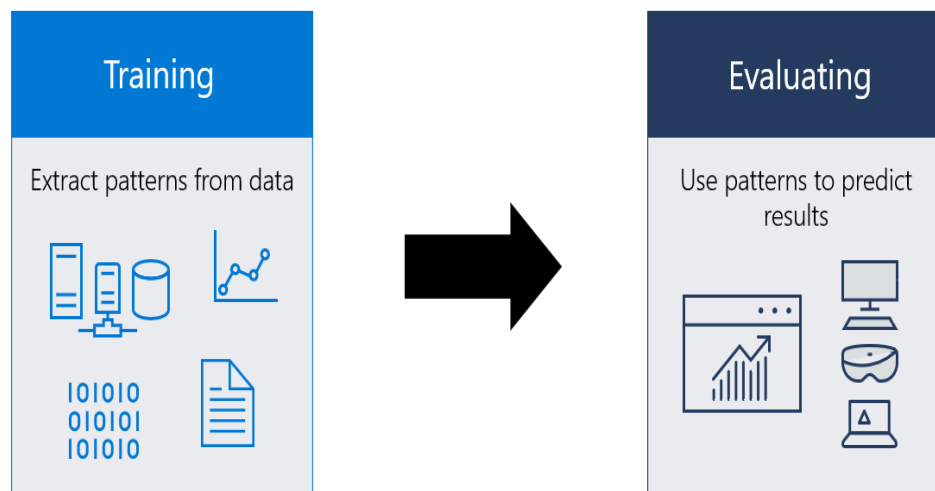


A. Stroke Risk Diagnosed:

Through our Web Application, the user will get to know about the outcome of its input data. In the case for “Stroke” as an outcome, it will be displayed as “Stroke Risk Diagnosed.

B. No Stroke Risk Diagnosed:

Through our Web Application, the user will get to know about the outcome of its input data. The outcome for “No Stroke” will be displayed as “No Stroke Risk Diagnosed”.



So in order to generate this we have to use artificial intelligence for the prediction of brain stroke. For prediction we use a machine learning and deep learning algorithms

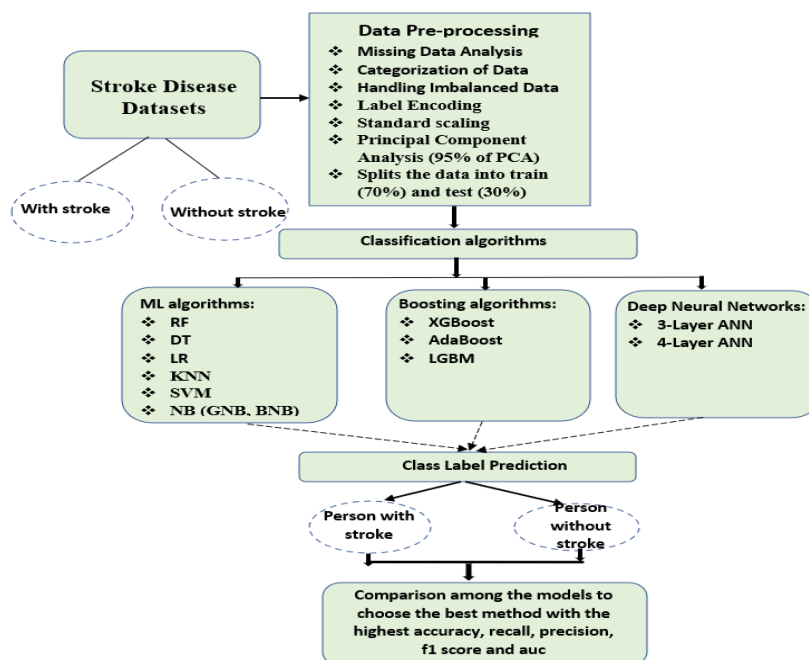
which is a part of artificial intelligence. Machine learning build the model on training data in order to make decision in less time.

10. Concept Development

Machine Learning algorithms are trained in python jupyter notebook. Firstly dataset is collected and then some EDA done on that. Some important features are considered by feature selection process for model making. This dataset is imported into the spyder environment. Python contains some in-built libraries such as pandas, scikit learn, Seaborn which is exclusively used for training machine learning algorithms. Machine Learning and deep learning methods are done on dataset and prediction can be done and you get a report on your device.

Machine learning not only predicts stroke results faster but also gives higher accuracy which is around 80 to 85 % which is greater when compared with the Neurologists. Stroke one such kind of disease which when treated early would save many lives. So machine learning is most sort after technique which is very useful in replacing present process.

11. Final Product prototype:



12. Product Details:

With the help of artificial intelligence, machine learning, and data science techniques, significant advancements have been made in the realm of clinical and medical services. The foundation of the present period is machine learning, which is utilized to anticipate numerous issues at an earlier stage. As a serious disease that may be treated if anticipated in the early stages, stroke is one of many that can be prevented if predicted early. In the health care sector, machine learning is crucial for the diagnosis and prognosis of diseases. Currently, stroke incidence is predicted using machine learning algorithms.

12.1 Machine Learning Algorithms

12.1.1 Decision Tree (DT)

Classification with DT addresses both regression and classification issues. This approach uses a supervised learning model and an output variable. It has tree-like characteristics. The decision tree's two components are the decision node and the leaf node. The data is split at the first node, then it is combined at the second node that generates the output. Since DT mimics the stages, a human goes through when creating a real-world object, it is simple to understand. The algorithm assumes that the existence or absence of a given feature in the dataset depends on others and helps to categorize the target towards a particular class.

12.1.2 Random Forest (RF)

Classification and regression problems can be resolved using ensemble learning techniques like Random Forests (sometimes referred to as random choice forests). They work through distributed training of a large number of decision trees. When attempting to solve classification problems, a random forest's output is the class that majority of the trees select. In order to tackle complex issues and improve the performance of a model, ensemble learning, which is a technique, employs several classifiers. The Random Forest classifier takes into account predictions from all of the trees rather than just one, as stated in its name, "combining a large number of decision trees on different subsets of a given dataset and taking an average to boost the projected accuracy". To determine the outcome, the most popular forecasts are used. The Random Forest demonstrated the highest accuracy (criterion=entropy) with the utilized Stroke prediction dataset.

12.1.3 Naive Bayes (NB)

The supervised learning technique known as naive Bayes, which is based on the naïve theorem, is used in machine learning (ML). The Naive Bayes algorithm is founded on the notion that the presence of one feature or parameter does not preclude the presence of another, i.e., that one feature's existence is unrelated to that of other features. The Bayes theorem is a conditional probability theorem in mathematics that estimates the chance that a certain event will take place

assuming that a particular condition has already been satisfied. With the use of the Bayes theorem, conditional probability is contrasted; it is the likelihood that a particular event has occurred, given the premise that some event has already occurred. There are several variations of NB found in the literature [19], with the key distinction being how the likelihood of the intended class is calculated. These variations include simple Naive Bayes, Gaussian Naive Bayes, Multinomial Naive Bayes, Bernoulli Naive Bayes, and Multivariate Poisson Naive Bayes. Gaussian Naive Bayes and Bernoulli Naive Bayes were used for this study.

12.1.4 K-Nearest Neighbors (KNN)

K-NN is a kind of slow learning in which there is no specific pre-processing stage and all computations are saved for classification. The nearest training data points on the feature map are used to make decisions using this data categorization technique. The Euclidean distance metric is used by the K-NN classifier to predict the target class. The dataset defines the best value for the parameter k , which controls how well the classifier performs. The ideal value is then determined once the impacts have been studied. $K = 3$ was applied in our investigation.

12.1.5 Support Vector Machine (SVM)

The Support Vector Machine is a type of supervised learning system that uses labeled data to categorize unknown data. To express decision boundaries, it uses the concept of decision planes or hyperplanes. A hyperplane is used to separate the collection of data objects into the various classifications. The Radial Basis Function (RBF) kernel, with a setting of 1, was applied. SVM attempts to classify the data by creating a function that assigns each data point to its appropriate label with the minimum amount of error and the largest (possible) margin.

12.1.6 Logistic Regression

(LR) LR is one of the most widely used ML algorithms in the supervised learning approach. It is a forecasting technique that forecasts a categorical dependent variable using a number of independent factors. Logistic regression and linear regression are fairly similar, with the exception of how they are used. While logistic regression is used to address classification issues, regression issues are addressed by linear regression. A model tuning method called ridge regression can be used to analyze any multi-collinear data. This method performs L2 regularization. For this task, we utilized `solver='lib linear'` and `max iter=100`.

12.2 Deep Learning Approaches

Deep learning is a term used to describe artificial neural networks used in machine learning. Examples of deep learning architectures include convolutional neural networks, deep belief networks, recurrent neural networks, and deep neural networks (DNN). A group of algorithms known as neural networks are created to recognize patterns and are roughly modeled after the

human brain. These are widely used in many different research areas, including as speech recognition, natural language processing, audio recognition, computer vision, gaming, and many more. A DNN is made up of an input layer, many hidden layers, and an output layer. Back propagation is used to train the network and reduce the difference between the desired and actual output. Two ANN models have been put into practice for the analysis. The output of the two models, three- and four-layer ANNs has a sigmoid function. The data collection and preprocessing are the same as for machine learning approaches.

12.3 Team required to develop:

1. Data Analyst
2. Machine Learning Engineer
3. Cloud Engineer
4. Test Engineer
5. Software Engineer

Conclusion

Stroke is a potentially fatal medical condition that needs to be treated right away to prevent future consequences. We 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 Machine Learning algorithm and Deep Learning Techniques given a highest accuracy.

