



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Machine Learning Approaches to Predict Stroke Disease-A Review

Dr. Harish B.G* , Noorul Huda Khanum

Department of Master of Computer Applications, University BDT College of Engineering, Visvesvaraya Technological University, Davanagere. Karnataka, India

ABSTRACT: Machine learning is a part of artificial intelligence that make software application to gain the exact accuracy to predict the end result not having to be directly involved to get the work done. This article or review is about prediction of stroke, and the type of stroke the individual is suffering. The application of machine learning is used to detect many clinical diseases for rapid diagnosis. For the prediction of stroke there are commercially available machine learning algorithms. There are new algorithms which are emerging to have accuracy in diagnosis of stroke prediction or any other clinical diseases. For diagnosis of any clinical diseases the data is collected from studying physiological data, environmental influences and genetic factors allow practitioners to diagnose diseases early and effectively. The prediction of disease helps the doctors and clinical researchers reach out to potential participants for the trial. That's because machine learning technology can access hospital databases of patients EHRs. Then they find people who fit the conditions for a particular trial.

I. INTRODUCTION

Health is essential aspect for every living being in a world. Stroke is a threatening disease for every human regardless of race, gender etc. Many people fall victim to stroke and the number are increasing more in the developing countries. Stroke is a leading cause of disabilities in adults and the elderly which can result in numerous social or economic difficulties. Stroke is the subsequent driving reason for death and was a serious, long haul inability. Stroke is the unexpected passing of cerebrum cells because of oxygen misfortune brought about by hindering the dissemination or wrecking a flexible way to the brain. Blood vessels carries oxygen and nutrients to the brain when stroke occurs the blood vessel is either blocked by a clot or burst. When the clot or burst occur, part of the brain cannot get the blood needed, so blood cells dies.

In 2019 there were 12.2 million incident cases of stroke and 6.55 million deaths from stroke. This occurs when a blood vessel in the brain ruptures and bleeds, the rupture or blockage prevent the reaching of blood and oxygen to the brain's tissue. According to the World Health Organization, 15 million people suffer stroke worldwide each year, 5 million die and another 5 million are permanently disabled. High blood pressure contributes to more than 12.7 million strokes worldwide. In developed countries, the incidence of stroke is declining, largely due to efforts to lower blood pressure and reduce smoking. However, the overall rate of stroke remains high due to the aging of the population. In 2013 approximately 6.9 million people had an ischemic stroke and 3.4 million people had a hemorrhagic stroke. In 2015 there were about 42.4 million people who had previously had a stroke and were still alive. Between 1990 and 2010 the number of strokes which occurred each year decreased by approximately 10% in the developed world and increased by 10% in the developing world. In 2015, stroke was the second most frequent cause of death after coronary artery disease, accounting for 6.3 million deaths. About 3.0 million deaths resulted from ischemic stroke while 3.3 million deaths resulted from hemorrhagic stroke. About half of people who have had a stroke live less than one year. Overall, two thirds of strokes occurred in those over 65 years old.

Europe averages approximately 650,000 stroke death each year. Stroke is the fifth leading cause of death in the United States and a significant cause of severe disability in adults. Each year, around 800,000 American experience a new or recurrent stroke.

World Health Organization indicated that the stroke will keep on increasing death rate in coming years. So, earnest condition treatment must be quick as could be expected under the circumstances. Consistently a million individuals overall suffer from stroke. Inability incorporates confront deformity, loss of vision, loss of motion, and discourse. Stroke is a significant unsafe ailment which hurts the cerebrum, like heart strike which hurts the heart. It doesn't circle blood and enough oxygen to the brain cells. Everyone overcomes with some stroke peril. The stroke may cause loss of movement, sudden torment in the chest, talk inability, loss of memory and thinking limit, daze like state, or passing.

Stroke impacts the person of all ages. It can be balanced through the helpful control and modifiable risk factors are essential. The report says the most surely understood remedial error happens in light of expiry of meds, off kilter medicines, misguided estimations and treatment given to the wrong patient. Stroke for the most part three composes:

- Ischemic stroke
- Hemorrhagic stroke
- Transient ischemic stroke

A. Ischemic stroke

Ischemic stroke is the most common type. It is usually caused by a blood clot that blocks or plugs a blood vessel in the brain. This keeps blood from following to the brain. Within minutes, brain cells begin to die. Ischemic stroke is a third biggest purpose behind the death.

B. Hemorrhagic stroke

Hemorrhagic strokes make up about 13% of stroke cases. It is caused by a weakened vessel that ruptures and bleeds into the surrounding brain. In this case, blood puts excessive weight on the cerebrum and harm the brain cells.

C. Transient Ischemic Attack

This sort of stroke is called smaller than mini-stroke. Transient Ischemic Attack is not same as another stroke composes. The blood streams to the cerebrum, so it is blocked just for the brief time, not more than 10 minutes. Diagnosing and taking great treatment transient ischemic assault can diminish the fundamental stroke.

Ischemic stroke is most common in stroke. 87% are Ischemic stroke (IS), 10-15% are Transient Ischemic stroke (TIS), which is named as 'mini -stroke' .TIA caused by temporary blockage and symptoms will disappear within 23 hours. TIA is a warning sign of a future stroke. Hemorrhagic stroke occurs when an artery in the brain leaks blood.

Global prevalence of stroke in 2019 was 101.5 million people, whereas that of ischemic stroke was 77.2 million, that of intracerebral hemorrhage was 20.7 million, and that of subarachnoid hemorrhage was 8.4 million.

II. MACHINE LEARNING IN STROKE PREDICTION

Machine learning is sub-field of Artificial Intelligence (AI). Now a day's machine learning is a very important. Machine learning gives the ability to learn and improve from experience without being explicitly programmed. Machine learning is used to make a system learn from data identifying patterns and making decisions with minimal human intervention. In healthcare machine learning is implemented that benefits the doctors, where computers are programmed to learn information without human intervention. Machine learning uses programmed algorithms that receive and analyze input data to predict output values within an acceptable range. Predicting disease in healthcare using machine learning needs powerful data tools which are needed to extract useful information from the huge amount of data, where patient's profile like name, age, and blood sugar and blood pressure, etc.

Machine learning uses K-Nearest Neighbor (KNN) and Convolutional neural network (CNN) algorithms for accurate prediction of disease. For disease prediction we require the symptoms dataset. Along with the above mentioned algorithms machine learning uses support vector machine (SVM), random forest methods which have been successfully applied in disease prediction based on clinical data. Machine learning is changing the healthcare sector by collecting patient data such as past medical records, psychological behavior, family history, this makes more easy to predict the disease of patient.

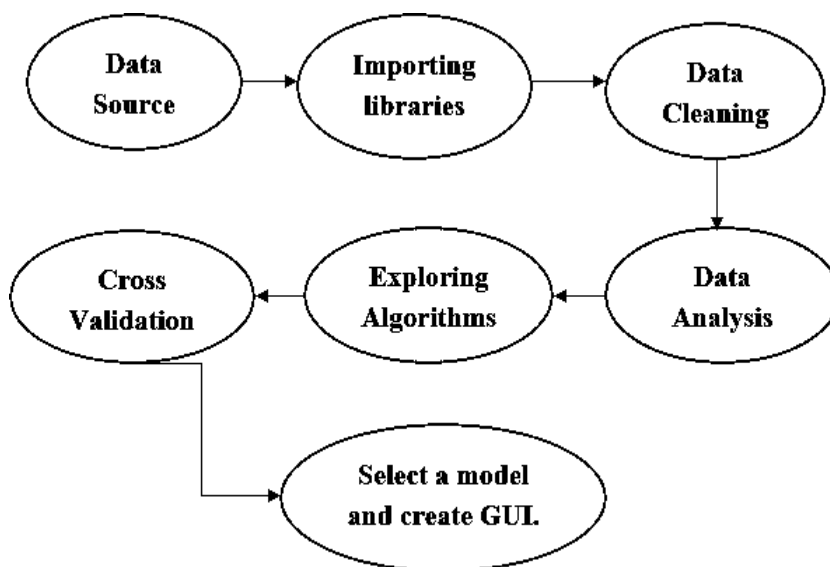
III. LITERATURE SURVEY

Geethanjali et al, In their paper, stroke attack can be predicted accurately. They have used three classifiers such as logistics regression, Decision tree classifier and support vector classifier for the prediction of stroke. The classification model is based on 5110 records. And they have provided the result with 95.49% of accuracy. **Veena potdaret et al**, In their paper they have described different methods of prediction of stroke and concluded that each technique for prediction of stroke has its own advantages and disadvantages. One must perform statistical analysis and initialization to decide on the specific technique to use. However, they suggested random forest technique is considered as one of the most precise technique for analyzing which shows promising result. **Shraddha Mainali et al**, In their paper they have described the stroke diagnosis and outcome prediction using supervised and unsupervised machine learning types. **Ankitha et al**, In their paper they have used machine learning or AI intelligence algorithm to predict the type of stroke a patient is suffering through an application where admin, user and interaction with particular doctor come into play. And finally gives the result for the user which type of stroke he is suffering. **Vida Abediet et al**, In their paper they

have used artificial intelligence for improving stroke diagnosis in emergency department. They have mentioned the key step for stroke ML-enabled decision support system for EDs. **Harshitha et al**, In this paper, they have constructed a model for predicting stroke using machine learning algorithms and given the accuracy for each model.

IV. METHODOLOGY

Fig 1: Methods of stroke prediction using Machine Learning Algorithms.



1.1 Data source:

- Primary Data:
The data, reference materials and features/attributes collected by consulting neurologists and cardiologists.
- Secondary Data:
The data is collected from Kaggle website. “Healthcare - dataset-stroke data” which is publicly available.

1.2 Libraries Imported:

Numpy, Pandas, MatplotlibPyplot, Seaborn, Tkinter are the libraries which can be imported.

1.3 Data cleaning:

Data was cleaned for missing data and null values. Missing data was dealt by removing the rows with null values or redundant values.

1.4 Data Analysis:

There are three types of data analysis which is performed i.e., Categorical feature analysis, Numerical feature analysis and Multicollinearity analysis. Data analysis is done to show us the hidden relationships and attributes present in the dataset which help the machine learning model to perform better.

1.5 Implementing Algorithms:

There are eight different algorithms for stroke prediction where one or more algorithms are implemented for diagnosis of different types of stroke.

1.6 Cross Validation:

Effectiveness of all the models is verified to solve overfitting problems. Overall assessment on how the model will perform for an independent test dataset.

Finally, the best performing model will be used to predict stroke using the input data given by the user.



1.7 Creating GUI:

The graphical user interface is a user interface which permits users to interact with digital gadgets through graphics and audio indicator, as opposed to textual content, mostly based on user interfaces, written command labels or textual content. In response to an impressive curve of command-line interfaces that requires directions to be typed on a pc-keyboard, GUIs were provided.

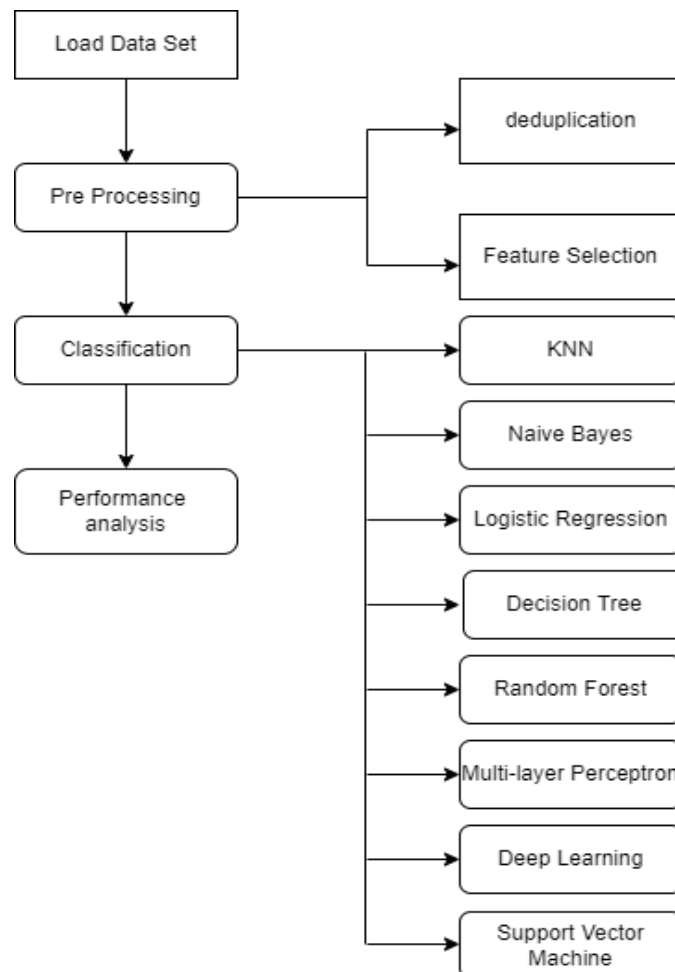
1.8 Software Specification:

The programming language which is used is Python. Python is an extensively helpful translated, intuitive, object-orchestrated, and irrefutable level programming language. Python is used for web improvement, machine learning, AI, working systems, compact application headway, and computer games. Like Perl, Python source code is in like manner available under the GNU General Public License (GPL). This educational exercise gives adequate understanding on Python programming language. Python is expected to be extraordinarily comprehensible. It uses English expressions routinely where as various tongues use highlight, and it has less semantic improvements than variouslingos.

Jupyter notebook:

In the past it was referred as IPython Notebooks. It is an online intelligent computational platform for making Jupyter journal records. The "notebook" term can conversationally make reference to various elements, primarily the Jupyter web application, Jupyter Python web server, or Jupyter report design contingent upon setting.

Data flow diagram for Prediction of stroke using machine learning:



- **Load Data set** in which all parameters of stroke disease is available.
- **Pre Processing** The data preprocessing in machine learning refers to the technique of preparing i.e., cleaning and organizing the raw data to make it suitable for building and training machine learning models.
- **Classification** is based on the type of one or more algorithms. This classification has 8 algorithms which can be used for the prediction of stroke disease.
- **Performance analysis** is according to the data entered to the parameters of stroke by the user to predict the stroke disease, through machine learning approaches and algorithms the performance is analyzed.

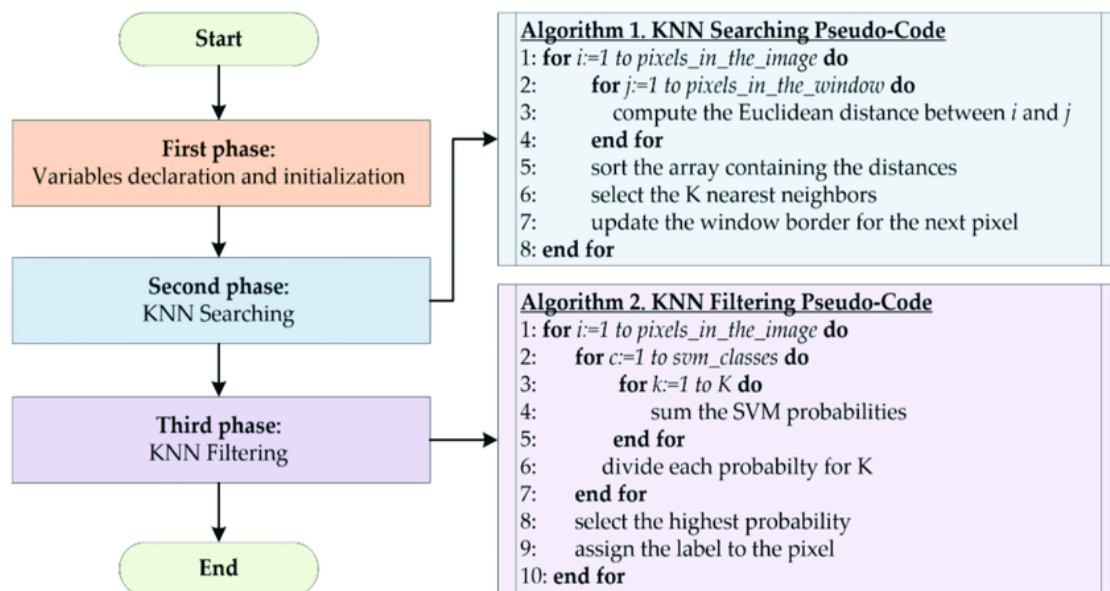
Machine Learning Algorithms used in Stroke Prediction:

Eight machine learning algorithms used in stroke prediction are:

- K-Nearest Neighbors
- Naive Bayes
- Logistic Regression
- Decision Tree
- Random Forest
- Multi-layer perceptron(MLP-NN)
- Deep Learning
- Support Vector Machine

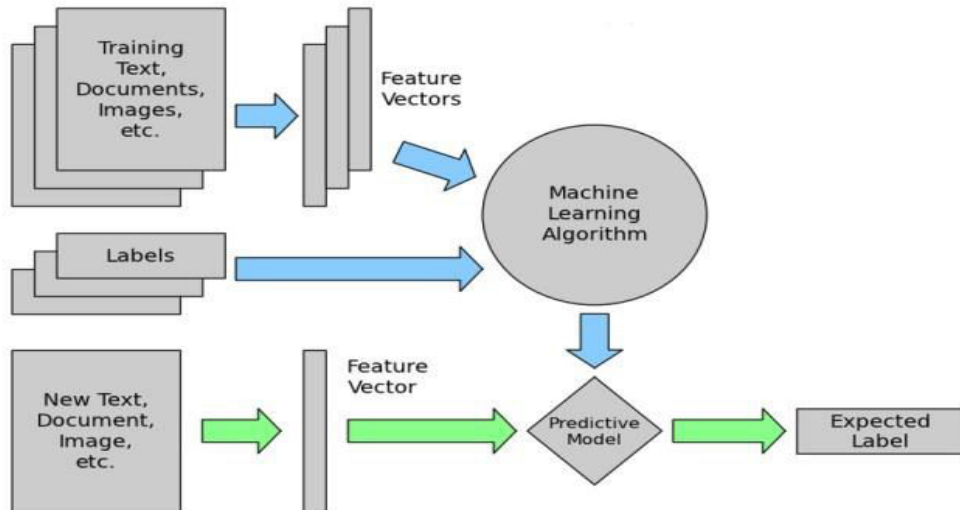
1. K-Nearest Neighbors:

This is a simple algorithm that stores all the available cases and classifies the new data or case based on a similarity measures. This algorithm is based Supervised Learning technique.



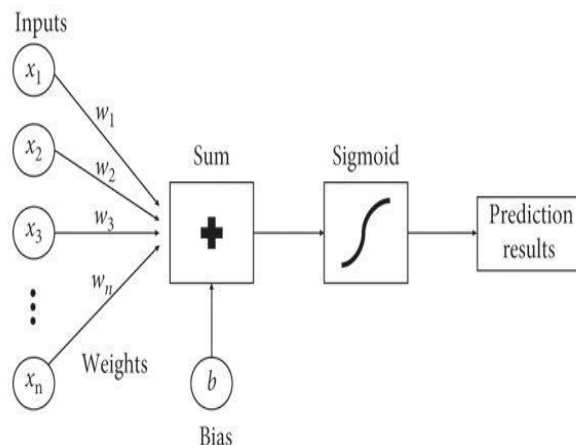
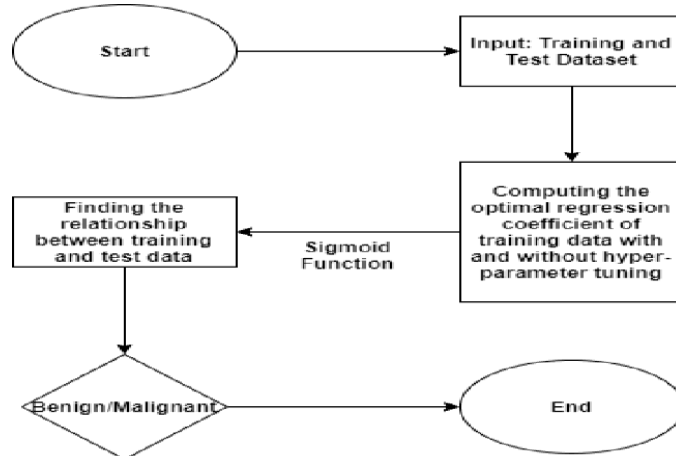
2. Naive Bayes:

This algorithm uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems having multiple classes.



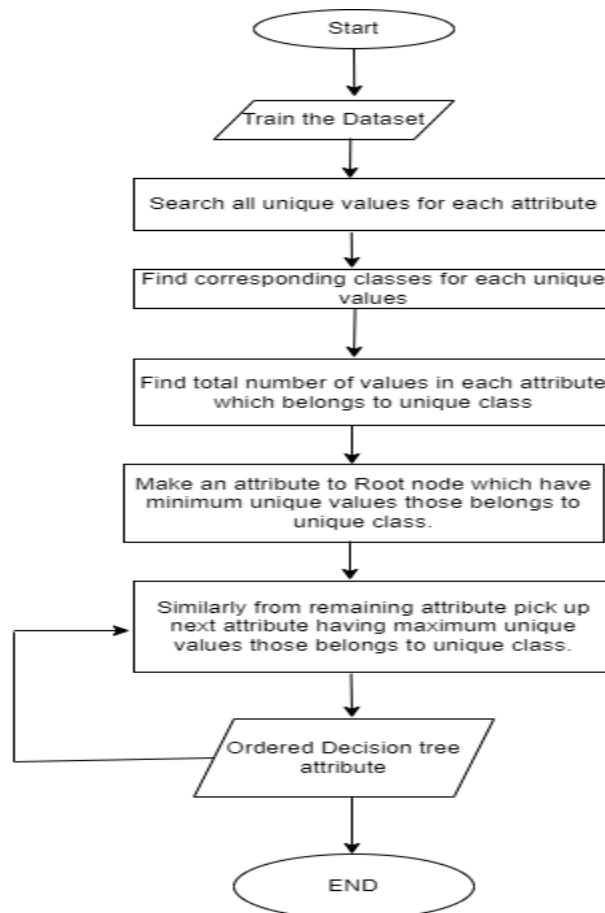
3. Logistic Regression:

This algorithm is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. This algorithm is based on Supervised Learning technique.



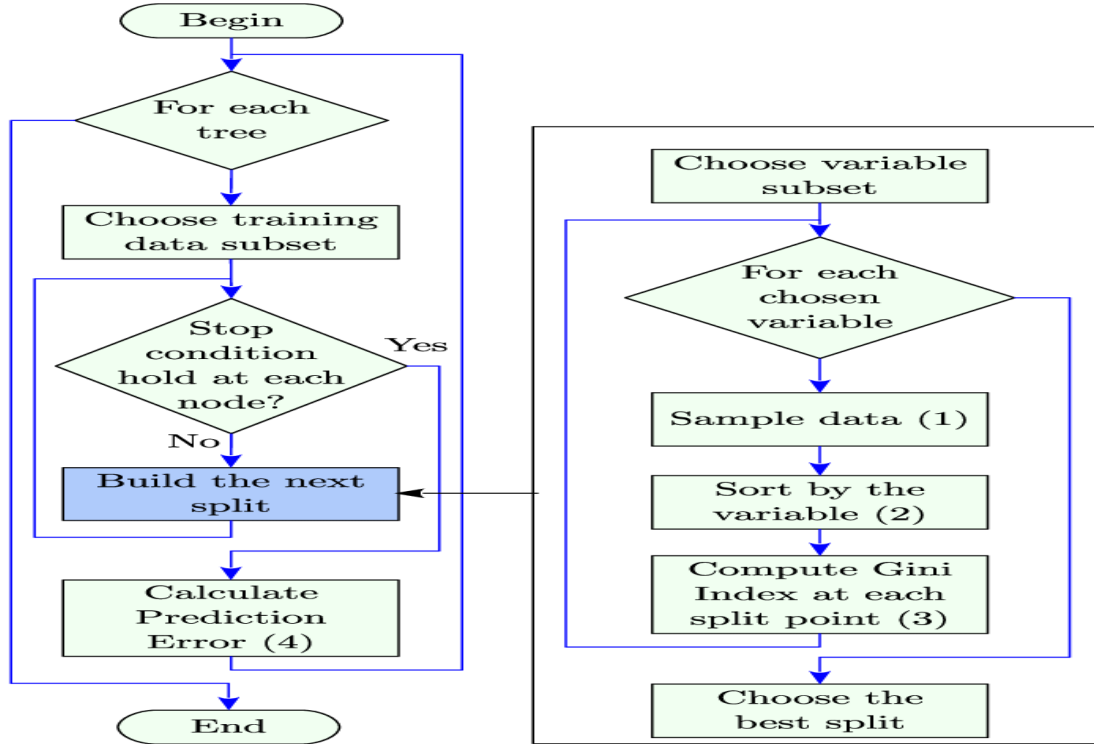
4. Decision Tree:

This algorithm is used to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data).



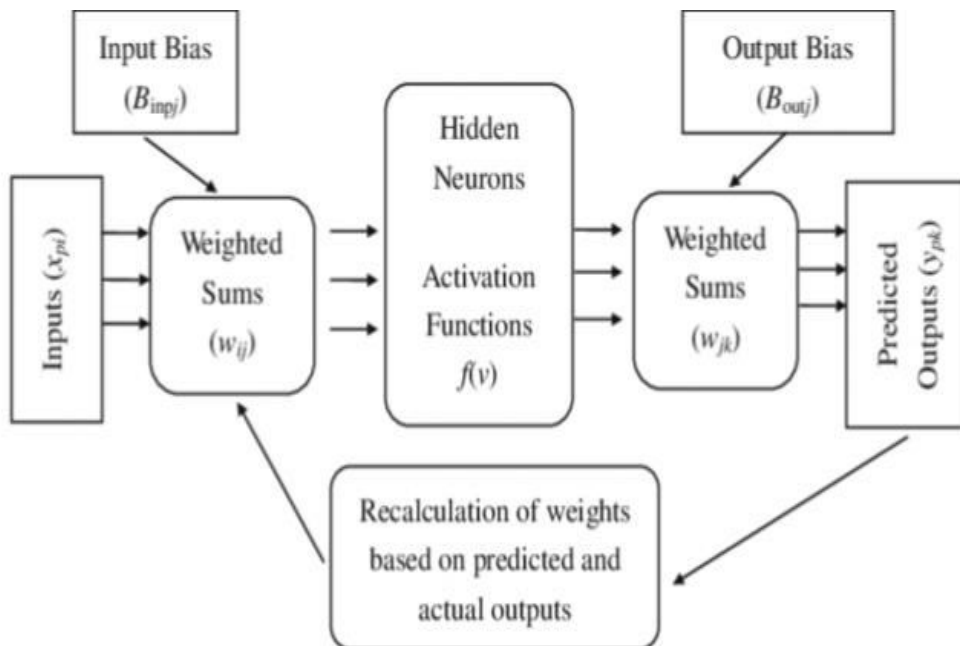
5. Random Forest:

This algorithm is based on supervised learning technique is used widely in classification and regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.



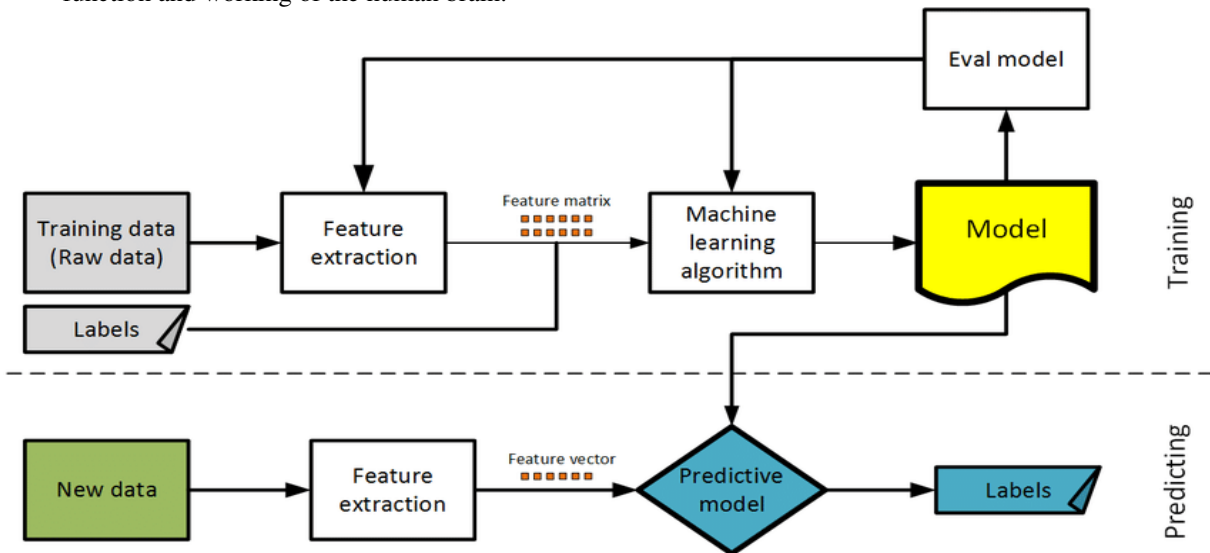
6. Multi-layer perceptron(MLP-NN):

This algorithm is a supervised learning algorithm, it is a feedforward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses backpropagation for training the network.



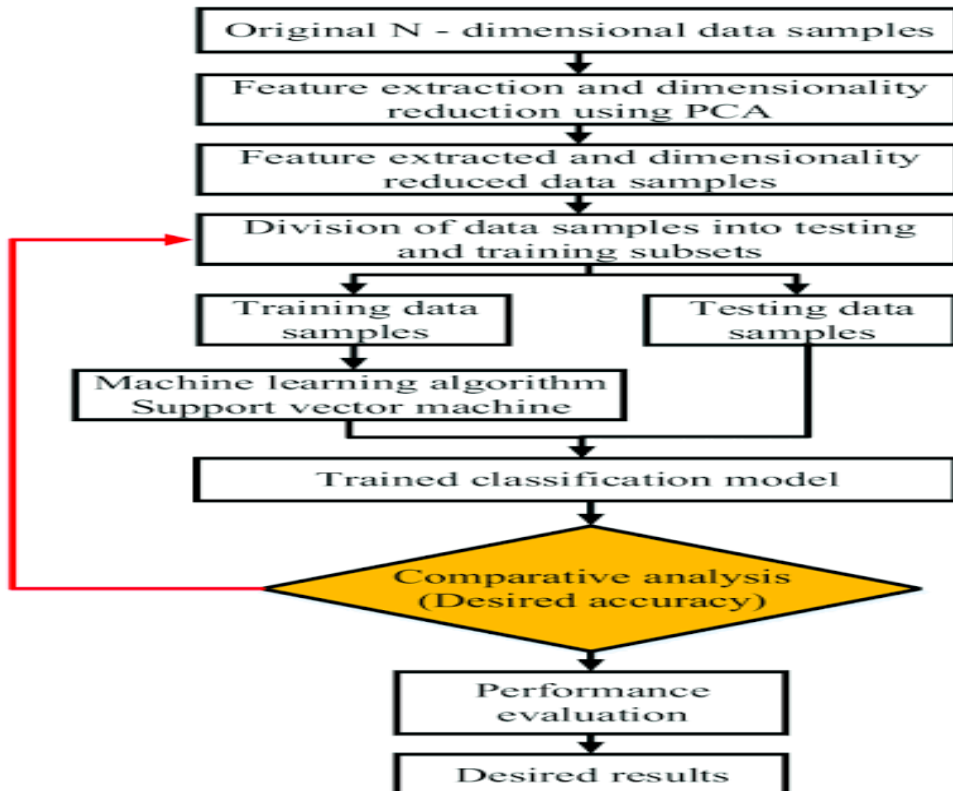
7. Deep Learning:

This algorithm is a subset of a Machine Learning algorithm that uses multiple layers of neural networks to perform in processing data and computation on a large amount of data. This algorithm works based on the function and working of the human brain.



8. Support Vector Machine:

This algorithm is a supervised machine learning algorithm used for both classification and regression. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points.



V. CONCLUSION

This review paper has information about machine learning approaches, types of stroke disease, machine learning used for prediction of stroke all this info is collected after reviewing many review and research papers. This paper also contains the types of algorithms and technique used in prediction of type of stroke and has the methodology for prediction of stroke using machine learning.

REFERENCES

1. VeenaPotdar, LavanyaSanthosg, Yashu Raj Gowda CY, A survey on stroke disease classification and prediction using machine learning algorithms,IJERT,Vol. 10, 2021.
2. T.M,Geethanjali, Divyashree M.D, Monisha S.K, Sahana M.K, Stroke prediction using machine learning, JETIR,2349-5162,2014.
3. Ankitha S,Jhanvi V, Harshavardhan N, Deepthi M, Remanth M, An artificial approach for predicting different types of stroke, IJert,ISSN:2278-0181,Vol. 8 Issue 12, 2019.
4. Shraddha Mainali, Marin E. Darsie and Keaton S. Smetana, frontiers in Neurology (fneur), 2021.
5. Vida Abedi, Ayesha khan, Using AI for improving stroke diagnosis in emergency departments: a practical framework, The AdvNeurolDisord. Vol. 13, 2020.
6. Harshitha K V, Harshitha P, Stroke prediction using machine learning algorithms, IJIREM, ISSN:2350-0557,Volume-8, Issue-4, 2021.
7. GiordanaFlorimbi et al,accelerating the K-NN Filtering Algorithm to optimize the real time classification of human brain tumours in hyperspectral images, DOI:10.3390/s1872314
8. Charles Cooper, Basic concepts and methods of machine learning, HBBL meeting February 20th, 2014.
9. Taki Hasan Rafi*, A comparative analysis of hyper-parameter tuned supervised machine learning algorithm on breast cancer prediction.
10. Mohammad Monirujjaman khan et al, A comparative analysis of machine learning algorithm to predict Alzheimer's Disease, Journal of Healthcare Engineering 2021(7) DOI:10.1155/2021/9917919
11. N.M Patel, Divakar Singh, An algorithm to construct Decision Tree for machine learning based on similar factor, DOI:10.5120/19575-1376.
12. Nileshkunjare et al, particle swarm optimization and feature selection for intrusion detection system, DOI:10.1007/s12046-020-1308-5
13. Yuan Y Liu et al, A comparision of logistics regression,classification and regression tree, and Neural networks models in predicting violent Re-Offending, DOI:10.1007/s10940-011-9137-7
14. Dong Nguyen et al, Joint network coding and machine learning for error-prone wireless broadcast, DOI:10.1109/CCWC.2017.77868415
15. RaheelMuzzammel, Ali Raza, A support vector machine learning-based protection technique for MT-HUDC systems, Energies 2020,13(24),6668, DOI:10.3390/en3246668



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

doi[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details