e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

STROKE PREDICTION SYSTEM BY USING MACHINE LEARNING ALGORITHM

Aishwarya Raj*[1], Daniel Jebaraj. P*[2], Divagar. M*[3], Pradeep Kumar. K.*[4], Sivanesan. V*[5]

*[1]Assistant Professor, Biomedical Department, Agni College Of Technology, Chennai, Tamil Nadu,India.

*[2,3,4,5]Student, Biomedical Department, Agni College Of Technology, Chennai, Tamil Nadu, India.

DOI : https://www.doi.org/10.56726/IRJMETS52009

ABSTRACT

A stroke is a medical condition that causes damage to the blood vessels in the brain. Its severe condition can lead to death and if treated on time can save the lives of the patients. The strokes are basically of two types – ischemic, and hemorrhagic.The objective of the study is to construct a prediction model for predicting stroke to assess the accuracy of the model and manually detect the glucose level and heart rate using the respective sensors. The project mainly aims at predicting the chances of the occurrence of stroke using emerging machine-learning techniques with various Datasets. To accomplish these four main types of algorithms used according to statistics about 185000 strokes occur every year in India, so necessary conditions must be taken.

Keywords: Glucose Level, Heart Rate, Machine Learning Algorithm, Stroke Prediction.

I. INTRODUCTION

A stroke is a blood clot or bleeding in the brain that can cause permanent damage that affects mobility cognition sight or communication. A stroke could be a medical condition in which poor blood flow to the brain causes cell death. There are two main kinds of stroke: ischemic, because of lack of blood flow, and hemorrhagic, where blood spill directly into the brain. Both cause parts of the brain problems understanding or speaking, dizziness, or loss of vision to at least one side. Signs and symptoms often appear soon after the stroke has occurred. Two principal types of stroke are CVA and hemorrhage. Ischemic stroke happens due to the absence of a blood stream and hemorrhagic stroke happens due to bleeding. Hemorrhage is classified into two kinds' subarachnoid hemorrhage and intracerebral hemorrhage.

Transient ischemic assault is otherwise called "ministroke". Stroke denies an individual's mind oxygen and supplements, which ends up in the death of dead cells when a stroke occurs. It's not only very expensive for medical treatments and a permanent disability but can ultimately prompt demise. A machine learning model would take the patient's information andpropose a bunch of suitable Expectations. The framework can remove concealed information from a chronicled clinical data set and may anticipate patients with infection and utilize the clinical profiles like Age, force per unit area, Glucose, and then forth it can foresee the probability of patients getting an illness. Grouping calculations are utilized with the amount of properties for the expectation of illness. Prevention of stroke - over 70% of strokes are first events, thus making primary stroke prevention a very important aspect. Interventions should be targeted at behavior therapy, which however requires information about the baseline perceptions, knowledge, and prevalence of risk factors in defined populations. The goal of our project is to use principles of machine learning over large existing data sets to effectively predict the stroke-supported potentially modifiable risk factors. Then it intended to develop the appliance to produce a personalized warning on the premise of every user's level of stroke risk.

II. EXISTING SYSTEM

The development of a prediction method to assess high stroke risk while handling incomplete data involves several key steps. Initially, a comprehensive datasets comprising relevant medical, demographic, and lifestyle factors is collected and preprocessed to address missing values and outliers. Feature selection techniques are then applied to identify significant risk factors. Machine learning or statistical models like logistic regression or random forests are trained using labeled data to learn patterns and relationships for stroke risk prediction addition to the technical aspects, ensuring the prediction method's interpretability and explainability is crucial in medical contexts for gaining trust and facilitating adoption by healthcare professionals. Transparent models and clear explanations of predictions, such as using feature importance rankings and understandable visualizations, help clinicians understand the underlying reasoning behind high stroke risk predictions. Continuous monitoring and validation processes not only maintain the model's accuracy over time but also allow for updates and improvements as new data and insights which emerges. Collaborative efforts between data scientists, clinicians, and domain experts ensure that the prediction method aligns with clinical needs, enhances patient care, and contributes meaningfully to stroke risk assessment and management strategies in healthcare settings. The theory of belief functions also referred to as evidence theory or Dempster–Shafer theory (DST), is a general framework for reasoning with uncertainty, with understood connections to other frameworks such as probability, possibility, and imprecise probability theories. First introduced by Arthur P. Dempster in the context of statistical inference, the theory was later developed by Glenn Shafer into a general framework for modeling epistemic uncertainty—a mathematical theory of evidence. Thetheory allows one to combine evidence from different sources and arrive at a degree of belief (represented by a mathematical object called belief function) that takes into account all the available evidence.

III. PROPOSED SYSTEM

Our project is multifaceted, aiming to revolutionize healthcare by applying machine learning principles to predict strokes based on modifiable risk factors and detect heart abnormalities using heart rate sensor data and Glucose sensors data. Firstly, we collect a comprehensive datasets encompassing demographic details, medical history, lifestyle behaviors like smoking and exercise habits, as well as physiological measurements such as blood pressure and cholesterol levels. Throughrigorous data preprocessing and feature engineering, we ensure data quality and extract meaningful features crucial for predictive modeling. Machine learning algorithms such as logistic regression, decision trees, random forests, or neural networks are then employed to train models that can effectively predict stroke occurrences based on the identified risk factors, and also we manually detect the Heart rate and glucose levels with their respective sensors and hardware's with model performance rigorously evaluated using standard metrics. Glucose level and heart rate can both play roles in the risk and outcomes of a brain stroke,

Glucose Level (Blood Sugar):

High blood glucose levels, such as those found in diabetes, can increase the risk of stroke. Diabetes is a significant risk factor for stroke because persistently high blood sugar levels can damage blood vessels over time, leading to atherosclerosis (hardening and narrowing of the arteries), which can ultimately result in a stroke.Acute changes in glucose levels, such as those experienced during hyperglycemia (high blood sugar) or hypoglycemia (low blood sugar), can also affect stroke risk. Hypoglycemia, for example, can lead to neurological symptoms such as confusion, dizziness, and loss of consciousness, which could potentially increase the risk of accidents or other medical complications that could trigger a stroke.

Heart Rate:

Resting heart rate, as well as heart rate variability (the variation in time intervals between heartbeats), have been linked to cardiovascular health and stroke risk. Elevated resting heart rate has been associated with an increased risk of stroke and other cardiovascular events.

Atrial fibrillation (an irregular and often rapid heart rate) is a known risk factor for stroke. In atrial fibrillation, the heart's upper chambers (atria) beat chaotically and out of coordination with the lower chambers (ventricles), which can lead to the formation of blood clots in the atria. If a clot dislodges and travels to the brain, it can cause a stroke.

Sensors used:

Heart rate sensor:

A heart rate sensor is a device designed to measure the heartbeat rate in real-time. Typically, it uses various technologies such as photoplethysmography (PPG) or electrocardiography (ECG) to detect the heart's electrical activity or changes in blood volume to determine heart rate. These sensors can be integrated into wearable devices like smartwatches, fitness trackers, and chest straps, allowing users to monitor their heart rate during physical activity, exercise, or throughout the day. Heart rate sensors provide valuable insights into cardiovascular health, helping individuals track their fitness levels, monitor stress levels, and even detect potential abnormalities such as arrhythmia's. They play a significant role inpromoting health awareness and encouraging users to adopt healthier lifestyles.

Near Infra Red Sensor:

Near-infrared (NIR) sensors are devices that detect and measure light in the near-infrared range of the electromagnetic spectrum. These sensors utilize the unique properties of near-infrared light to analyze the composition and characteristics of materials. NIR sensors are commonly used in various fields such as agriculture, food processing, pharmaceuticals, and environmental monitoring. Simultaneously, we focus on heart health by gathering heart rate data from sensors or wearable devices capable of continuous monitoring. The heart rate data undergoes preprocessing to handle noise and missing values,followed by feature extraction to capture relevant information related to heart rate variability, rhythm irregularities, and waveform characteristics. Leveraging machine learning techniques for anomaly detection, such as clustering or supervised learning methods, we develop models capable of identifying deviations from normal heart rate patterns indicative of potential heartabnormalities. Model validation and tuning using labeled data ensure the accuracy and reliability of our anomaly detections

Secondly Developing a comprehensive health monitoring system involves integrating multiple components, including stroke prediction using machine learning algorithms and detecting glucose levels using a Near-Infrared (NIR) sensor, alongwith monitoring heart rate using a heart rate sensor. For stroke prediction, data collection involves gathering relevant medical and lifestyle data, followed by feature selection/engineering and model training/testing using algorithms like logistic regression or neural networks. The NIR sensor is calibrated to measure glucose levels accurately, with data preprocessing and algorithm development linking sensor readings to glucose levels. Similarly, the heart rate sensorcaptures heart rate data, which is processed using signal processing techniques to monitor heart rate in real-time. Integrating these systems allows for a holistic health monitoring approach, with potential applications in personalized healthcare and early intervention strategies. Collaboration with experts in data science and healthcare ensures the accuracy, reliability, and ethical handling of sensitive health data within this framework.

The integration of these predictive models and anomaly detection systems into a unified platform allows for real-time monitoring and analysis of patient data. By providing actionable insights and timely alerts, our system empowers healthcare professionals to intervene early, personalize treatment plans, and implement preventive measures tailored to individual patient needs.

Algorithms Used:

1) Naive Bayes:

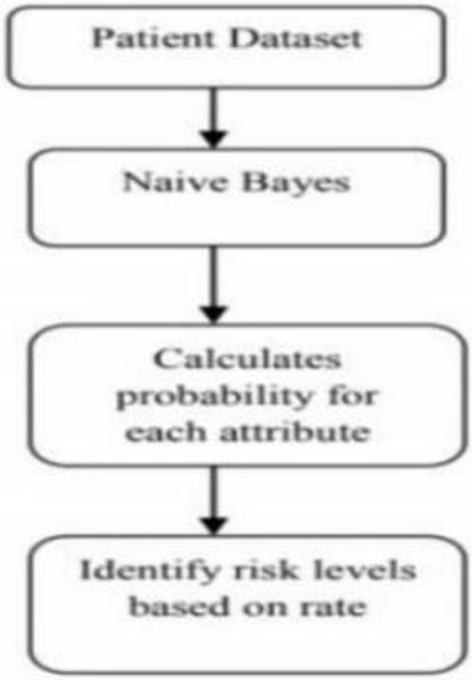

Fig 1: Probabilistic Machine learning model

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

2) Logistic Regression:

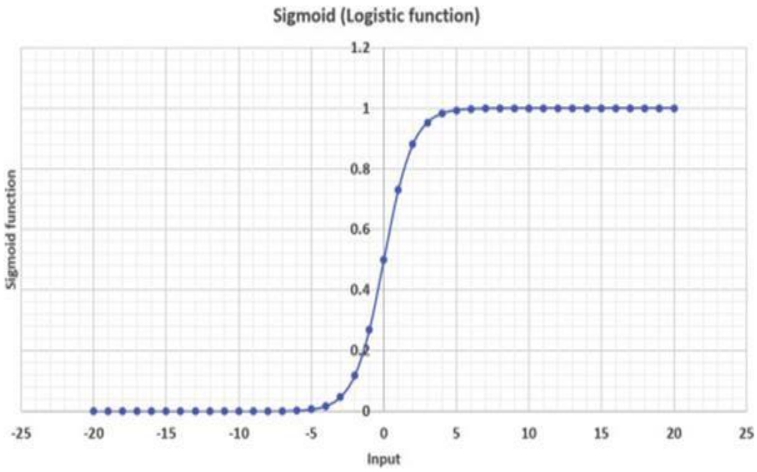

Fig 2: Probability of a discrete outcome

3) Decision Tree:

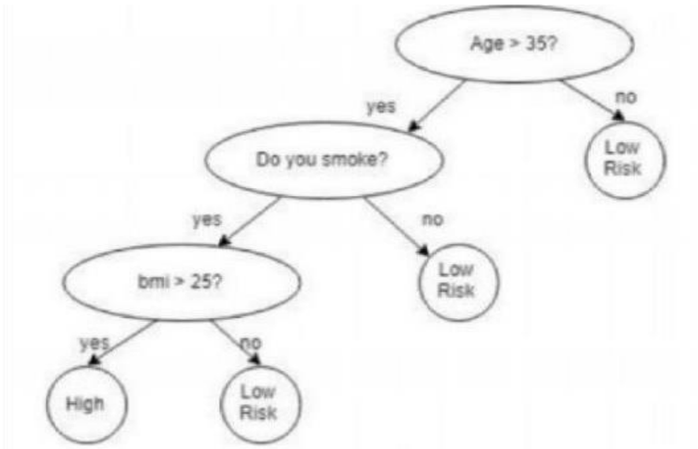

Fig 3: Decision Support Tool

4) Random Forest:

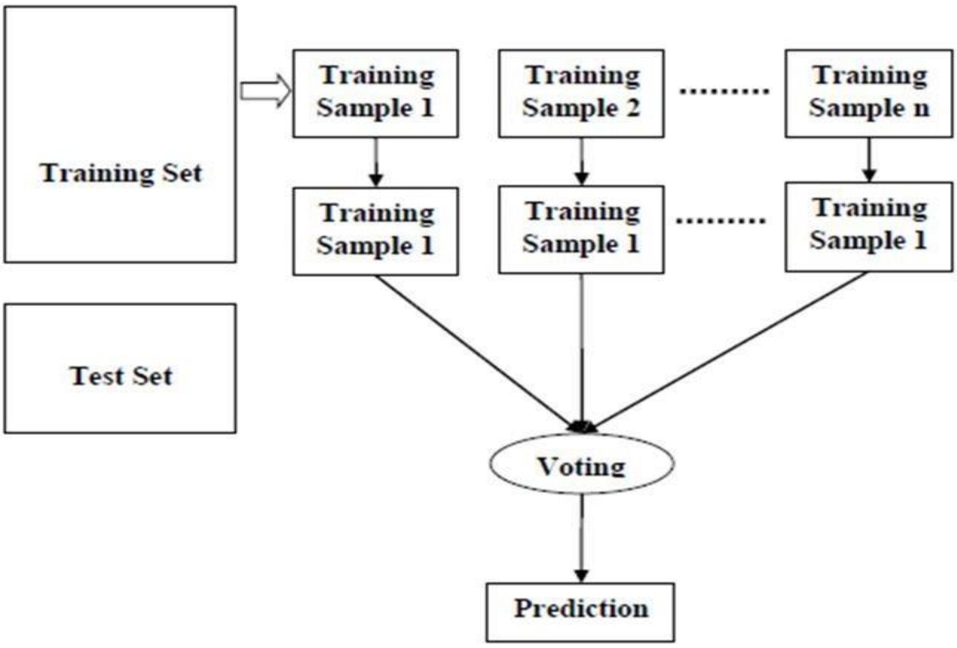

Fig 4: Decision tree on various dataset

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

5) Block Diagram:

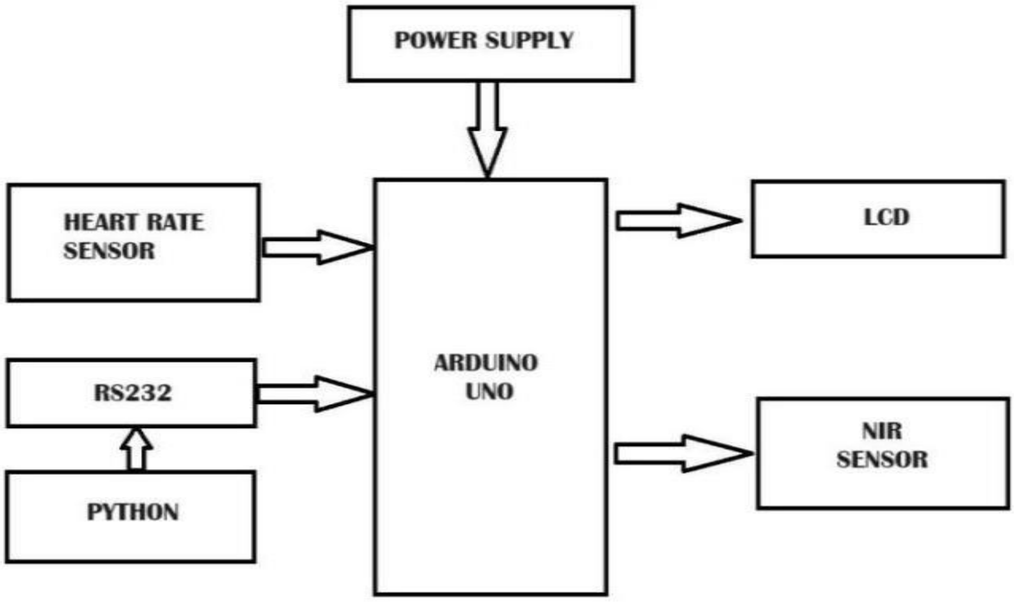

Fig 5: Glucose Level Monitor

6) Types of Datasets used (eg):

Gender	Age	Hypertension	Heart_disease	Ever_married	Work_type	Residence_type	Avg_glucose_level	Bmi	Smoking_status	Stroke
Male	67	0	1	Yes	Private	Urban	228.69	36.6	Formerly smoked	1
Male	80	0	1	Yes	Private	Rural	105.92	32.5	Never smoked	1
Fema le	49	0	0	Yes	Private	Urban	171.23	34.4	Smokes	1
Fema le	79	1	0	Yes	Self-employed	Rural	174.12	24	Never smoked	1
Male	81	0	0	Yes	Private	Urban	186.21	29	Formerly smoked	1
Male	74	1	1	Yes	Private	Rural	70.09	27.4	Never smoked	1
Fema le	69	0	0	No	Private	Urban	94.39	22.8	Never smoked	1

IV. CONCLUSION

In this paper, four learning techniques were explored to predict Stroke by manually detecting two main datasets such as Heart rate and Glucose level. We made the following observations after significant analysis. After detecting the datasets manually they are transferred to those machine learning algorithms by accepting the request by the program whether to accept the manually detected heart rate and glucose rate in the main datasets All five models are compared and the best-performing model is considered for the occurrence of prediction with various datasets. Conclusion Several assessments and prediction models, Decision Tree, Logistic Regression, Naïve Bayes, and Random Forest, showed acceptable accuracy in identifying stroke-prone patients. This project hence helps to predict stroke risk using manually detected datasets such as glucose level and heart rate using the respective sensors and provides personalized warning and lifestyle correction

ACKNOWLEDGEMENTS

The completion of this research work has been made possible through the support and contributions of various people for whom we would like to express our sincere gratitude. We extend our deepest appreciation to Aishwarya raj, Assistant Professor of the Department of Biomedical, Agni College of Technology for her continuous support. Her technical expertise greatly influenced the direction and quality of our research. Her mentorship significantly contributed to the success of this project. Her guidance played a pivotal role in shaping our methodology and analysis. Special thanks to the Principal, of Agni College of Technology for providing us with the necessary equipment for theresearch process

V. REFERENCES

[1] Jeena R.S. and Dr.Sukesh Kumar used SVM with appropriate kernel functions which had been investigated for analysis of stroke. Pre-processing was done to remove redundant and incompatible data, and 350 inputs were taken for the prediction [2020]

[2] Prediction of Stroke Using Machine Learning Kunder Akash Mahesh, Shashank H N, Srikanth S and Thejas A M from Dept. of Computer Science & Engineering CMRIT, Bangalore Karnataka, India, used principles of machine learning over large existing data sets to effectively predict stroke based on potentially modifiable risk factors [2020]

[3] "Stroke prediction using artificial intelligence"- M. Sheetal Singh, Prakash Choudhary - In this paper, Here, decision tree algorithm is used for the feature selection process, principle component analysis algorithm is used for reducing thedimension [2024]

[4] Stroke Prediction Using Machine Learning Algorithms by Harshitha K V, Harshitha P, Gunjan Gupta, Vaishak P, and Prajna KB from Department of Electronics & Communication Engg, Nitte Meenakshi Institute of Technology,Bangalore, Karnataka, India. The project mainly aims at predicting the chances of occurrence of stroke using the emerging Machine Learning techniques [2022]

[5] Predicting Stroke Risk with an Interpretable Classifier, The work of Sergio Peñafiel was supported by the Conicyt (Chile) Master Scholarship under Grant 22180506. The development of a prediction method that gives information about the most probable causes of a high stroke risk and can deal with incomplete data records. It is based on the Dempster-Shafer theory of plausibility [2021]

[6] Stroke Risk Prediction with Machine Learning Techniques Georgios D. Barmparis, Academic Editor, Maria E. Marketou, Academic Editor, and Giorgos P. Tsironis, Academic Editor [2022]

[7] A predictive analytics approach for stroke prediction using machine learning and neural networks. Author links open overlaypanel Soumyabrata Dev a b, Hewei Wang c d, ChidozieShamrock Nwosu e, Nishtha Jain a, Bharadwaj Veeravalli f, Deepu John [2022]

[8] Predicting Risk of Stroke From Lab Tests Using Machine Learning Algorithms: Development and Evaluation of Prediction Models Reviewed by Tiago Taveira-Gomes, Klaus Kades, and Anupam Sule [2021]

[9] Analysis and Prediction of Stroke using Machine Learning Algorithms by V.Vijayalakshmi, V.Geetha, M.Mohammedljaz [2020]

[10] Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study With a Web Application for Early Intervention by Krishna Mridha, Sandhesh Ghimire, Jungpil Shin, Md. Mezbah Uddin [2023]