Brain Stroke Prediction Using Machine Learning

Latharani T R
Dept of CSE
Jit, Davangere
lathaquick@gmail.com

Divya G C
Dept of CSE
Jit, Davangere
divyagc10@gmail.com

Tejashwini B R
Dept of CSE
Jit, Davangere
tejaswinibr89@gmail.com

Roja D C
Dept of CSE
Jit, Davangere
dcroja953@gmail.com

Madhusudhan Hovale
Dept of CSE
Jit, Davangere
madhusudhanhovale0667@gmail.com

Abstract-**A stroke, often referred to as a cerebrovascular accident (CVA), occurs when a portion of the brain losses blood flow, which causes the area of the body that those brain cells control to become dysfunctional. Because of poor blood flow or bleeding into the brain tissue, this decrease of blood supply may be ischemic or hemorrhagic. Due to the possibility of fatality or lifelong impairment, a stroke is a medical emergency. Ischemic strokes may be managed, however this treatment must begin within a few hours of the onset of stroke symptoms. If a stroke is suspected, the patient, their family, or witnesses should call emergency medical assistance right away. A transient ischemic attack (TIA or mini-stroke) is a brief ischemic stroke in which the symptoms disappear on their own. This situation also demands an immediate evaluation to lower the risk of a future stroke. If all symptoms go away within 24 hours, that would be a stroke by definition, not a TIA. Stroke is the second leading cause of mortality globally, claims the World Health Organization (WHO). accounting for around 11% of all fatalities. Our ML model uses a dataset for survival prediction to determine a patient's likelihood of suffering a stroke based on inputs including gender, age, various illnesses, and smoking status. Our dataset, in contrast to most others, concentrates on characteristics that would be significant risk factors for a brain stroke.**

Keywords - *Computer learning, brain damage. Transient ischemia attack, ischemic stroke.*

I. INTRODUCTION

Atrial fibrillation can result in stroke, which has the potential to be fatal. For doctors, predicting a stroke takes time and is tiresome. Stroke is a debilitating condition that mainly affects those over the age of 65. Similar to "coronary episodes," It harms the brain and affects the heart, making cardiovascular diseases the third largest cause of death in the US and other wealthy nations.

When there is a disruption or reduction in the blood flow to the brain, a stroke occurs. Hemorrhagic and ischemic strokes are the two main types. Both hemorrhagic and ischemic strokes are caused by a disruption in the blood supply The two kinds of hemorrhagic stroke are intracerebral and subarachnoid haemorrhage. "Ministroke" is another name for transient ischemic attack.. Therefore, the purpose of this work is to identify and forecast employing machine learning (ML) methods like logistic regression, SVM, KNN, decision trees, and random forests, one may estimate the risk of brain strokes. 39 studies on ML for brain stroke were found in the ScienceDirect online scientific database between 2007 and 2019. [2]. In ten investigations for stroke issues, Support Vector Machine (SVM) was found to be the best models. In addition, the majority of studies are in stroke diagnosis whereas the majority of studies are in stroke treatment, indicating a research gap that needs to be filled. Similar to this, CT pictures are a common dataset in stroke. Finally, effective methods employed for each category include SVM and Random Forests [2]. The significance of various ML approaches employed in brain stroke is highlighted by the current study.

II. LITERATURE REVIEW

Five machine learning techniques were applied to the Cardiovascular Health Study (CHS) dataset to forecast strokes. To get the best results, the authors combined the Decision Tree with the C4.5 approach, Principal Component Analysis, Artificial Neural Networks, and Support Vector Machine. However, there were fewer input parameters in the CHS Dataset used for this study.

The writers of this article examined numerous stroke-related issues using cutting-edge research. Based on their similarities, the examined research were divided into numerous categories.

According to the study, Studies that employed different performance metrics for different tasks while taking into account different datasets, techniques, and tuning variables make it difficult to compare them. As a result, only the research topics that were the focus of multiple investigations and those that had the best categorization accuracy in each section are mentioned [1].

In their work, Harish Kamal, Victor Lopez, and Sunil A. Sheth discuss how Machine Learning (ML) using pattern recognition algorithms is increasingly being used to diagnose, treat, and forecast problems and results for patients with various neurological diseases Recent developments in acute ischemic stroke (AIS) evaluation and treatment have enhanced the need for expanded use of neuroimaging in decision-making. This article discusses current advancements and uses of ML in neuroimaging with an emphasis on acute ischemic stroke. Numerous applications of computer learning have been made, including the analysis of cerebral edoema, Early detection of imaging diagnostic findings, assessment of the timing of treatment beginning, lesion segmentation, and the fate of salvageable tissue, as well as the prediction of side effects and patient outcomes following therapy.

The final section of the study looks at how machine learning applications are growing in the medical sector for both therapeutic and diagnostic purposes., with AIS proving to be a promising area due to its rapid growth and increasing reliance on neuroimaging. This industry has a special need for ML solutions because it is struggling with the problem of increasingly complicated data and has a shortage of human expert personnel. Building a solid dataset for effective ML network training may involve teamwork across various institutions in the future of ML for AIS. [2]

The postings that people make on social media have been used to predict strokes. To identify the many symptoms connected to stroke illness in this specific experiment, the scientists employed the DRFS approach. The model's overall execution time grows, which is not optimal, as a result of the usage of Text from social media posts is extracted using natural language processing.

To build the data frequency matrix, all text data were processed into tokens using the "quanteda" NLP programmer. Ten-fold cross validation was used to correct the training set's bias. Manual labelling for AIS was done to identify clinical notes. For the binary classifiers In addition to support vector machines, they also employed single decision trees, binary logistic regression, and naïve Bayesian classification. The F1-measure was used to assess how well the algorithms performed. Additionally, the method's efficiency was evaluated in light of n-grams and word weighting documents based on frequency-inverse frequency. The paper's conclusion focused on how supervised ML-based NLP algorithms can be used to automatically classify brain MRI reports in order to identify patients with AIS. One decision tree served as the best classifier for categorizing brain MRI data with AIS. [3].

In order to predict stroke, Joon Nyung Heo et al. took into account three machine learning models: deep neural networks, random forests, and logistic regression. After reading this study,

we found that Deep Neural Network (DNN) is frequently utilized for ischemia or acute stroke patients and also affects long-term prediction. The DNN model outperformed the other models with an accuracy of 88% in relation to the inputs. In order to enhance the model, automated calculations with higher accuracy are performed less often [2].

Using a Bayesian model known as Bayesian Rule Lists (BRL), Benjamin Letham et al. predicted a predictive analysis that generates a distribution of permutations from a sizable, processed set of data. The pre-processed data minimizes the model space for different sets of fragments, and as a result, the method scales with the smallest amount of the data set with the fewest characteristics. High levels of accuracy, precision, and tractability may be attained with the aid of the BRL approach [6].

which included 287 stroke patients—of whom 16 were omitted because they showed no signs of a stroke—were obtained to enable the research. After eliminating the information from the 60 patients whose NIHSS questionnaires included missing values or answers that were outliers, the final NIHSS data contained 227 individuals. The seniors over 65 who participated in the research comprised In this experiment, a suggested system is used to classify and forecast Employing representative categorization and prediction models created using data mining and machine learning approaches, the stroke severity score was divided into four categories. The suggested system's experiment accuracy is assessed using recall and precision as the measures. The project provided speedier and more accurate predictions of stroke severity as well as effective system functioning through the application of multiple Machine Learning algorithms, C4.5 decision tree, and Random Forest categorization and prediction.

III.   PROPOSED SYSTEM

The proposed system functions as a machine that supports predictions and can help the user with diagnosis. The output prediction algorithms may be able to achieve substantially higher accuracy than the current system. In the suggested system, it has been found that the practical application of various gathered data proceeds more quickly.

.Advantages:
1. Exceptional efficiency and precision
2. Users have easy access to the data and information gathered for predictions.

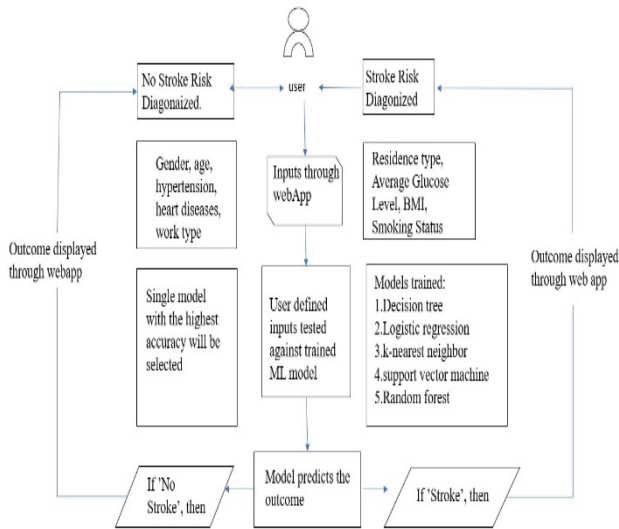The system offers consumers precautions they may take to lower the danger factor.

315

Figure 1: A system's architecture.

Description in Detail of the architecture of a system [1]

USER: The user of our online application will be someone who is curious about their chance of contracting a brain ailment.

WebApp inputs: The user will be prompted to provide information about their gender, age, hypertension, heart conditions, marital status, occupation, type of residence, average blood glucose level, BMI, and smoking status. All of these details are required in order to forecast the likelihood of a stroke occurring in that person.

User-defined inputs are compared to the ML model: The trained ML model that would assist predict the likelihood of stroke based on new user-provided data was picked from a total of five machine learning methods. The algorithm that got the best accuracy score was chosen. Logistic Regression, Decision Trees, and other machine learning

A. Collecting user input using our web application. Our web application's initial step will be to take some possibility of fresh data from the user's perspective. Decision Tree, Logistic Regression, K-Nearest Neighbour, Support Vector Machine, and Random Forest are examples of machine learning algorithms.

Model predicts the Outcome: Using a trained machine learning model, the likelihood that a user will experience a stroke is calculated. If the user is at risk for a brain stroke, the model will predict the outcome based on that risk, and vice versa if they do not.

No Stroke Risk Diagnosed: The user will learn about the results of the web application's input data through our web application. "No Stroke Risk Diagnosed" will be the result for "No Stroke".

Stroke Risk Diagnosis: The user will learn about the results of the web application's input data through our web application. When "Stroke" is selected as an outcome, the text "Stroke Risk Diagnosed" will appear.

B. Comparing the input data to the practice data.

The user data will be compared to the learned data in a processing step, and thereafter , as was said in the later section of the preceding module, will result in accurate results.

C. Receiving test findings.

The final phase of our web application is to provide the user with accurate and exact results, allowing them to proceed as necessary in light of the findings.

IV. METHODOLOGY

To achieve the highest results and accuracy, the system has been constructed utilizing 5 different machine learning algorithms. The algorithms used to construct the machine learning model include Logistic Regression, Support Vector Machine (SVM), K Nearest Neighbour (KNN), Decision Tree, and Random Forest.

• Front end: CSS (Cascading Style Sheets), Bootstrap, and HTML (Hyper Text Markup Language).
• Framework: Python API for creating web apps called Flask.

• Using Google Collaboration as the Runtime Environment Google Research created Colaboratory, sometimes known as "Colab" for short. Colab excels in three areas: data analysis, instruction, and machine learning. Anyone may create and execute arbitrary Python code using the browser. Colab is a hosted Jupyter notebook service that offers free access to computer resources, including GPUs, and doesn't need to be installed.

• Dataset
The Prediction of Brain Strokes 5110 rows and 11 columns make up the data collection, which includes attributes like "id," "gender," "age," "hypertension," "heart disease," "ever married," "work_type," "Residence_type," "avg_glucose_level," "BMI," "smoking_status," and "stroke."

• Libraries
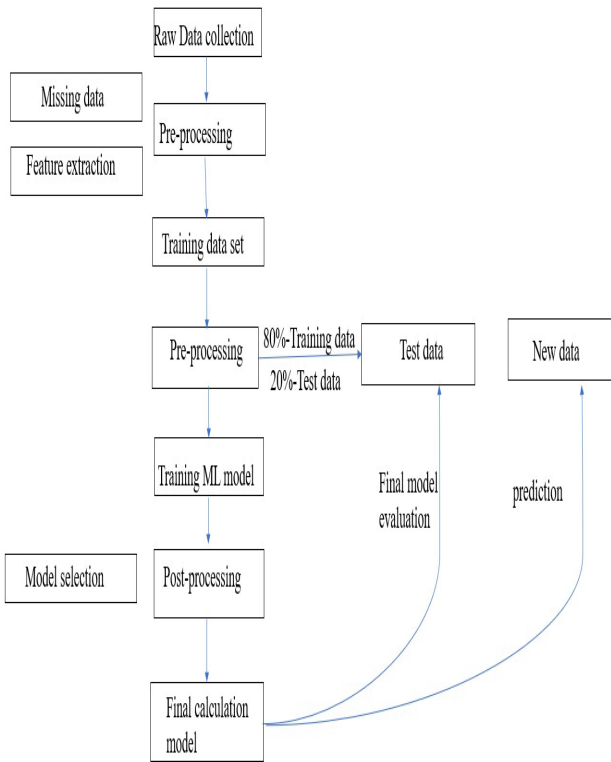Numpy, Pandas, Seaborn, Matplotlib, Sklearn/Scikit-learn, Pickle,

Figure 2 – Workflow [1]

1) Remove any missing values from the training and test data.
2) Converting objects into integers using Label Encoder.
3) Distinguishing between training and test data for the data.
4) ML Model Training:
  ➢ Decision Tree
  ➢ K-nearest neighbour (KNN)
  ➢ Random Forest
  ➢ Support Vector Machine (SVM)
  ➢ Logistic Regression
5) the evaluation of each model's degree of correctness.
6) Choosing the model with the best accuracy score.
7) Build a GUI, then add the model to the GUI module.
8) If a stroke prediction is needed, enter the revised data.
9) Outcome: Predicted data based on chosen model.

The system will finally produce the desired result after the approach, modules, algorithms, and codes have been implemented. The connected stroke prediction system's information entry process may be made easier by using the site, and the GUI is designed to be user-friendly for normal people. It will be simpler to reach your goals if you are aware of them. Five distinct machine learning (ML) techniques have been used to construct the system utilizing our dataset, as indicated in the Implementation.

V.   RESULTS AND DISCUSSION

The following outcomes were produced:
• The Decision Tree Algorithm provides the lowest accuracy, or (92.08%).
• The Logistic Regression algorithm provides the maximum accuracy, or (94.89%).
The results of the evaluation parameters vs. models table are shown below (this was done using the best training and testing dataset split, which was 70% training and 30% testing):

Evaluation parameters / Models	Accuracy Rating	Recall Rating	Precision Rating	F1 Rating
Decision Tree	92.08	100.00	94.95	97.41
KNN	94.63	82.22	73.87	77.82
Random Forest	94.76	100.00	85.86	92.16
SVM	94.89	88.84	77.73	82.92
Logistic Regression	**94.89**	**100.00**	**97.54**	**98.56**

Table 1- Comparison of model results and evaluation parameters

Figure 3 uses the Python Matplotlib package to display a comparison of the accuracy scores of all trained models.
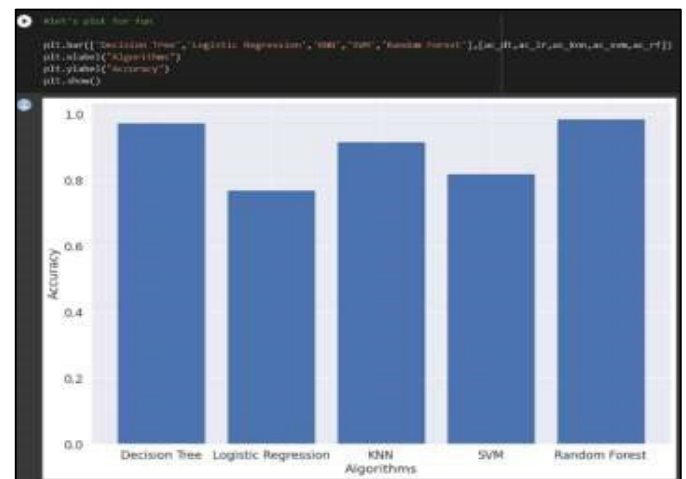

Figure. 3 – Trained Models Accuracy Score comparison

For the purpose of predicting the risk of stroke, The data supplied to the GUI was utilised with the Random Forest model, which was trained using the dataset. and the newly supplied user data was compared to the trained model.As illustrated in

317

Figure 5, the suggested approach helped us analyse the most effective method for obtaining user inputs for the GUI development portion of our project.

Figure 4 – GUI

Figure 6 shows No Stroke Risk Diagnosed: The user will learn about the results of the web application's input data through our web application. "No Stroke Risk Diagnosed" will be the result for "No Stroke".
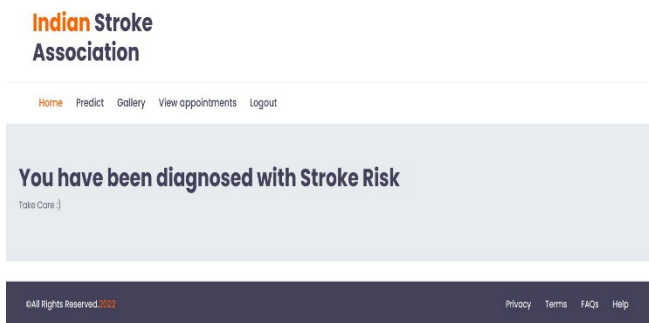

Figure 5- For Stroke

Figure 7 shows Stroke Risk Diagnosis: The user will learn about the results of the web application's input data through our web application. When "Stroke" is selected as an outcome, the text "Stroke Risk Diagnosed" will appear.
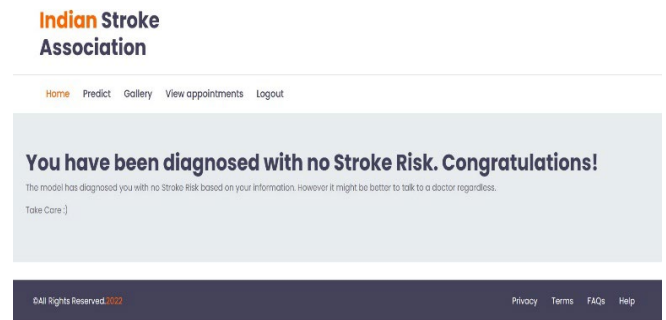

Figure 6 - For No Stroke

VI.   CONCLUSION

As a result, we proposed a system that uses a few user-provided inputs and trained machine learning algorithms to help with the cost-effective and efficient prediction of brain strokes. With a maximum accuracy of 98.56%, a system for anticipating brain strokes has been developed using five machine learning algorithms. The system was developed in this way to give a user interface that is both simple and efficient while also demonstrating empathy for both users and patients. The system's capacity for future growth offers the chance to improve user experience and outcome. The user will thus save a lot of time and be more ready. The implemented system's potential future range could include: Improving the model's accuracy. More details on brain strokes can be explained. Providing people with the option to see their outputs based on their inputs.

VII.   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

319