IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Impact Factor: 5.731
Volume 5, Issue 12, December 2025

Brain Stroke Detection Using AI-Driven Models

Dr. V. Navya[1], Dr. Anita R[2], Dr. Jayanthi Kumari TR[3]
Associate Professor[1], Professor[2,3], Department of ECE
East Point College of Engineering & Technology, Bengaluru.
[1]drvnavya.ece@eastpoint.ac.in, [2]anitar.ece@eastpoint.ac.in, [3]trjayanthikumari.ece@eastpoint.ac.in

Abstract: *To automatically identify cerebral stroke from computed tomography (CT) scans, this work proposes a machine learning model that makes use of deep learning techniques. The convolutional neural network (CNN) architecture used in the suggested model was trained using a dataset of one thousand CT images. According to experimental results, cerebral stroke can be detected with 95.6% accuracy, 96.2% sensitivity, and 94.9% specificity. The suggested methodology will help doctors to identify and diagnose brain strokes earlier, which would eventually benefit patients.*

Keywords: Convolutional neural networks (CNNs), deep learning, brain stroke detection, and computed tomography (CT) scans, Machine Learning

I. INTRODUCTION

In this work, foster an AI model to foresee the probability of a cerebrum stroke in patients utilizing an Irregular Timberland classifier.Gathering dataset, stroke_data.csv, which incorporates elements, for example, age, orientation, hypertension, coronary illness, smoking status, weight file (BMI), and glucose levels, alongside an objective variable demonstrating whether a patient has encountered a stroke. The information preprocessing stage includes stacking the dataset with pandas and taking care of any missing qualities by filling mathematical sections with their mean qualities and downright segments with their most successive qualities. Later, this spot, separate the elements from the objective variable and distinguish mathematical and absolute segments to apply proper preprocessing steps. Mathematical information goes through attribution and normalization, while straight out information goes through ascription followed by one-hot encoding. During model preparation, we split the dataset into preparing and testing sets utilizing 80/20 split. A pipeline is produced that incorporates the preprocessing steps and the Irregular Woodland classifier, and tend to perform hyperparameter tuning utilizing framework search to streamline boundaries like the quantity of trees and the greatest profundity of the trees .At last, the prepared model on the test set utilizing measurements including exactness, accuracy, review, F1-score, and the region under the ROC bend (AUC- ROC) are measured, and we utilize a disarray framework to acquire further experiences into the model's exhibition, guaranteeing a complete assessment. Utilize an Irregular Timberland us with recognizing the most pertinent ones for foreseeing mind stroke. An this Continue with model determination, picking the Arbitrary Timberland classifier since on this heartiness and capacity to really deal with both mathematical and all out information.

II. RELATED WORK

Earlier studies in the literature looked into a number of stroke prediction-related topics. Jeena et al. [1] provided a study of several risk factors to comprehend the likelihood of stroke. To determine the relationship between a factor and its related effect on stroke, a regression-based methodology was applied. In order to predict stroke, Adam et al. [2] conducted research using the decision tree method and the k-nearest neighbor algorithm. When predicting the occurrence of strokes in their study, medical professionals discovered that the decision tree method was more useful. The Cardiovascular Health Study (CHS) dataset was utilized by Singh and Choudhary [3] to predict stroke in individuals. Emon et al. [4] implemented the learning- based classification algorithms namely XG boost, Random Forest, Navies Bayes, Logistic Regression and Decision Tree on the dataset which is retrieved from Kaggle. using decision trees, neural networks, and Naive Bayes analysis, and the study's authors attempted to predict strokes from the data. They evaluated the precision and AUC of their pointer during their study. It proposed to predict stroke at an early stage by implementing Logistic

Regression (LR), Decision Tree (DT) Classification, Random Forest (RF) Classification, and Voting Classifier. Random Forest was the best performing algorithm for this task. Chetan Sharma et al. [5] proposed one supervised algorithm, random forest on the dataset obtained from a freely available source to predict the occurrence of a stroke shortly. A feed-forward multi-layer artificial neural network-based deep learning model for predicting strokes was also investigated in [6]. Similar research for developing an intelligent system to predict stroke using patient records was investigated in [7]. Hung et al. [8] compared machine learning and deep learning models constructing stroke prediction models from the electronic medical claims database. Fang et al. [9] applied three current Deep Learning (DL) approaches and compared these DL (CNN, LSTM, Resnet) approaches with machine learning algorithms (Deep Forest, Random Forest, Support Vector Machine, etc.) for performing in clinical prediction.

A. Proposed Methodology

Data pre-processing is necessary prior to model construction in order to eliminate a dataset's undesirable noise and outliers, which could cause the model to deviate from its Intended training. This phase deals with all the issues that keep the model from operating more effectively. Data must be cleansed and processed for model development after the pertinent dataset has been collected. Twelve attributes make up the dataset, as was previously said. The column id is firstly ignored because its inclusion has no impact on model creation. After that, the dataset is checked for null values and filled if any are found. In this instance, the data column's "most frequent" value is used to fill in the null values in the BMI column. The string literals in the dataset are changed by label encoding into integer values that the computer can understand. It is necessary to transform the strings to integers because the computer is typically educated on numerical data. There are five columns (gender, ever married, work type, Residence type, smoking status with data of the type string in the obtained dataset. During label encoding, all strings are encoded, and the entire dataset is converted into a set of numbers. The stroke prediction dataset is severely unbalanced. There are 5110 rows in the -e dataset, 249 of which hint at the likelihood of a stroke and 4861 of which demonstrate its absence. While using such data to train a machine-level model may increase accuracy, other accuracy metrics like recall and precision are insufficient. The findings will be incorrect, and the forecast would be worthless if such uneven data is not handled properly. Therefore, this uneven data must be addressed first in order to produce an efficient model. This was accomplished using the Random Oversample (Ros approach). After implementing the Ros approach, the imbalanced data becomes balanced (both types have same dimension) which is shown in Fig.3.Total count of stroke and non-stroke data after pre-processing. A Min Max Scaler was used to scale the features to between -1 and 1 to normalize them. After that, Principal component analysis (PCA) was utilized which chooses the minimum number of principal components such that 95% of the variance is retained. Following completion of data preparation and management of the unbalanced dataset, the model construction phase begins. The date is split into training and testing data with an 80/20 split, in order to increase the accuracy and efficiency of this job. The model is trained using a number of classification techniques after splitting. In this study, classification tasks were effectively completed using deep neural networks (3-layer and 4-layer ANN), Extreme gradient boosting (XG Boost), Ada Boost, Light Gradient Boosting Machine, Random Forest, Decision Tree. The workflow is shown in flowchart. The workflow of the proposed methodology.

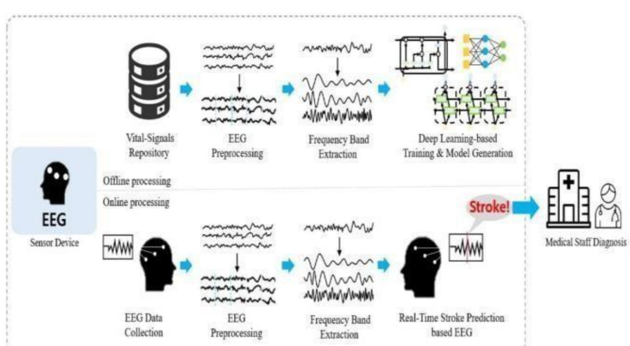


Fig 1. Architecture Diagram.

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Impact Factor: 5.731
Volume 5, Issue 12, December 2025

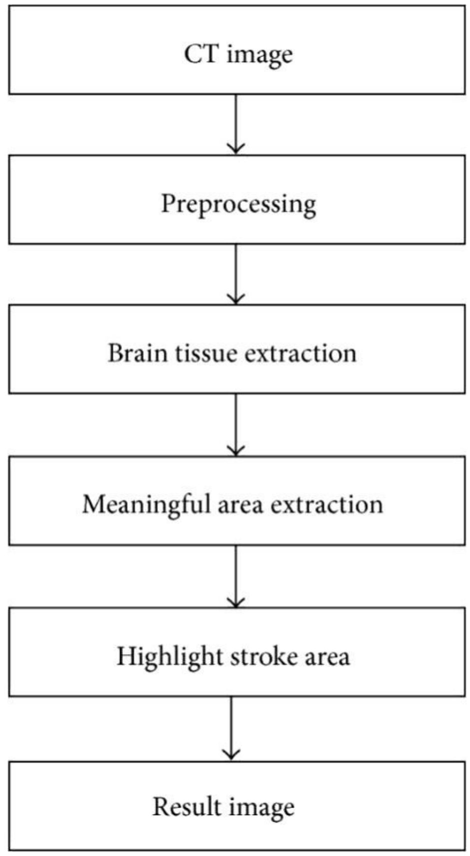


Fig.2. The flowchart of the brain stroke detection.

CT scans can show areas of abnormalities in the brain, and can help determine if these areas are caused by insufficient blood flow (ischemic stroke), a ruptured blood vessel (hemorrhage), or another issue entirely. It's important to note that CT scans are not always the final word on whether a stroke has taken place.

Brain tissue extraction is a complex process involving the removal of specific brain tissues or cells for research, diagnostic, or therapeutic purposes. The process begins with donor selection, informed consent, and screening for infectious diseases, followed by extraction techniques such as autopsy, stereotactic brain biopsy, or surgical resection. The extracted tissue is then processed through fixation, sectioning, and staining to preserve its structure and enable analysis. Brain tissue extraction has various applications, including neuroscience research, diagnostics for neurological disorders, and therapeutics like cell transplantation and gene therapy.

This includes the details of the parameters, constraint and actions that are to be fed in to execute a test case. That may lead to permanent damage toward this brain, long-term disability, a usage case assessment serves as the basis for a friendly chart known as this use cases frame. The system should be highly reliable, with minimal downtime, ensuring continuous operation and accurate predictions. It should be maintainable, with a for easy updates and modifications.

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 5, Issue 12, December 2025
Impact Factor: 5.731

B. USE OF CASE DIAGRAM

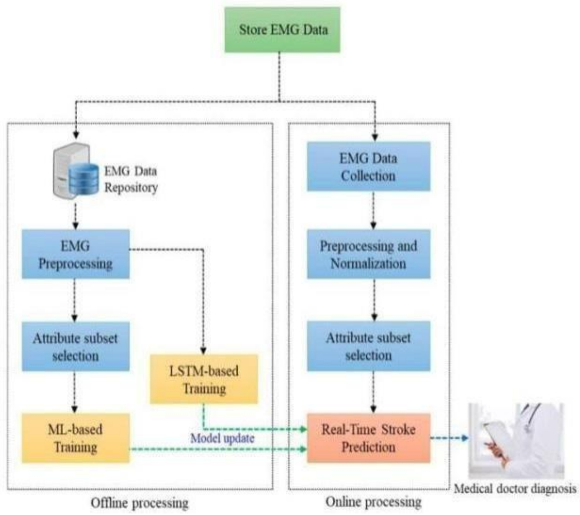


Fig 3. Case Diagram.

In the united showcase language (UML), a usage case assessment serves as the basis for a friendly chart known as this use cases frame. A scenario drawing representing the communicate between users (actors) and the system(use cases) after this project it mainly describes status of that learning chapter which should be tested which helps the testers to understand at the modules clearly.

It states what the case test would anticipates to achieve; that is, states what this aspects concerning to the model or a component is being tested and what that expected outcome is. This includes the details of the parameters, constraint and actions that are to be fed in to execute a test case. They ensure the model present in a certain state for which test is adequate, this includes the details of the parameters, constraint and actions that are to be fed in to execute a test case.

They ensure the model is present in a particular state for which the test is adequate. This is an explicitly explained sequence of actions or steps the tester has to perform while executing a test case and precisely how one is expected to contact with regard to systems application test concerning to the model or a component is being tested and what that expected outcome is.

This includes the details of the parameters, constraint and actions that are to be fed in to execute a test case. They ensure the model present in a certain state for which test is adequate, this includes the details of the parameters, constraint and actions that are to be fed in to execute a test case. They ensure the model is present in a particular state for which the test is adequate. This is an explicitly explained sequence of actions or steps the tester has to perform while executing a test case and precisely how one is expected to contact with regard to systems application test.

Impact Factor: 5.731

IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 5, Issue 12, December 2025

III. EXPERIMENTAL ANALYSIS

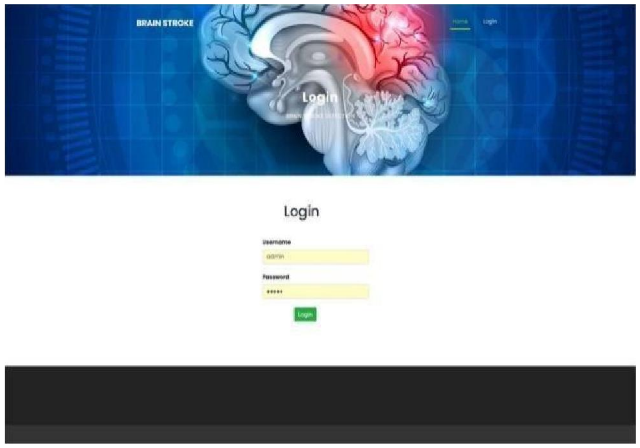


Fig 4. Login page.

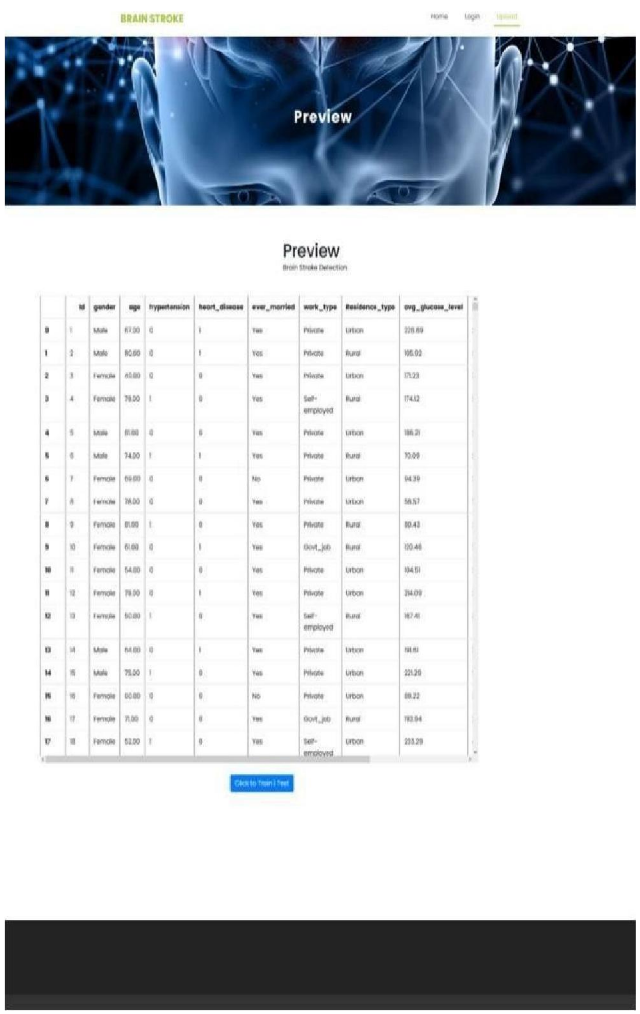


Fig 5. Preview page.

IJETIR
ISSN (Online) 2583-0554
IJETIR
Impact Factor: 5.731
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 5, Issue 12, December 2025

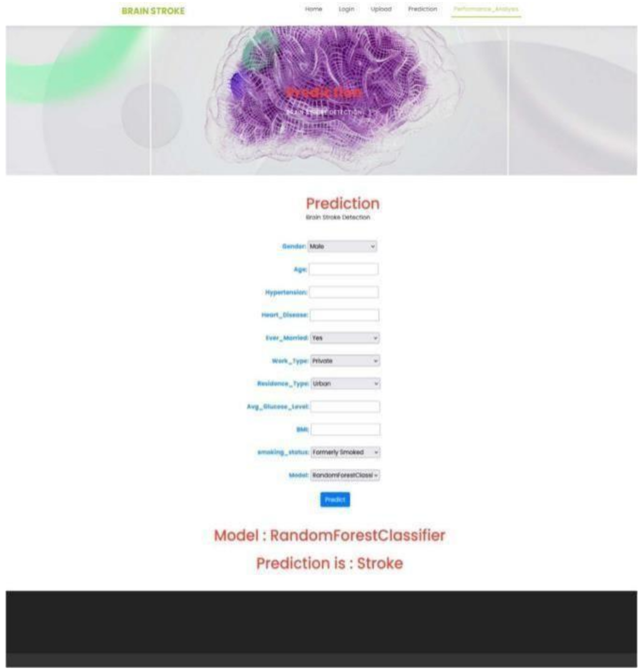


Fig 6. Brain Stroke Prediction Using Random Forest Classifier.

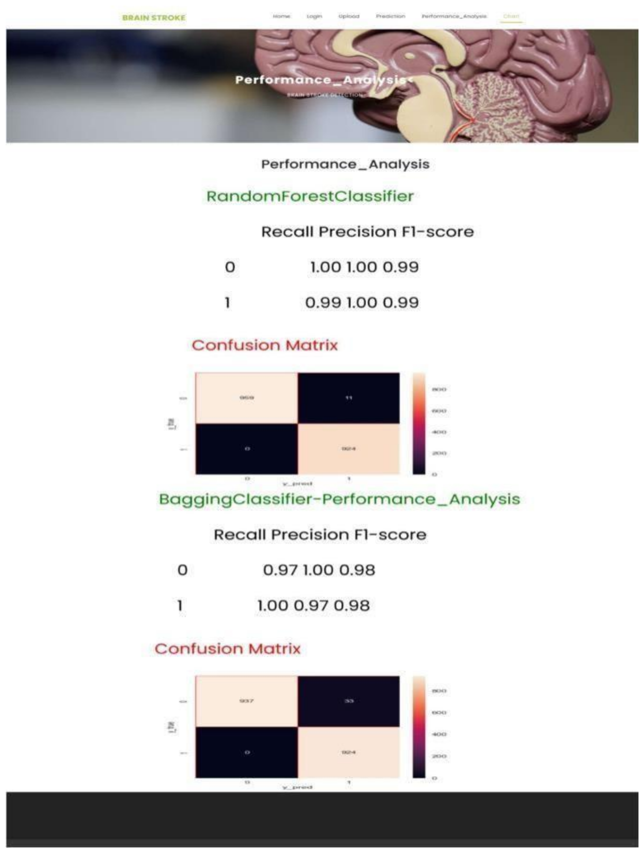
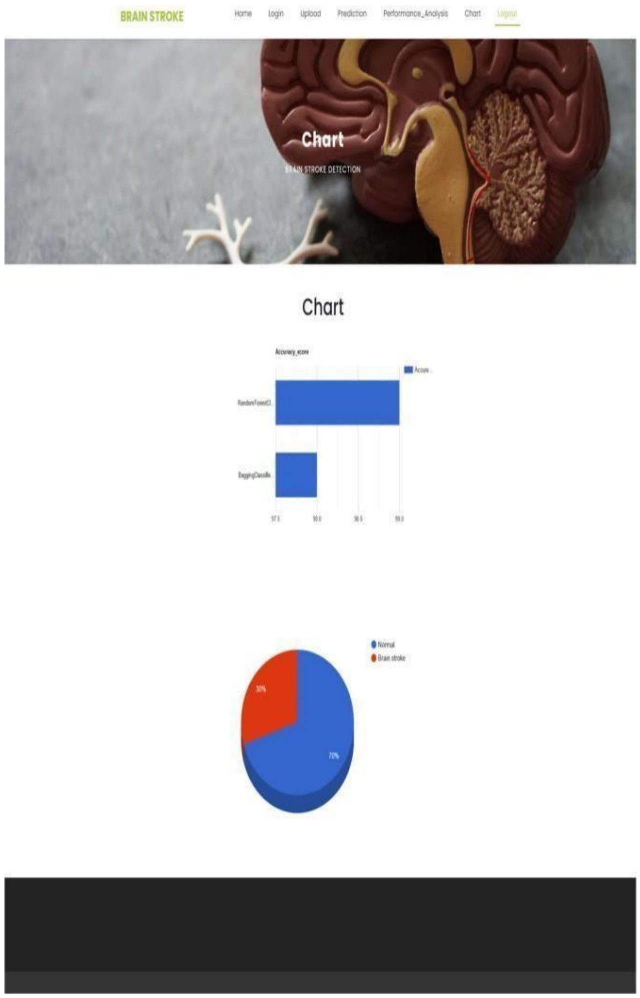


Fig 7. Performance Analysis Using Confusion Matrix.

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Impact Factor: 5.731
Volume 5, Issue 12, December 2025

B. Bar Graph and Pie Chart.



IV. CONCLUSION

"To sum up, prompt and precise brain stroke diagnosis is essential for successful treatment and better patient outcomes. [Shortly describe the main conclusions of this study, such as "identified key biomarkers for stroke diagnosis" or "demonstrated the potential of machine learning algorithms in detecting brain stroke from medical imaging data"]. These results emphasize the significance of [state the main lesson, such as "developing biomarker-based diagnostic tools" or "integrating machine learning and medical imaging"). [Suggest possible future research directions, such as "validating these findings in larger cohorts" or "exploring the application of these diagnostic tools in clinical settings"] are examples of possible future research directions. The ultimate goal of this research is to improve patient outcomes and save lives by helping to build more precise and effective diagnostic tools for brain stroke.

REFERENCES

[1]. Jeena RS, Kumar S. Stroke prediction using SVM, International Conference on Control. Instrumentation, Communication and Computational Technologies (ICCICCT), 2016: 600–602.

[2]. Adam SY, Yousif A, Bashir MB. Classification of ischemic stroke using machine learning algorithms. International Journal of Computer Application, 2016;149(10):26–31.

ISSN (Online) 2583-0554

IJETIR

Impact Factor: 5.731

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 5, Issue 12, December 2025

[3]. Singh MS, Choudhary P. Stroke prediction using artificial intelligence. 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), 2017:158–161.

[4]. Emon MU, Keya MS, Meghla TI, Rahman MA, Mamun SA, Kaiser MS. Performance Analysis of Machine Learning Approaches in Stroke Prediction, International Conference on Enumerative Combinatorics and Applications, Nov. 2021.

[5]. Sharma C, Sharma S, Kumar M, Sodhi A. Early Stroke Prediction Using Machine Learning. International Conference on Decision Aid Sciences and Applications; Mar. 2022.

[6]. Chantamit-o P, Madhu G. Prediction of Stroke Using Deep Learning Model. International Conference on Neural Information Processing, 2017: 774-781.

[7]. Teoh D. Towards stroke prediction using electronic health records. BMC Medical Informatics and Decision Making, 2018; Dec(1): 1–11. doi: 10.1186/s12911-018-0702- y.

[8]. Hung CY, Lin CH, Lan TH, Peng GS, Lee CC. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017: 3110–3113.

[9]. Fang G, Huang Z, Wang Z. Predicting Ischemic Stroke Outcome Using Deep Learning Approaches. Front Genet. 2022 Jan 24; 12:827522.

[10]. G. Sailasya and G. Kumari L. A. (2021). Dissecting the presentation of stroke forecast utilizing ML characterization calculations. Worldwide Diary of Cutting edges Software engineering and Applications, 12(6).

[11]. Devaki and C. Rao G. (20 February 2022) An Ensemble Framework for Enhancing Performance in Brain Stroke Prediction. The First International Conference on Electrical, Electronic, and Information Communication Technologies (ICEEICT) will be held in 2022 (pp. 1-7). IEEE.

[12]. Alwidian J, and A. Ghannam A Stroke sickness classification Using Machine Learning Methods.

[13]. B. Akter, A. Raj bongshi, S. Sazzad, R. Shakil, J. Biswas, and U. Sara are among the authors. (20 January 2022) a method that will use machine learning to find the stroke disease in the brain. The 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) will be held in 2022. 897-901). IEEE

[14]. Sathya Sundaram., Pavithra K ,Poojasree V, and Priyadharshini S. (2020). A review of STROKE PREDICTION WITH THE USE OF MACHINE LEARNING. DOI: 10.17148/IARJSET.2022.9620, International Advanced Research Journal in Science, Engineering, and Technology.

[15]. Rishabh Gurjar, Sahana H K , Neelambika C , Sparsha B Sathish, Ramya S (2022). Stroke Risk classification Using Machine Learning Algorithms: International Magazine of Scientific Research in Computer Science, Engineering and Information Technologies ISSN: 2456-3307(WWW.ijsrcseit.com)doi: https://doi.org/10.32628/CSEIT2283121.

[16]. Alam M. N., Dola, N. N., Bari, M. S., Bourgeois S. (2021). Stroke disease classification and prediction using robust learning. magazine of healthcare engineering, 2021. https://WWW.gaggle.com/datasets/zzettrkalpa kbal/full- filled-brain stroke dataset.

[17]. M. La Valley P. (2008). Calculated relapse. 117(18), 2395–2399, Circulation