iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** VI **Month of publication:** June 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.44565

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue VI June 2022- Available at www.ijraset.com*

# A Generic Real Time Application for Prediction and Categorization of Stroke Using Machine Learning Techniques

Mr. Anand M[1], Geethashree L[2], Impana S[3], Supriya KN[4]
[1]*Assistant Professor of Dept of ISE, GSSS Institute Of Engineering & Technology For Women*
[2, 3, 4]*Dept of ISE GSSS Institute of Engineering & Technology for Women*

***Abstract:*** ***Content-based Strokes are the second greatest cause of death, resulting in serious, long-term disability. A stroke occurs when the cerebrum dies suddenly owing to a shortage of oxygen, which can be caused by a blockage in the circulation or a rupture in a supply line to the brain. The World Health Organization (WHO) stated in. The death rate will continue to rise in the future year's stroke rate. Many projects have been completed. Detecting stroke illnesses is a difficult task. A computer programmer with artificial intelligence. A technique for predicting stroke and its kinds has been created using deep learning. There are two forms of stroke: ischemic and hemorrhagic. A form of stroke known as a transient ischemic attack (TIA). We use datasets from the medical institute's collection in our work.***
***Keywords: Stroke prediction, Machine learning approaches, Sensitivity and Specificity, Comparison Analysis.***

## I. INTRODUCTION

Strokes are the second greatest cause of mortality, resulting insignificant, long-term disability. Stroke is the sudden death of cerebrum cells owing to a shortage of oxygen, which is caused by a blockage in the circulation or a rupture in the brain's supply line. Stroke is a dangerous, life-threatening brain disorder akin to heart attack, which affects the heart. The brain cells are not getting enough blood and oxygen.This stroke prediction web browser for stoke prediction and its types using machine learning algorithms like KNN. Disease registries can give information on disease burden, patient characteristics, care patterns, and outcomes by collecting ongoing data about many aspects of a patient's sickness. These data may be used to investigate illness causation, assess intervention program, enhance treatment quality, and aid in health policy decision- making. Almost all of these registry systems now gather data retrospectively and manually by research coordinators who check medical file records after patients are released. It's a lengthy and error-prone procedure. Obtaining funds to cover the costs of data gathering and registration, as well as the large commitment of resources required to run a register system, is becoming an increasingly difficult task.

The goal of this work is to describe a low-cost, long-term registry system that uses a highly automated data gathering and input method and runs indefinitely with no expiration date. Due to color, size, larger intra-class variability, and light regions other than OD, these characteristics cannot adequately describe glaucoma zones, leading in disappointing outcomes of computer-aided diagnostic methods. Stroke is a vascular irregularity in the brain that causes neurological symptoms such as muscular weakness, numbness, and sometimes death. Ischemic strokes and hemorrhagic strokes are the two forms of strokes. Stroke affects daily functions such as memory, mobility, vision, speaking, and literal ability. Stroke detection is arduous and time-consuming for medical professionals. Patients' demographic information includes their gender, age, and educational level. As a result, an automated method for predicting stroke symptoms based on patient demographic data is required. Methodology for stroke prediction Currently, various research on risk prediction in stroke patients are available.

## II. RELATED MACHINE LEARNING APPROACHES

In this part, previously published publications pertaining to studies on stroke disease prediction utilizing various machine learning methodologies and algorithms are analyzed and reviewed. At the very least, works from the previous ten years were examined for the review. The following are the details:

In 2010, Adithya Khosla and colleagues published a novel automated feature selection approach that selects robust features based on a recommended heuristic: conservative mean. They used it with SVM support vector machines to expand the area under the ROC Curve (AUC). On the cardiovascular Health Study (CHS) dataset, they compared the cox proportional hazards model with a machine learning technique for stroke prediction.

They also suggested a margin-based censored regression method that combined margin-based classifiers with censored regression to producea finer concordance index than the cox model. In terms of AUC and concordance index, this method exceeded the previous state-of-the-art. This strategy may be used to predictthe clinical outcome of a variety of illnesses with inadequate data and poorly understood risk variables. They realized, however, that while their feature selection technique examined the performance of each feature separately, it would not function well in other datasets with strongly linkedcharacteristics. They overcame this problem by pruning the features using an L1 regularised feature selection method before fine-tuning with conservative mean feature selection.

Sudha. A presented a stroke predicting model employingclassification approaches in 2012, with the help of her teachers N.Jaisankar and P Gayathra. For predicting the stroke using relevant attributes, they employed classification techniques such as Decision Tree, Naive Bayes, and Neural Networks. For dimension reduction, they used the principal component analysis technique. They investigated and evaluated sensitivity 7 accuracy indicators. The decision treehas a sensitivity of 95.29 percent and an accuracy of 98.01 percent. 8710 percent and 91.30 percent for the Bayesian classifier, respectively. They examined various approaches and determined that the decision tree was the most effective categorization method. The suggested model takes the patient's information and compares it to reduced characteristics (glucose level, blood pressure level, family history, etc.) to determine whether or not the patient has stroke illness. Sensitivity and specificity are used to determine accuracy. When compared to the other two classification strategies, the performance of neural networks was shown to be more accurate.

In 2013, the study was conducted in the Esfan Al-Zahra and Mashhad Ghaem hospitals between 2010 and 2011. They collected data on 807 healthy and unwell patients using a standard checklist that covers 50 risk factors for stroke, such as cardiovascular disease, diabetes, hyperlipidemia, smoking,and alcohol consumption. They used data mining techniques including k-nearest neighbor and C4.5 decision tree, as well as the WEKA tool, to evaluate the data. Age, gender, sleep duration, hours of exercise, hypertension, hyperlipidemia,smoking, alcohol, narcotics, stimulants, and additional risk variables that had not been examined previously were all retrieved using the C4.5 and k-nearest algorithms in WEKA

3.6 to evaluate stroke data. Finally, the C4.5 technique produced the best results, outperforming the k-nearest neighbor strategy in terms of accuracy, precision, and specificity by a little margin. As a consequence of its greater accuracy, the decision tree was chosen as the stroke- prediction approach.

In 2014, Hamed Asadi, Richard Dowling, Bernard Yan, and Peter Mitchell conducted a retrospective study on a database of suspected acute ischemic strokes. They looked at a variety of machine learning methods for predicting the effectiveness of cardiovascular interventions in patients with acute anterior circulation ischemic stroke. There were 107 people in the study who had a conservative acute anterior circulation Ischemic stroke and were treated with a cardiovascular strategy. The model incorporated all available data on demographic, procedural, and clinical aspects. They designeda supervised machine capable of categorising these variables into possible outcomes and terrible outcomes using SPSS, MATLAB, Rapid miner, classical statistics, artificial neural networks, and support vector methods. Using supervised machine learning, despite the limited dataset employed, there was promising accuracy, nearing 70% of predicting result.

They presented a powerful machine learning method that might improve the process of choosing between endovascular and medicinal therapy for acute stroke.

In 2015, Balar Khalid and Naji Abdelwahab proposed a data mining approach for predicting ischemic stroke using classification and logistic regression. They investigated what causes ischemic stroke. After that, the data was pre-processed,cleaned, and analysed with the WEKA 3.6 and C4.5 algorithms, as well as logistic regression. They analyze the sample data they had obtained using Microsoft "XLSTAT." The logistic regression model allowed them to analyse the linkbetween the occurrence of ischemic stroke and its risk factorsin their case study. The XLSTAT programme offers a high sensitivity of 77.58 percent and an 83 percent specificity. The ROC Curve evaluates specificity and sensitivity.

They discovered, however, that the prediction model had a 19.7% inaccuracy rate. To establish the parameters used in glaucoma prediction analysis, we offer an autonomous technique for segmenting the cup region from the optical disc (OD) region in RGB channels. A important step in calculating the cup-to- disc area ratio is correctly segmenting the cup region and the optical disc from the retinal fundus image (ACDR). This prediction technique was quantitatively evaluated on a publicly available database, with an accuracy of 83.168 percent.

In 2016, Ahmet Kadir, Cemi Colak, and Mehmet Ediz Sariham gathered data from Turget Ozal Medical Center, which comprised records of 112 healthy people and80 patients, as well as two target factors for applying Data Mining techniques to predict Ischemic stroke. Support Vector Machine (SVM), Stochastic Gradient Boosing (SGB), and Penalized Logistic Regression were employed (PLR).

They used a resampling procedure with a 10 fold cross validation. Area, Accuracy Sensitivity, specificity, positive predictive value, and negative predictive value were the performance assessment criteria under the RoC Curve (AUC). The study discovered that SVM performed the bestwhen compared to other models for predicting Ischemic stroke. SVM had a 95 percent confidence interval of 0.9789,SGB had a 95 percent confidence interval of 0.9737, and PLR had a 95 percent confidence interval of 0.8947.

The AUC values with 95 percent confidence intervals for SVM,SGB, and PLR were 0.9783, 0.9757, and 0.8953, respectively. SVM and SGB performed exceptionally well in predicting Ischemic stroke by approximating the circularoptic disc border with a circular model. The proposed technique is tested on two separate data sets, one local database and the other, MESSIDOR, a public database. Thesuggested optic disc segmentation approach achieved an average overlapping score of 99.45 percent and 99.47 percent between true OD and segmented OD for the local and public databases, respectively. Furthermore, success rates of 92.06 percent and 92 percent, respectively, were observed in both the local and public databases.

A group of Taiwanese medical university researchers developed a technique to automate the early identification ofischemic stroke in 2017. The CNN deep learning approach was used to create this model. They performed brain CT scans to assess whether stoke was a possibility.

The CT pictures were preprocessed by the technology to eliminate the impossible region where a stroke may occur. The patch photos were then chosen, and the number of patch images was increased using a data augmentation approach. They then used patch photos as input to train and test the convolutional neural network. They trained and tested aCNN module that could detect an ischemic stroke using 256patch images. It was determined that the proposed produced a result of above 90%.

A survey on AI applications in stroke was conducted in2018 by a team from Karnataka's National Institute of Engineering with the objective of properly predicting the start of stroke. They employed predictive algorithms and parameters to analyze these parameters, which included patient characteristics such as gender, age, height, BMW, andso on. To examine these factors, they created a data model using the decision tree approach. The findings were assessedusing the confusion matrix, and the accuracy was 95%. Theyachieved this by using a training model that facilitated in thecomparison of freshly given data with survey data. The report was prepared on the basis of this comparison. For the identification and segmentation of the optic disc in this work,an upgraded classification based on a hierarchical method was devised. The exact border of the optic disc can bedetermined by calculating the region of interest and using a new morphological transformation-based adaptive thresholding method. The suggested approach aids in the reduction of the process area required for segmentation techniques, resulting in significant performance improvements and a reduction in the amount of computing cost required for each retinal fundus picture. The proposed method was evaluated using publicly available retinal imaging data sets such DIARETDB1, DRIVE, HRF, DRIONS-DB, IDRiD, and STARE, and it outperformed earlier strategies in terms of accuracy and processing time.

In 2019, the Department of Computer Architecture and Automation at the Universidad Complutense de Madrid investigated whether state-of-the-art machine learning-based modelling methods could be used to test the hypothesis in collaboration with the hospital Universitario de La Princesa in Madrid, Spain. They investigated if non-invasive monitoring devices may aid in stroke type diagnosis.

These tools can even be used to anticipate future hazards, such as the patient's ultimate death. They collected data from 119 people's medical records, which included seven predictors and two target variables: stroke type prediction and death prediction. They used 7 different machine learning algorithmsto look at 6 different metrics, including Decision Tree, KNN,logistic regression; Nave Bayes, Neural Network, Random Forest, and Support Vector Machines. They also utilised a 10-fold cross validation re-sampling technique for both a guaranteed validation set from training one and the validationof the learnt classifier against any unknown sample. Random Forest models outperformed other techniques in terms ofstroke and mortality prediction, with values of 0.93+0.03 and0.970.01, respectively, when compared to other methods.

In 2019, JoonNyung Heo, Hyungjong Park, young Dae Kim, Hyo Suk Nam, and Ji Hoe Heo developed a model to predict Ischemic stroke long-term outcomes. They looked at the use of machine learning algorithms to predict ischemic stroke outcomes. They conducted a retrospective study usinga prospective cohort of people who had ischemic stroke. Deepneural networks, random forests, and logistic regression wereconstructed as machine learning models.

They compared all of their predictability after that. They also compared these models to an acute stroke registry and a Lausanne score analysis to see how accurate they were. There were 2604 patients in all, and 78 percent of them had positive results. Although the AUC curve of the deep neural network model was higher than the ASTRAL score, the AUC curves of the logistic regression and random forest models were not. Deep neural networks outperformed all other models, according to the results. It was more suited to forecasting outcomes.

## III. COMPARISON

| PAPER | YEAR | APROACH | DESCRIPTION |
|---|---|---|---|
| An IntegratedMachine learning Approach to Stroke Prediction | 2021 | L1 regularised logistic regression is an uniqueprediction approach that uses conservative mean feature selection. | In both AUC and Concordance index metrics, this technique exceeded the existing state-of-the-art. |
| Prediction and Control of Strokeby Data Mining. | 2020 | K-Nearest Neighbor and C4.5 DecisionTree Using WEKA are two data mining algorithms. | The accuracy of the C4.5 decision tree technique and K- Nearest Neighbor in stroke prediction was 95.42 percent and 94.18 percent,respectively. |
| Effective Analysis and Predictive Model of Stroke Diseaseusing Classification Methods | 2019 | Bayesian Classifier, Decision Tree,Neural Networks | The decision tree has a sensitivity of 95.29 percent and an accuracy of 98.01 percent. The Bayesian Classifier scored 87.10 percent and the Bayesian Classifier scored 91.30 percent, respectively. The results for neural networks were 94.82 percent and 97.87 percent, respectively. |
| Review on Machine Learning Approaches used for Stroke Prediction | 2019 | Decision Tree,Bayesian Classifier, Neural Networks | It demonstrated a promising accuracy of up to 70% in predictingresult. |
| Automated Ischemic Stroke Subtyping Based on Machine Learning Approach | 2018 | Pearson Correlation analysis and Shapiro-Wilkranking wereused. | Prediction of one type of the stroke is done in the proposed system, that is Ischemic Stroke. |

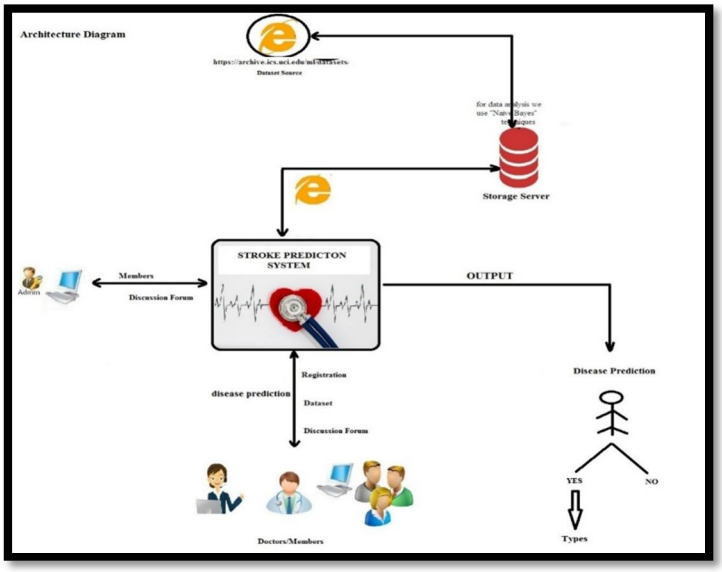
| | | | |
|---|---|---|---|
| Prediction of Brain Stroke Severity UsingMachine Learning | 2020 | Gaussian Naïve Bayes, Linear Regression & Logistic regression | The suggested model predicts stroke severity in three separate level factors: low risk,moderate risk, and high risk, based on essential attributes and acategorization technique. |
| Detection of Brain Stroke using Electroencephalography (EEG) | 2019 | EEG data to detect the type of strokeand region of lesion more accurate. | The only static data available on the UCI Machine Learning Repository was utilized for implementation. Data mining software was utilized. |
| The Use of Deep Learning to Predict StrokePatient Mortality | 2019 | DNN/scaled principal component analysis (PCA)to automation | With average values of 0.93+0.03 and 0.90.01, the Random Forest Model performed the best. |
| Machine Learning Approach toIdentify Stroke Within 4.5 Hours | 2018 | Logistic Regression, Support Vectormachine(SVM ) and RandomForest. . | An automated machine learning system can detect patients with acute ischemic stroke within 4.5 hours of the onset of symptoms. Machine learning techniques to choosing candidates for therapy in patients with an uncertain stroke start time may be practical and beneficial. |
| Expert SystemDetect Stroke with DempsterShafer Method | 2018 | Dempster Shafer technique withExpert System | In addition to consultinga doctor, this research might be a viable optionfor people. The dempster Shafertechnique will computethe patient's ischemic orhemorrhagic stroke if the patient is adequatelyinformed of the disease's symptoms. Knowing about the condition will help you deal with it more effectively. |

| | | | |
|---|---|---|---|
| Machine learning based model for prediction of outcomes in acute stroke | 2018 | Deep Neural Network, Random Forest.Logistic Regression | Deep neural networkshowed the highest accuracy. |
| An automatedEarly IschemicStroke DetectionSystem using CNN Deeplearning algorithm | 2018 | Computer- assisted identification of images, data augmentation, and convolutionalneural networks | It was extremely accurate,with a score of above 90%. |
| A Model for Predicting Ischemic Stroke Using DataMining Algorithms | 2016 | Data Mining, Classification, Logistic Regression, WEKA 3.6 | The results were obtainedwith the "XLSTAT" software. They showed the sensitivity of 77.58% and specificity of 83% |
| Stroke Predictionusing Decision trees in AI | 2015 | AI Decision Tree | The Decision tree algorithm showed the 95% accuracy in prediction of stroke |
| Different MedicalData Mining Approaches basedPrediction of Ischemic Stroke | 2015 | Some of the approaches employed includeSupport Vector Machine (SVM),Stochastic Gradient Boosting (SGB), and penalised logistic regression (PLR). | SVM, SGB, and PLR had AUC values of0.9783, 0.9757, and 0.853, respectively, with 95percent CI. |

## IV. METHODOLOGY

Step 1: Raw data Collected
Step 2: Extract and Segment Data (Data Preprocessing)Step 3: Train Data
Step 4: ML Technique for Disease Prediction
a. Machine Learning
b. Supervised Learning Technique
Step 5: Stroke Disease Prediction
Step 6: Results
Step 7: Visual Representation

Supervised Learning Technique is a predictive model used for the tasks where it involves prediction of one value using other values in the data-set. Supervised learning will have predefined labels. It classifies an object based on the parameters to one of the predefined set of labels. Depending of the requirement, labels, parameters and data-set we select the appropriate algorithm for predictions. Algorithm is used to build a model that makes predictions based on evidence in the presence of uncertainty. In this project, for prediction we make use to "KNN Algorithm and Naïve Bayes Algorithm" which is an efficient and works fine for all different sets of parameters. It also generates accurate results.



## V. REMARKS

*1)* Comparison to other models, SVM exhibited the greatest predictive performance, according to Ahmet Kadir Arslan et al. Comprehensive simulation, on the other hand, is required to generate a more precise and robust comparison.

*2)* Chiun-Li-Chin et al. trained the model with a small number of patch pictures (approximately 256), which reduced the system's efficiency. However, clinicians can utilize this suggested approach to identify illnesses successfully.

*3)* Aishwarya Roy and colleagues suggested a methodology to aid doctors in clinical trials. The dataset they utilised for the prediction model, as well as the approach they employed.

*4)* Luis Garcfa-Terriza et al. employed several methods to predict the kind of stroke (hemorrhagic vs. ischemic) and subsequent disease consequences. As a result, doctors will be able to utilize preventative therapies to avert adversities.

## VI. CONCLUSION

In a number of settings, machine learning algorithms havehelped predict stroke. Scenarios, data sets, parameters, andother analyses should all be considered when deciding which machine learning technique to use. We haven't beenable to agree on the best method for predicting strokes.

Every strategy has its own set of advantages and disadvantages. One of them should be chosen depending on the significance of the specific problem statement. To choose a certain approach or model, statistical analysis and initialization are required. However, because it produces promising results, random forest is one of the most common and powerful techniques for predicting a number from a data sample.

## REFERENCES

[1] Luis Garcfa-Terriza, Risco Martin, Ayala and Gemma Reig Rosello, “Comparison of different Machine Learning approaches to model stroke subtype classification and risk prediction”, Society for Modeling & Simulation International (SCS), 2019 April 29-May2.

[2] JoonNyung Heo, Jihoon G. Yoon, Hyungjong Park, Young Dae Kim, Hyo Suk Nam, Ji Hoe Heo, “Machine LearningBased Model for Prediction of Outcomes in AcuteStroke”,2019 February 1,doi:10.1161/strokeaha.118.024293

[3] Aishwarya Roy, Anwesh Kumar, Navin Kumar Singh and Shashank D, “Stroke Prediction using Decision Trees in Artificial Intelligence”, IJARIIT, Vol. 4, Issue 2, 2018, pp: 1636-1642

[4] Chiun-Li-Chin, Guei-Ru Wu, Bing-Jhang Lin, Tzu- Chieh Weng, Cheng-Shiun Yang, Rui-Cih Su and Yu-Jen Pan, “An Automated Early Ischemic Stroke Detection System using CNN Deep Learning Algorithm”, IEEE 8th International Conference on Awareness Science and Technology, 2017.

[5] Ahmet Kadir Arslan, Cemil Colak, Mehmet Ediz Sarihan, “Different medical data mining approaches based prediction of ischemic stroke”, Elsevier, Computer Methods and Programs in Biomedicine 2016 March 18.

[6] Balar Khalid and Naji Abdelwahab, “A model for predicting Ischemic stroke using Data Mining algorithms”,IJISET, Vol. 2 Issue 11, Nov 2015, ISSN: 2348-7968.

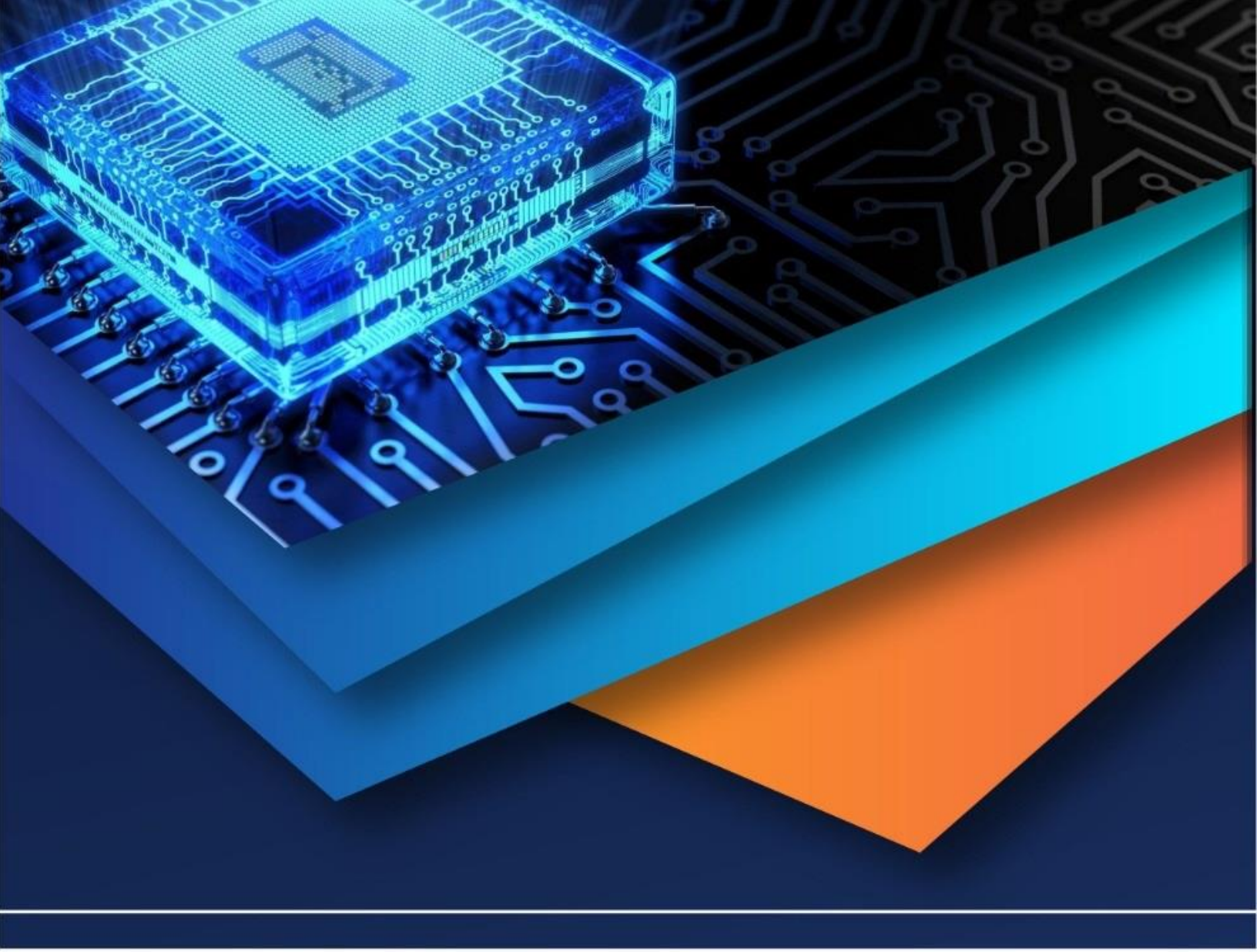
[7] A[.1S2u]dha, P. Gayathri, “Effective analysis & predictivemodel of stroke disease using classification methods”,

[8] Leila Amini, Reza, Rasul Norouzi & Associates, “Prediction and Control of Stroke by Data Mining”, IJPM, 8th Iranian Neurology Congress, Vol. 4, 23 Feb 2013.

[9] Aditya Khosla, Yu cao, Honglak Lee & Associates, “An integrated machine learning approach to stroke prediction”, 25-28 July 2010, Washington, DC, USA.

[10] Automated Ischemic Stroke Subtyping Based on Machine Learning Approach: IEEE Paper accepted on June4th 2020.

.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY