International Journal of Computing and Artificial Intelligence

**E-ISSN:** 2707-658X
**P-ISSN:** 2707-6571
Impact Factor (RJIF): 5.57
www.computersciencejournals.com/ijcai
IJCAI 2025; 6(2): 248-253
Received: 21-09-2025
Accepted: 30-10-2025

**Greeshma Madhukumari**
SRU, Bagar Rajput, Rajasthan, India

**Corresponding Author:**
**Greeshma Madhukumari**
SRU, Bagar Rajput, Rajasthan, India

# XAI-SDCNN Based brain stroke detection and risk factor identification using EBtrapP-FUZZY

**Greeshma Madhukumari**

**DOI:** https://www.doi.org/10.33545/27076571.2025.v6.i2c.204

**Abstract**
The brain is the most complex organ in the human body. Brain Stroke is a long-term disability disease that occurs all over the world and is the leading cause of death. A stroke occurs when the brain's blood supply is cut off and it ceases to function. Nevertheless, none of the traditional mechanisms focused on finding Brain Stroke along with Risk Factor Identification by combining Clinical Data and CT Image Data. Therefore, this paper proposes an innovative model named XAI-SDCNN-based brain stroke detection and risk factor identification using EBtrapP-FUZZY. Initially, the input clinical dataset is gathered and pre-processed for missing value imputation and one hot encoding, thus improving the data quality. Next, the data is augmented via the proposed QuanGutileSMOTE. Thereafter, the Clinical Features are extracted and will be given to the XAI-SDCNN. If the classified result is a stroke, then the corresponding CT brain image is taken and pre-processed for Noise removal and sharpening. Next, the CT image data is augmented via the geometric transformations. Thereafter, the lesion voxels are segmented using REWS-D2KMC. Next, from the segmented image, features are extracted, and these extracted features are given to REWSO for feature selection. After that, the reduced features are given to the XAI-SDCNN classifier to classify the different types of brain stroke. Finally, EBtrapP-FUZZY-based rules are generated to classify the risk factors.

**Keywords:** QuantileGuttman synthetic minority oversampling technique (QuanGutile SMOTE), explainable artificial intelligence sech deep convolutional neural networks (XAI-SDCNN), Weiner filter (WF), renyi the entropy white shark disper dennis KMEANS clustering (REWS-D2KMC), brain stroke classification, and Risk Factor

## Introduction
Stroke is classically characterized as a neurological deficit attributed to an acute focal injury of the central nervous system (CNS) by a vascular cause, including cerebral infarction, intracerebral hemorrhage (ICH), and subarachnoid hemorrhage (SAH). It is a significant cause of disability and death worldwide. Despite its global impact, the term "stroke" is not consistently defined in clinical practice, in clinical research, or in assessments of public health. Advances in basic science, neuropathology, and neuroimaging have improved the understanding of ischemia, infarction, and hemorrhage in the CNS.

A stroke is important in the CNS because it stops blood flow to a part of the brain, which causes acute, focal neurological deficits and brain cell death because the cells don't get enough oxygen and nutrients. This vascular injury causes serious and permanent brain damage, long-term disability, or death. Strokes are a major cause of death and disability worldwide. This shows how important the brain's vascular supply is and how important it is to get treatment right away to keep brain function (Brea *et al*., 2021) [12].

### Computerized tomography (CT)
CT is an important imaging technique for the early diagnosis, identification, and treatment of stroke, as it is fast, non-invasive, and extremely valuable in delineating ischemic and hemorrhagic strokes (Van Der Hoeven *et al*., 2015) [13]. Because a patient with stroke symptoms must be treated urgently, CT scanning is often the first imaging assessment performed, because it is fast and provides critical diagnostic information to identify possible brain injury and its type and extent. Although CT may appear normal early (within the first few hours) in ischemic strokes, it is crucial in ruling out bleeding, which precludes the use of clot-busting agents, such as tissue plasminogen activator (tPA). With hemorrhagic strokes, CT will often show the presence of bleeding within or adjacent to the brain and is essential

for determining medical or surgical interventions in these situations (Wu *et al*., 2019) [14]

CT image is mostly preferred by doctors because of its low cost and ability to identify brain strokes. Most of the traditional works used Machine Learning (ML) approaches, such as Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), neural network classifier, Naïve Bayes (NB), K-Nearest Neighbors (KNN), and Decision Tree (DT) for brain stroke classification. Likewise, the conventional works employed Deep Learning (DL) techniques like Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Artificial Neural Network (ANN), and Deep Neural Network (DNN) for classifying brain stroke. Additionally, some existing works used Visual Geometry Group-16 (VGG-16), Residual Network-50 (ResNet-50), and Auto Encoders (AE) to classify the brain stroke. Also, the risk factor of brain stroke is analyzed by using a Fuzzy Logic System (FLS) and an Adaptive Neuro-Fuzzy Inference System (ANFIS). However, improvement is still needed for the classification of brain stroke based on CT images. Therefore, an efficient model, which is based on XAI-SDCNN and EBtrap Pseudo Fuzzy, was used to superiorly classify brain stroke and to analyze the risk factors of brain stroke.

## 2. Literature Review

In this section, the aim, methods used, advantages, and limitations of the prevailing works related to brain stroke classification and its risk factor analysis are described.

**Table 1:** Existing works related to brain stroke classification and its risk factor analyzation

| Author's name | Aim | Methods | Advantages | Disadvantages |
|---|---|---|---|---|
| (Srinivas & Mosiganti, 2023) [5] | Brain stroke detection | Ensemble machine learning classifier integrated several individual classifiers, such as RF, extremely randomized trees, and histogram-based gradient boosting. | The presented soft voting model improved the accuracy and robustness of the brain stroke detection compared to the single classifier. | However, the model used unbalanced data; thus, the findings were inaccurate and the classification was ineffective. |
| (Gautam & Raman, 2021) [1] | Effective classification of brain hemorrhagic and ischemic stroke | Multi-focus image fusion and CNN | The CNN model effectively classified brain hemorrhagic and ischemic stroke with high accuracy. | Yet, the lesions were segmented by threshold values, which might consider the white matter as the lesion in segmented results, thus affecting the final outcomes. |
| (Soltanpour *et al*., 2021) [4] | Segmentation of automatic ischemic stroke lesions in CT perfusion maps | DNN-based MultiRes U-Net | The presented research efficiently identified the ischemic stroke lesions with highly varying scales and representations. | Owing to the high False Alarm Rate, the lesion was also considered as the tumor. |
| (Rao *et al*., 2022) [3] | Automatic prediction of hemorrhagic stroke using deep transfer learning based on CT images | Automated transfer deep learning method that combined ResNet-50 and the dense layer | The presented model perfectly diagnosed hemorrhagic stroke, and it was efficient to use as a clinical support tool in diagnosing brain strokes. | Furthermore, the presented research took a huge amount of time for training, which affected the effectiveness of the model. |
| (Raghavendra et al., 2021) [2] | Non-linear index for the automated detection of hemorrhagic brain stroke using CT images | Supervised learning methods, such as KNN, Probabilistic Neural Network, and SVM | The presented research achieved high classification accuracy in the identification of intracerebral hemorrhage stroke and classified the brain stroke as intracerebral hemorrhage or normal. | But, the research had computational complexity and interpretability issues. |

## 3. Problem Statement

Some of the existing related works have the following limitations,

- None of the works concentrate on finding Brain Stroke along with Risk Factor Identification by combining Clinical Data and CT Image Data.
- This work [Srinivas & Mosiganti, 2023] [5] used unbalanced data. If such unbalanced data is not dealt with properly, then the findings are inaccurate and the classification will be ineffective.
- In this paper [Gautam & Raman, 2021] [1], the lesions are segmented with the help of Threshold values. Thus, it is possible to consider the white matter is also the lesion in segmented results.
- In this work [Soltanpour *et al*., 2021] [4], the images are directly given to the model to find the lesions in the image. But, due to a high False Alarm Rate, the lesion is also considered as a tumor.
- Most of the existing work did not concentrate on the clinical data, which contains noise or missing values, therefore, such kind of data takes a reasonable amount of time to make it trainable. In image data do not concentrate on sharpening the image. Images are not sharpened, so some significant information may be lost.

## 4. Significance of the study

Strokes are one of the most important public health problems in the world because they cause a lot of deaths and disabilities. To come up with ways to prevent strokes, you need to know what causes them, such as high blood pressure, diabetes, smoking, not getting enough exercise, and family history. This research is essential for reducing these risks, identifying the risk in a timely manner, and, in the end, minimizing the occurrence of strokes.

By addressing these risk factors, lives may be saved, and the healthcare costs of stroke recovery and rehabilitation may be lowered.

The central contribution of this research is given as follows,

- To introduce novel pre-processing steps to transform the data into a format that is more easily and effectively processed to improve the data quality.

- To explain feature extraction, which is a way to change raw data into a set of useful, informative, and non-redundant features that can be used to make deep learning models work better.
- To implement effective data balancing, enabling the model to learn uniformly from all classes, thereby enhancing accuracy, precision, recall, and F1-score for each category, rather than solely the predominant one.
- To add new steps to the pre-processing stage that change the image data into a format that is easier and more efficient to work with in image processing. Noise removal and sharpening are important steps in image and signal processing that make data better and make later analyses or visualizations work better.
- To introduce efficient segmentation that divides brain CT images of stroke patients into voxels with lesions and those without.
- To introduce efficient risk factors, which efficiently analyze the risks for enhancing the interpretation of CT images by providing context regarding the likelihood of a stroke based on a patient's medical history.
- To compare the proposed technique with the existing technique using the result parameters like Rule Generation Time, Fuzzification Time, Defuzzification Time, Classification accuracy, precision, recall, f1-measure, Sensitivity, Specificity, False Positive Rate (FPR), False Negative Rate (FNR), True Negative Rate (TNR), Training Time, Dice score, Jaccard index, etc.

## 5. Objective of the study

The research objectives of the proposed brain stroke with risk factors are described further,

- To mitigate the underfitting problem in deep learning models, the Quantile Guttman Synthetic Minority Oversampling Technique (QuanGutile SMOTE) is proposed, which augments the data and increases the data size.
- The Renyi Entropy White Shark Disper Dennis Kmeans Clustering (REWS-D$^2$KMC) algorithm is introduced in this work to group the lesions' regions.
- To reduce the classifiers' complexity and increase their interpretability, the Cumulative distribution function Renyi Entropy White Shark Optimizer (RE-WSO) is presented to select the optimal features from the high dimensional attribute spaces.
- Explainable Artificial Intelligence Sech Deep Convolutional Neural Networks (XAI-SDCNN) classifier is utilized as a deep learning classification model in the proposed work to categorize the types of brain stroke.
- Exponential Bell Trapezoidal Pseudo Fuzzy (EBtrapP-FUZZY) is utilized in the risk factor classification of the image.

## 6. Hypothesis/research question

The hypothesis or research question that is familiar to the subject and helps to define an appropriate research question for a subject area or field of study is given as follows,

- What specific imaging characteristics on CT scans are most indicative of the type and severity of stroke?
- How do the characteristics of lesions identified on CT images differ between ischemic and hemorrhagic stroke patients?
- How can Deep learning techniques be applied to integrate clinical and imaging data for improved stroke prediction?
- What are the most significant clinical risk factors associated with the incidence of strokes in the study population?
- How do demographic factors (age, gender, and ethnicity) influence the risk profiles of patients experiencing strokes?
- How do changes in risk factors over time correlate with variations in lesion characteristics on CT scans?

## 7. Research Design

A stroke is a medical condition in which poor blood flow to the brain results in cell death. Nowadays, it is a leading cause of death all over the world. Several risk factors believed to be related to the cause of stroke have been found by inspecting the affected individuals. This research methodology proposed a novel (XAI-SDCNN)-based brain stroke detection with risk factor assessment. The proposed work comprises the following stages: preprocessing, data balancing, anonymous detection, feature extraction, feature selection, and classification of types of brain stroke, followed by risk factor identification.

The proposed system starts with the Clinical Dataset. It contains patient clinical information. Then, preprocessing is carried out to improve the data quality. Preprocessing steps are missing value imputation and one hot encoding. Then, the next step is to increase the amount of data by using the Quantile Guttman Synthetic Minority Oversampling Technique (QuanGutile SMOTE) algorithm. SMOTE (Synthetic Minority Oversampling Technique) is a data augmentation technique that generates synthetic examples for the minority class to mitigate the effects of class imbalance. However, SMOTE also has disadvantages, such as increased computational complexity. To solve that problem, this research methodology uses the QuanGutile function. After that, the features, such as age, hypertension, heart disease, smoking status, etc., are extracted from the data. Then, selected features are given as input to the Explainable Artificial Intelligence Sech Deep Convolutional Neural Networks (XAI-SDCNN) classifier to classify the result stroke or non-stroke.

In this scenario, if the classified result is a stroke, then the corresponding CT brain image is taken. Then, preprocessing is carried out to improve the image quality. Preprocessing steps, such as noise removal using a Weiner filter and sharpening are carried out. Then, the next step is to increase the amount of data by using the scaling, cropping, flipping, rotation, and translation of image augmentation techniques. Then, the anonymous detection in the CT image uses a two-step technique. In the first step, Gabor texture information is used to evaluate the probability of each voxel belonging to one of two groups, such as lesion or non-lesion voxels. Then, the lesion voxels are segmented by using Renyi the Entropy White Shark Disper Dennis Kmeans Clustering (REWS-D2KMC) algorithm. The core ideas and underpinnings of the WSO are inspired by the behaviors of great white sharks, including their exceptional senses of hearing and smell while navigating and foraging. The search agents of WSO randomly update their position in connection with best-so-far solutions, which leads to a premature convergence problem. To solve those problems, this research methodology uses the Renyi entropy function. Here, the maximum probability is considered as the fitness function. Gabor filters whose textures are computed on the non-lesion voxels. Then, the lesion voxels were segmented using Disper Dennis KMeans. The K-means algorithm is very helpful for the shape-based process, so this research uses K-means clustering. But, it is not guaranteed to find the optimal global solution, and some data points are not grouped into clusters and may fall into the same cluster group. The Euclidean distance cannot always find more accurate cluster boundaries. To solve this issue, DisperDennis distance is used. Further, from the resultant anonymous detection images, features, such as the Grey-Level Co-occurrence Matrix (GLCM), skewness, kurtosis, shape, size, etc. are extracted.

In the next phase, the extracted features are inputted into the Renyi entropy White Shark Optimizer (REWSO) to select the optimal features based on classification accuracy as a fitness measure. Then, the selected features are given as input to the Explainable Artificial Intelligence Sech Deep Convolutional Neural Networks (XAI-SDCNN) classifier to classify the different types of brain stroke, such as Ischemic and hemorrhagic. Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. In this work, the Shapely Additive Explanation (SHAP) explanation is included with SDCNN. DCNNs are a type of neural network that is commonly used for image analysis. They are designed to learn hierarchical representations of the data by using multiple layers of convolutional filters. Also, it uses the softmax activation function, which converts all logits into probabilities, and the sum of all probabilities will always be zero. This means that if the input doesn't belong to the class, then it will still give a result. To solve that problem, this research methodology uses the search activation function. Finally, the types of brain stroke are identified, and then Exponential Bell Trapezoidal Pseudo Fuzzy (EBtrapP-FUZZY) based rules are generated to classify the risk factors, namely low, moderate, and high. Risk factors identified by following the ABCD2 score are based on five parameters (age, blood pressure, clinical features, duration of TIA, and presence of diabetes). Scores for each item are added together to produce an overall result ranging between zero and seven. If the score is between "1-3", then it is denoted as "Low". If the score is between "4-5", then it is denoted as "Moderate".

If the score between "6-7", then it is denoted as "High". The Fuzzy algorithm is very easy and understandable, and it is capable of providing the most effective solution to complex issues. The Fuzzy algorithm presents tuning difficulty of membership functions and control rules. In order to solve this issue, the Exponential Bell Trapezoidal Pseudo membership function is used by the existing Fuzzy algorithm.

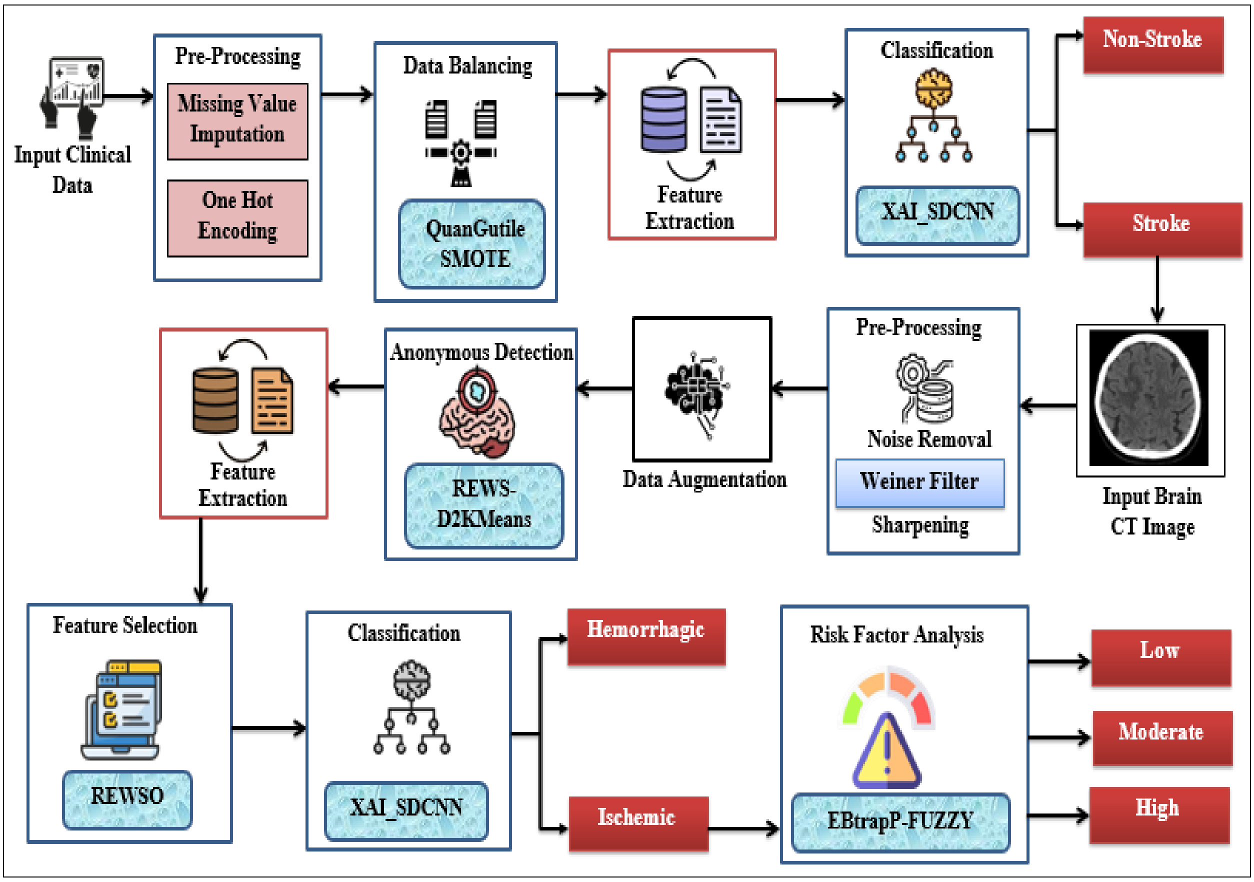


**Fig 1:** Block diagram of the proposed brain stroke classification model

## 8. Data Collection

### 8.1 Dataset Used

Some commonly used datasets for developing and evaluating XAI models for stroke detection include CT and MRI scans in addition to clinical data such as age, gender, stroke type, and outcome contained in a few datasets. Annotated MRI scans in the publicly available datasets, such as those for the ISLES challenges, serve as excellent training and testing data for models used in stroke segmentation, especially those focusing on XAI techniques like saliency mapping and feature attribution.

**8.2 Software Requirement**

The proposed work is implemented in the working platform of PYTHON. Python is a widely used general-purpose high-level programming language. It was mainly developed to emphasize code readability, and its syntax allows programmers to express concepts in fewer lines of code. Python is a programming language that lets you work quickly and integrate systems more efficiently. The virtual environment tool creates an isolated Python environment (in the form of a directory) that is completely separated from the system-wide Python environment.

**8.3 Hardware Requirements**

- **Processor:** Intel i5/core i7
- **CPU Speed:** 3.20 GHz
- **OS :** Windows 10
- **System Type:** 64 bit
- **RAM:** 8GB

**9. Proposed analysis**

**XAI-SDCNN-BASED BRAIN STROKE Detection and RISK Factor Identification**

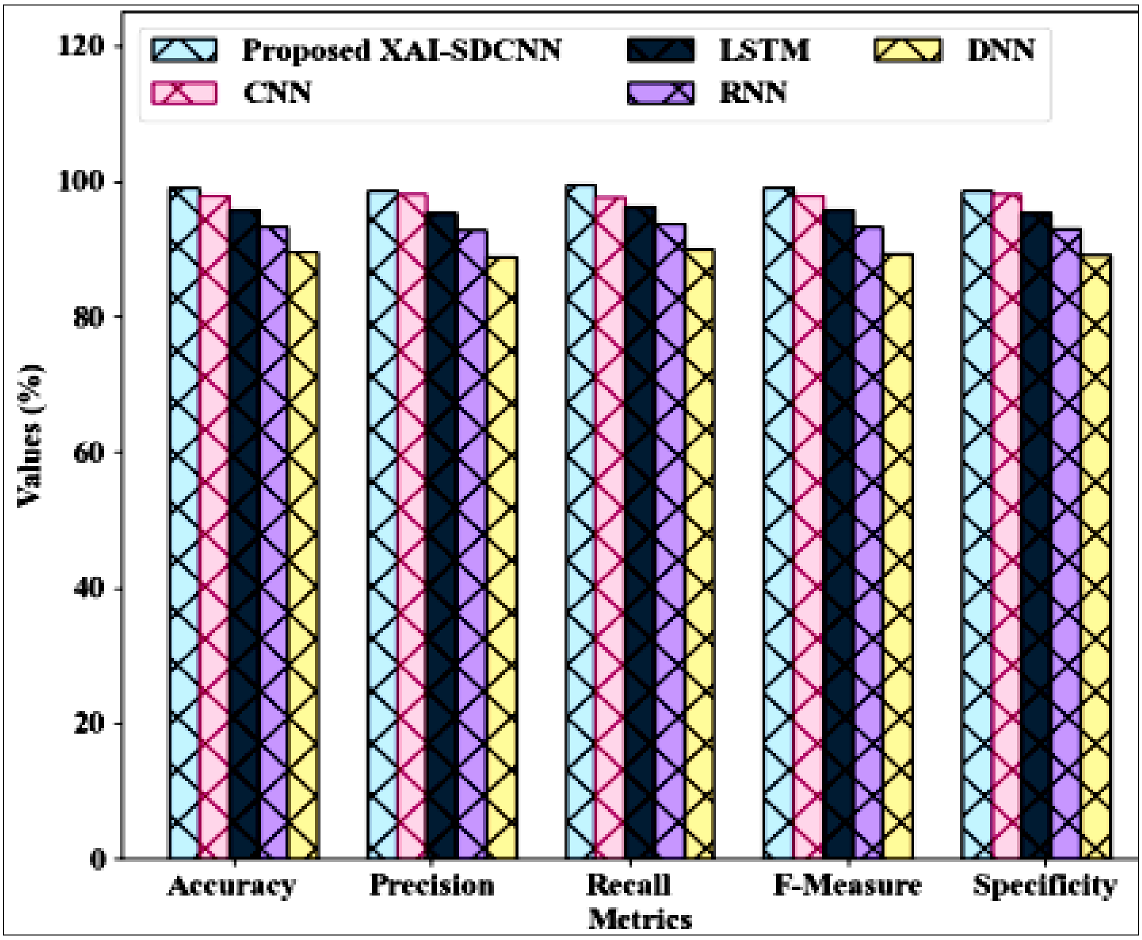


**Fig 2:** Brain Stroke Prediction Using Brain CT Image Dataset

The methodological process for classifying or identifying brian strokes and contributing risk factors is developed as a stepwise process, utilizing clinical data and CT images. In this solution the underlying clinical data whether incomplete, imbalanced or noisy is converted into a deterministic process to achieve reliable stroke calssification and risk analysis.

The clinical data can then be balanced using the Quantile Guttman Synthetic Minority Oversampling Technique SMOTE.

The proposed technique is evaluated based on the performance metrics, such as Rule Generation Time, Fuzzification Time, Defuzzification Time, Classification accuracy, precision, recall, f1-measure, Sensitivity, Specificity, False Positive Rate (FPR), False Negative Rate (FNR), True Negative Rate (TNR), Training Time, Dice score, Jaccard index, Fitness Vs Iteration, Feature Selection Time. These metrics evaluate the classification performance by comparing the proposed model with some other existing techniques like Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Deep Belief Network (DBN), and Deep Neural Networks (DNN).

**10. References**

1. Gautam A, Raman B. Towards effective classification of brain hemorrhagic and ischemic stroke using CNN. Biomedical Signal Processing and Control. 2021;63:1-13. https://doi.org/10.1016/j.bspc.2020.102178
2. Raghavendra U, Pham TH, Gudigar A, Vidhya V, Rao BN, Sabut S, *et al*. Novel and accurate non-linear index for the automated detection of haemorrhagic brain stroke using CT images. Complex and Intelligent Systems. 2021;7:929-940. https://doi.org/10.1007/s40747-020-00257-x

3. Rao BN, Mohanty S, Sen K, Acharya UR, Cheong KH, Sabut S. Deep transfer learning for automatic prediction of hemorrhagic stroke on CT images. Computational and Mathematical Methods in Medicine. 2022;2022:1-10. https://doi.org/10.1155/2022/3560507
4. Soltanpour M, Greiner R, Boulanger P, Buck B. Improvement of automatic ischemic stroke lesion segmentation in CT perfusion maps using a learned deep neural network. Computers in Biology and Medicine. 2021;137:1-8. https://doi.org/10.1016/j.compbiomed.2021.104849
5. Srinivas A, Mosiganti JP. A brain stroke detection model using soft voting-based ensemble machine learning classifier. Measurement: Sensors. 2023;29:1-7. https://doi.org/10.1016/j.measen.2023.100871
6. Holzinger A. From machine learning to explainable AI. In: IEEE World Symposium on Digital Intelligence for Systems and Machines (DISA 2018). 2018. p. 55-66. https://doi.org/10.1109/DISA.2018.8490530
7. Hossain MM, Ahmed MM, Rakib MRH, Zia MO, Hasan R, Islam MR, *et al*. Optimizing stroke risk prediction: a primary dataset-driven ensemble classifier with explainable artificial intelligence. Health Science Reports. 2025;8(5):1-15. https://doi.org/10.1002/hsr2.70799
8. Hossain MS, Saha S, Paul LC, Azim R, Al Suman A. Ischemic brain stroke detection from MRI image using logistic regression classifier. In: International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). 2021. p. 763-767. https://doi.org/10.1109/ICREST51555.2021.9331090
9. Hossain S, Bhuiyan AS, Rusho MA, Afride MR, Azad RI, Rahi AM, *et al*. Automated stroke prediction using machine learning approach. In: 5th International Conference on Electrical, Computer and Energy Technologies (ICECET 2025). 2025. p. 1-6. Available from: https://hal.science/hal-05068201/document
10. Hu N, Zhang T, Wu Y, Tang B, Li M, Song B, *et al*. Detecting brain lesions in suspected acute ischemic stroke with CT-based synthetic MRI using generative adversarial networks. Annals of Translational Medicine. 2022;10(2):1-16. https://doi.org/10.21037/atm-21-4056
11. Hussain I, Jany R. Interpreting stroke-impaired electromyography patterns through explainable artificial intelligence. Sensors. 2024;24(5):1-18. https://doi.org/10.3390/s24051392
12. Brea-Gomez B, Torres-Sanchez I, Ortiz-Rubio A, Calvache-Mateo A, Cabrera-Martos I, Lopez-Lopez L, Valenza MC. Virtual reality in the treatment of adults with chronic low back pain: a systematic review and meta-analysis of randomized clinical trials. International journal of environmental research and public health. 2021 Nov 11;18(22):11806.
13. Vennix S, Musters GD, Mulder IM, Swank HA, Consten EC, Belgers EH, van Geloven AA, Gerhards MF, Govaert MJ, Van Grevenstein WM, Hoofwijk AG. Laparoscopic peritoneal lavage or sigmoidectomy for perforated diverticulitis with purulent peritonitis: a multicentre, parallel-group, randomised, open-label trial. The Lancet. 2015 Sep 26;386(10000):1269-77.
14. Wu Z, Shen C, Van Den Hengel A. Wider or deeper: Revisiting the resnet model for visual recognition. Pattern recognition. 2019 Jun 1;90:119-33.