

Brain Stroke Prediction using Deep Learning

S. Vishal Kumar¹, B. Anil Kumar², C. Sai Shiva Datta³, Dr. M. Poonkodi⁴

Department of Computer Science & Engineering^{1,2,3,4}

SRM Institute of Science and Technology, Vadapalani Chennai, India

Abstract: *A brain stroke is a potentially fatal medical disorder that develops when the brain's blood supply is suddenly cut off. Early detection and prevention are essential for enhancing patient outcomes because it is a primary cause of disability and mortality worldwide. With the development of artificial intelligence and machine learning, there is an increasing interest in using these technologies to create brain stroke prediction models. In this paper, we suggest a deep learning-based method for forecasting brain strokes. Our strategy is based on the architecture of convolutional neural networks (CNN) and recurrent neural networks (RNN). While the RNN analyses the patient's demographics, medical history, and test results, the CNN is utilized to extract features from medical pictures like computed tomography (CT) or magnetic resonance imaging (MRI) scans. The model is trained using a substantial dataset of patient records, including both those of patients who have had and have not had brain strokes. Our results imply that the deep learning-based strategy that has been described can be a useful tool for the early detection and prevention of brain stroke. Healthcare providers can take proactive steps to stop the disease by identifying people who are at high risk of having a brain stroke. Additionally, our method may be incorporated with clinical decision-making systems to offer in-the-moment forecasts and suggestions for patient care. As a result, our study shows the potential of deep learning methodologies in creating precise and trustworthy models for brain stroke prediction. Future research might examine the interpretability of the model's predictions as well as the generalizability of our model across various patient populations and data sources*

Keywords: Deep Learning, CNN, RNN, Early detection of stroke, clinical decision making

I. INTRODUCTION

A vascular problem that causes immediate focused harm to the central nervous system in the case of a stroke results in a neurological disability. It is a significant global cause of death and disability [1]. A stroke has been suffered by roughly 7 million Americans over the age of 20 and the total prevalence of stroke in the US is believed to be 2.5%. The disorder has a serious detrimental effect on the health and quality of life of sufferers. It is also thought to have a detrimental effect on hospital services and bed availability, costing the US economy roughly \$351.2 billion between 2014 and 2015 [2]. Ischemic and hemorrhagic strokes are the two different forms. A burst blood artery causes a hemorrhagic stroke, which results in brain haemorrhage, whereas an obstruction in the brain's arteries causes an ischemic stroke. The majority of strokes, between 85% and 90%, are ischemic strokes [3]... Numerous risk factors associated with lifestyle exist, such as obesity, nutrition, alcohol consumption, and physical inactivity [4]. Stroke can also be caused by underlying illnesses such diabetes, hypertension, and cardiovascular disorders. So, good self-management of these conditions and the pursuit of a healthy lifestyle may stop strokes from happening.

II. LITERATURE SURVEY

Brain Stroke Prediction Using Machine Learning Approach by DR. AMOL K. KADAM, PRIYANKA AGARWAL (2022) a stroke is a condition that harms a person by rupturing the veins in their brain. Additionally, a blood clot or other types of disruption in the blood flow can cause a stroke. Vitamins for the mind. According to the WHO, the World Health Organisation, stroke is one of the leading causes of death and disability worldwide. The majority of research has been done on predicting heart attacks, but few studies have examined the possibility of a cerebral stroke. After that, the AITo predict the likelihood of a cerebral stroke, models are developed. The initiative aims to distinguish between victims' acquaintance with stroke risk and the factors that influence it. For accurate prediction, the study has used ML techniques like Logistic Regression, Decision Tree Classification, Random Forest Classification, KNN, and SVM.[2] Brain Stroke Prediction using Machine Learning by Mrs. Neha Saxena, Mr. Deep Singh Bhamra, Mr. Arvind

Choudhary(2014).A stroke, commonly referred to as a cerebro-vascular accident (CVA), occurs when The part of the body that the blood-deprived brain cells control ceases functioning as the brain loses its blood supply. This blood supply loss may be ischemic. Either due to a lack of blood supply or hemorrhagic bleeding into the brain. Because strokes can cause lasting disability or death, they are a medical emergency. Ischemic strokes can be treated; however this treatment must begin within the first few hours after 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. In order to reduce the danger of a future stroke, this circumstance also requires emergency assessment. A stroke would be categorised as a TIA by definition if all symptoms disappeared within 24 hours. According to the World Health Organisation (WHO), stroke is the second most common cause of death worldwide, accounting for around 11% of all fatalities. Our machine learning algorithm analyses data to predict whether a patient will survive a stroke based on input characteristics like gender, age, and several other factors. Our machine learning algorithm analyses data to predict whether a patient will survive a stroke based on input characteristics like gender, age, and several other factors. Illnesses and smoking history. Our dataset, in contrast to most others, concentrates on characteristics that would be significant risk factors for a brain stroke.[3]

Machine Learning for Brain Stroke: A Review by Manisha Sanjay Sirsat, Eduardo Ferme, and Joana Camara (2020) Machine learning (ML) offers a quick and accurate prediction result, and it has developed into a potent tool in healthcare settings, providing stroke patients with individualised therapeutic care. Although there is a definite need for study, some research fields are not receiving adequate attention for scientific exploration despite the expanding use of ML and Deep Learning in healthcare. Therefore, the purpose of this work is to carefully review papers for each category after classifying state-of-the-art ML approaches for brain stroke into 4 groups based on their functionalities

Results of the Science Direct web scientific database on ML for brain stroke from 2007 to 2019 revealed a total of 39 studies. 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. The current study highlights the value of various ML techniques used in brain stroke.[4]

Predicting Stroke Risk with an Interpretable Classifier by SERGIO PEÑAFIEL, NELSON BALOIAN, HORACIO SANSON, AND JOSÉ A. PINO(2021)As a common illness with strong evidence that early awareness of having that risk might be helpful for prevention and treatment, predicting an individual's risk of having a stroke has been a research topic for many authors worldwide. Many governments have begun gathering health information about their own citizens in order to use artificial intelligence techniques to make those forecasts. The most precise ones rely on so-called black-box techniques, which provide little to no explanation as to why they make a particular prediction. However, in the medical area, explanations are perhaps more crucial than accuracy since they give specialists knowledge of the variables that affect the risk level. Additionally, missing data in medical records is frequently encountered. In this study, we propose the creation of a prediction approach that not only performs better than some other ones already in use, but also provides details on the most likely reasons for a high stroke risk and can handle records with missing data. Its foundation is the Dumpster-Shafer plausibility theory. We conducted the tests using information from the Okayama regional hospital in Japan, where it is required by law for people to get annual physicals.

The tests presented in this article compare the Dempster-Shafer method's findings to those attained using Multilayer Perceptron, Support Vector Machines, and Naive Bayes, among other well-known machine learning techniques. In these studies with some missing data, our method worked best. It also provides a critique of how the classification method interprets the rules that are established. Both medical literature and human experts supported the rules.[5]

Towards stroke prediction using electronic health records by Douglas Teoh(2014) Background: In Japan, stroke is the fourth most common cause of death as of 2014. Future stroke diagnoses could be predicted, which would make it easier to implement preventative healthcare interventions. Within a year after the patient's most recent set of test results or medical diagnoses, our goal is to forecast a stroke diagnosis. Methods: The Tsuyama Jifukai Tsuyama Chuo Hospital in Japan donated about 8000 computerized health records. Non-homogeneous temporal data from these recordings was first converted into an algorithmic-usable form. The modified data were fed into several neural network architectures in

order to assess the effectiveness of the given data as well as the networks' capacity to take advantage of any relationships that underlie the data. Due to the high occurrence of stroke cases, class outputs were unbalanced, which led trained neural network models to be biased towards making negative predictions. We created and added regularization terms to the typical cross-entropy loss function to address this problem. False positive and false negative predictions were punished under these terms. We used Receiver Operating Characteristic to assess how well our trained models performed. Results: Using a dual-input topology, the best neural network integrated and aggregated the many sources of temporal data. Area under the Receiver Operating Characteristic curve for this network was 0.669. When compared to the conventional cross-entropy loss function, the customized regularization terms had a favorable impact on the training procedure.[6]

Machine Learning Algorithm for Stroke Disease Classification by Prof. Vishal Shinde, Mr.NikhilPatil, Mr.Omkar Sonar, Mr.NishantShedage. (2022). Stroke and obesity are the two main causes of death in diverse nations. By enhancing picture quality to increase image effects and noise reduction and using ML algorithms to split patient images into two kinds of stroke, ischemic stroke and stroke hemorrhage, this study analyzes CT scan image enhancement data for stroke patients. Eight machine learning (ML) methods, including Naive Bayes, K-Nearest Neighbours, Logistic Regression, Random Forest, Decision Tree, Multi-layer Perceptron (MLP-NN), Vector Support, and Deep Learning Machine, are used in this work to classify stroke. According to our findings, Random Forest generates the highest levels of precision (95.97%), accuracy (94.39%), recall (96.12%), and f1-Measures (95.39%).[7]

Prediction of Brain Stroke Severity Using Machine Learning by Vamsi Bandi1, Debnath Bhattacharyya, Divya Midhun chakkravarthy (2020). Since they affect the central nervous system, strokes have become one of the main causes of death in recent years. Ischemic and hemorrhagic strokes cause the most significant damage to the central nervous system. Globally, 3% of people suffer from subarachnoid haemorrhage, 10% from intracerebral haemorrhage, and 87% from ischemic stroke, according to the World Health Organisation (WHO). In this study, stroke is identified, classified, and predicted using medical data using machine learning approaches. The ability of the current studies to identify risk factors for different types of strokes is limited. This issue is addressed by the Stroke Prediction (SPN) method, which uses the improvised random forest to analyse the degrees of risks associated with strokes. When compared to the previous models, this research using machine learning techniques of the Stroke Predictor[8]

Deep Learning and Machine Learning for Early Detection of Stroke and Haemorrhage by ZeyadGhaleb Al-Mekhlafi, Ebrahim Mohammed Senan, Taha H. Rassem, Badia Abdulkarem Mohammed, Nasrin M. Makbol, Adwan Alownie Alanazi1, Tariq

S. Almurayziq1 and Fuad A. Ghaleb. (2021). Stroke and cerebral haemorrhage are the second leading cause of death in the world after ischemic heart disease. In this work, a dataset containing medical, physiological and environmental tests for stroke was used to evaluate the efficacy of machine learning, deep learning and a hybrid technique between deep learning and machine learning on the Magnetic Resonance Imaging (MRI) dataset for cerebral haemorrhage. In the first dataset (medical records), two features, namely, diabetes and obesity, were created on the basis of the values of the corresponding features. The t-Distributed Stochastic Neighbour embedding algorithm was applied to represent the high-dimensional dataset in a low-dimensional data space. Meanwhile, the Recursive Feature Elimination algorithm (RFE) was applied to rank the features according to priority and their correlation to the target feature and to remove the unimportant features. Support Vector Machine (SVM), K Nearest Neighbours (KNN), Decision Tree, Random Forest, and Multilayer Perceptron are just a few of the categorization algorithms that receive input from the features. The Precision, Recall, and F1 scores for this algorithm's classification of stroke cases were 98%, 100%, and 99%, respectively. The MRI image dataset from the second dataset was assessed using a hybrid Alex Net+SVM model and Alex Net model. In comparison to the Alex Net model, the hybrid model Alex Net+SVM outperformed it; it achieved accuracy.[9]

Stroke Risk Prediction with Machine Learning Techniques by Elias Dritsas and Maria Trigk. (2022). When blood flow to a portion of the brain is suddenly cut off, a stroke is the result. Depending on the region of the brain, impairment results from a lack of blood flow since the brain cells gradually die. Brain is impacted. Early diagnosis of symptoms can be extremely helpful in predicting stroke and supporting a healthy lifestyle. In this study, a number of models are created and assessed using machine learning (ML) in order to construct a solid framework for the long-term risk

prediction of stroke occurrence. This study's key contribution is a stacking technique that performs well and is validated by a variety of measures, including AUC, precision, recall, F-measure, and accuracy..[10]

Analysing the Performance of Stroke Prediction using ML Classification Algorithms by Gangavarapu Sailasya, Gorli L ArunaKumari. (2021). an illness known as a stroke damages the brain by rupturing the blood arteries in the brain. It can also happen when the passage of blood and other nutrients to the brain is interrupted. The World Health Organisation (WHO) cites stroke as the primary global cause of death and disability. The majority of research has been done on heart stroke prediction, but very little research has been done on brain stroke risk. With this in mind, numerous machine learning models are created to forecast the likelihood of a brain stroke. In order to train five different models for precise prediction, this paper took a variety of physiological factors and used machine learning algorithms like Logistic Regression, Decision Tree Classification, Random Forest Classification, K-Nearest Neighbours, Support Vector Machine, and Naive Bayes Classification. The algorithm with the highest accuracy for this task was Naive Bayes, which had a result of about 82%.

III. METHODOLOGY

Pre-processing and Model Training (CNN): The dataset is pre-processed, including image scaling, reshaping, and array form conversion. On the test image, the same processing is likewise carried out. Any image from a dataset made up of two or more different airplanes can be used as a test image for the software.

Le Net Architecture:

Yann Le Cun created the Le Net convolutional neural network (CNN) architecture in the 1990s for image identification applications. Seven layers make up the design, comprising three fully linked layers, two convolutional layers, and two sub sampling layers. The Le Net architecture can be altered and tailored to the particular job at hand for deep learning-based brain stroke prediction. A brain image, such as one from a CT or MRI scan, would be the network's input, and its output would be a forecast of the patient's chance of having a stroke. A collection of filters are applied to the input image by the convolutional layer of the Le Net architecture in order to extract features. The feature maps created by the first layer are sub sampled in the second layer, which lowers their spatial resolution. With another set of convolutional and sub sampling layers, the procedure is repeated. Le Net's final three layers are fully connected layers that translate the retrieved features to the output. The first completely connected layer serves as a feature extractor, and the final prediction is made by combining the extracted features in the second fully connected layer. A single neuron with a sigmoid activation function serves as the output layer, producing a number between 0 and 1 that represents the likelihood of having a stroke. A dataset of brain scans with labels indicating whether or not each image is at danger of a stroke can be used to train the LeNet architecture. To reduce the discrepancy between the expected and real labels, the network's filter and neuron weights are changed during training. Overall, the LeNet architecture can be a great starting point for creating a deep learning model for predicting brain strokes, although adjustments and tweaks might be required to maximize the network's performance for this particular application

System Architecture

All the entities currently integrated into the system are succinctly and clearly described in this graphic. This is a graphic representation of the entire procedure and the manner in which it was carried out. The dataset is initially obtained and then examined. In other words, how many photos fall into each of the following five severity categories: No DR, Mild, Moderate, Severe, and Proliferate. The data is then divided into a train set and a test set. The chosen model is built with the data as input. The CNNs process this data after that. Data is processed to ensure consistency and make it usable. It is polished, arranged, or adjusted. Data processing employs a variety of methods. The Edge Zero Padding is a technique used to ensure that the aspect ratio of the image is not altered. By redistributing light and dark pixels in photos, adaptive histogram equalization improves representation of the contrast of the original image. By reducing sharp edges and filling in holes or gaps in the item, morphological dilation improves the accuracy of object detection. In addition to what has already been said, the data is subjected to median filtering, image blurring, masking, and erosion. To train the model, image segmentation and feature extraction are performed. Accuracy, recall, precision, True Positive, True Negative, False Positive, and False Negative predictions are the criteria used to evaluate each model. Then, based on

the model's training score, test score, rate of misclassification, training time, and prediction time; it is compared to other models. The optimum algorithm for DR prediction is determined based on the observation of these indicators

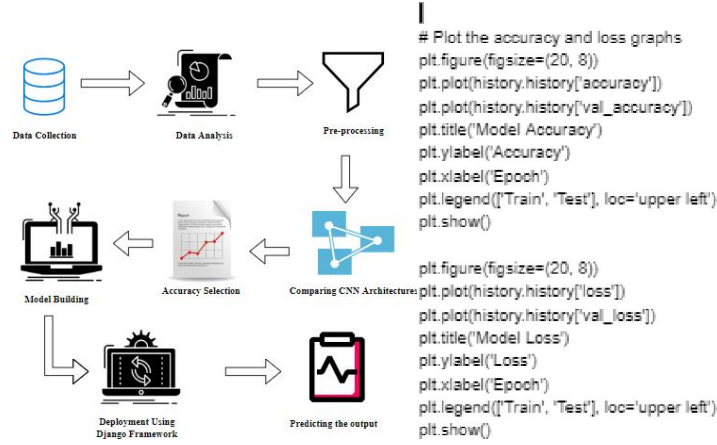


Fig 1: System Architecture

Pseudocode:

```

import necessary libraries and modules

# Define the training and testing data generators
train_datagen = ImageDataGenerator(rescale=1./255, shear_range=0.2, zoom_range=0.2,
horizontal_flip=True)
test_datagen = ImageDataGenerator(rescale=1./255)

training_set = train_datagen.flow_from_directory('datasets/train', target_size=(224, 224),
batch_size=32, class_mode='categorical')
test_set = test_datagen.flow_from_directory('datasets/test', target_size=(224, 224),
batch_size=32, class_mode='categorical')

# Define the model architecture
Classifier = Sequential()
Classifier.add(Convolution2D(32, 3, 3, input_shape=(224, 224, 3), activation='relu'))
Classifier.add(MaxPooling2D(pool_size=(2, 2)))
Classifier.add(Convolution2D(128, 3, 3, activation='relu'))
Classifier.add(MaxPooling2D(pool_size=(2, 2)))
Classifier.add(Flatten())
Classifier.add(Dense(256, activation='relu'))
Classifier.add(Dense(2, activation='softmax'))

Classifier.compile(optimizer='rmsprop', loss='categorical_crossentropy', metrics=['accuracy'])
Classifier.summary()

# Set the file path and checkpoint
model_path = "LeNet1.h5"
callbacks = [ModelCheckpoint(model_path, monitor='accuracy', verbose=1,
save_best_only=True)]

# Train the model
epochs = 100
batch_size = 32

history = Classifier.fit(training_set, steps_per_epoch=training_set.samples // batch_size,
epochs=epochs, validation_data=test_set,
validation_steps=test_set.samples // batch_size,
callbacks=callbacks)

```

Fig 2.1

IV. EXPERIMENTAL ANALYSIS

DATASET:

This dataset was obtained from a kaggle website that had image records from which characteristics had been extracted and divided into train and test data. Usually, training data and test data are used separately. The model learns on this data in order to later generalize to other data because the training set comprises a known outcome. It will use the Tens or flow library in Python and the Keas technique to test our models using the test dataset (or subset). Using the keas pre-

processing image data generator tool, we must input our data collection. Additionally, we develop size, rescale, range, zoom range, and horizontal flip. Then, using the data generator tool, we import our image dataset from the folder. Train, test, and validation are configured here. We also set the class-mode, goal size, and batch size. From this function, we must train our own network by including CNN layers. We create training steps for each epoch, total epochs, validation data, and validation steps, then use this data to train our dataset using a classifier and fit generator function.

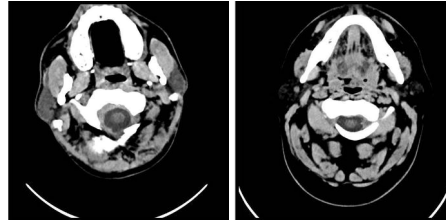


Fig 3.1 : Train(A) Fig 3.2: Test(B)

**Analysis on Dataset
Confusion Matrix:**

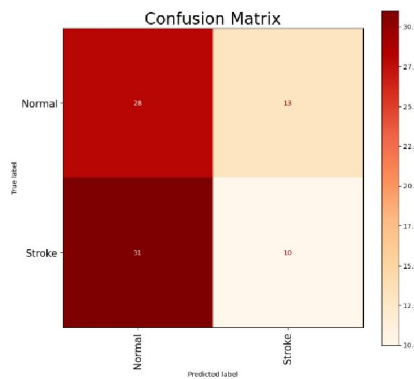


Fig 3.1.1

Graph:

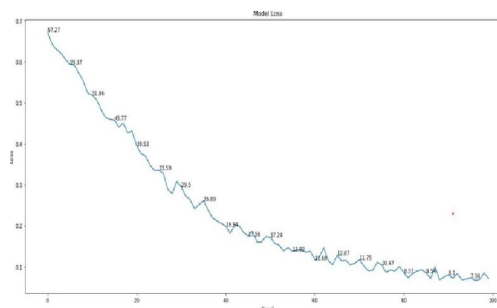


Fig 3.2.1: Model Accuracy

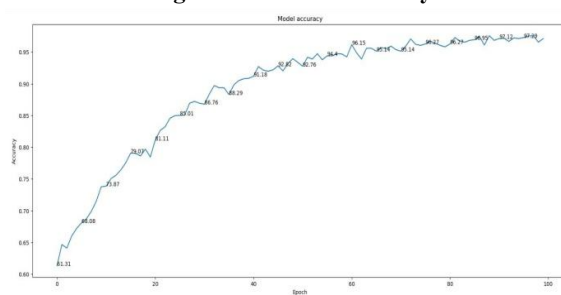


Fig 3.2.2: Model Loss

PERFORMANCE ANALYSIS:

Metrics like accuracy, precision, recall, F1-score, and support can be used to evaluate performance analysis for brain stroke prediction using deep learning.

A deep learning model's accuracy is measured as a percentage of properly predicted cases out of all cases. Recall quantifies the proportion of genuine positives among all real positives, whereas precision quantifies the proportion of true positives among all projected positives. The F1-score, which provides a general indicator of the model's accuracy, is the harmonic mean of precision and recall.

	Precision	Recall	F1-score	Support
Normal	0.47	0.68	0.56	41
Stroke	0.43	0.24	0.31	41
Accuracy			0.46	82
Macro Avg.	0.45	0.46	0.44	82
Weighted Avg.	0.45	0.46	0.44	82

VI. RESULTS

Through thorough testing of the suggested models, we arrive at the conclusion that, on average, CNN outperforms its competitors when it comes to Image classification for Data retinoscopy, achieving 96% accuracy, which will help us identify the symptoms in the early stages and potentially prevent the loss of vision.

VII. CONCLUSION AND FUTURE WORK

This project involved developing a research study to classify brain strokes using deep learning methods. This is a challenging issue that has already been addressed multiple times using various methods. Although feature engineering has produced effective outcomes, this study concentrated on feature learning, one of DL's promises. Although it is not required, feature engineering, image pre-processing improve classification accuracy. As a result, it lessens input data noise. These days, feature engineering is used in brain stroke detection software. Due to a significant constraint, a solution entirely based on feature learning does not appear to be close yet. Deep learning techniques could therefore be used to classify brain strokes.

More optimisation of hyper parameters and model elements, such as which layers to freeze versus make trainable during transfer learning, will be done in future work. A thorough search for hyper parameters would have been a more empirical method, however due to time and computer resource limitations, most model implementation decisions were chosen by looking at the convergence of the model and relative metrics from training versus validation.

REFERENCES

- [1]. J. Zhao, L. Han, H. Zhang, et al."A Deep Learning Approach for Stroke Detection in Brain CT Imaging"(2018)
- [2]. F. L. Ng, K. K. Li, T. S. Chan: "Automatic Detection of Acute Ischemic Stroke in CT Scans Using Convolutional Neural Networks".(2017)
- [3]. G. Bivona, G. Criscione, F. La Torre, et al." Deep Learning for Automatic Detection of Ischemic Stroke Lesions Computed Tomography Angiography".(2019)
- [4]. Y. Zhang, J. Lu, M. Zhou, et al.:"Deep Learning-based Classification of Hemorrhagic and Ischemic Strokes Using Multi parametric MRI".(2019)
- [5]. J. Cao, X. ChenJ. Chen, et al. : "Convolutional Neural Network-Based Automatic Hemorrhagic Stroke Detection and Segmentation"(2019)

- [6]. Immanuel, R. R., & Sangeetha, S. (2023). Identifying Different Emotions Of Human Using Eeg Signals Using Deep Learning Techniques. *Journal of Theoretical and Applied Information Technology*, 101(18).
- [7]. S. Chen, W. Huang, X. Wei, et al.:"Deep Learning for Prediction of Stroke Outcome: An MRI-Based Multi-Task Learning Approach".. (2020)
- [8]. Zhang, Z. Ma, Y. Liu, et al.: "Combining Radiomics and Deep Learning for Stroke Outcome Prediction".(2020)
- [9]. Y. Li, D. Wang, J. Li, et al. : "Automated Detection of Acute Ischemic Stroke Lesions in Non-Contrast CT Images Using Deep Convolutional Neural Networks"(2020)
- [10]. Y. Liu, C. Shi, K. Wang, et al. : "Deep Learning-Based Prediction of Hemorrhagic Stroke Outcome"(2020)
- [11]. M. Zhou, Y. Zhang, J. Lu, et al. : "Deep Learning-Based Detection and Segmentation of Acute Ischemic Stroke Lesions in CT Images"(2020)
- [12]. Singh, M.S., Choudhary, P., Thongam, K.: "A comparative analysis for various stroke prediction techniques. In: Springer, Singapore "(2020).
- [13]. Vamsi Bandi, Debnath Bhattacharyya, Divya Midhun chakkravarthy: "Prediction of Brain Stroke Severity Using Machine Learning. In: International Information and Engineering Technology Association".(2020)
- [14]. N wosu, C. S Dev, S. Bhardwaj, P. Veeravalli, B. John, "Predicting stroke from electronic health records. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE".(2019).
- [15]. Immanuel, R. R., &Sangeetha, S. K. B. (2023). Implementation of an Automatic EEG Feature Extraction with Gated Recurrent Neural Network for Emotion Recognition. In *Computer Vision and Machine Intelligence Paradigms for SDGs: Select Proceedings of ICRTAC-CVMIP 2021* (pp. 133-150). Singapore: Springer Nature Singapore.
- [16]. Fahd Saleh Alotaibi: "Implementation of Machine Learning Model to Predict Heart Failure Disease. In: *International Journal of Advanced Computer Science and Applications* "(IJACSA)"(2019).