SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I8P111

Volume 11 Issue 8, 98-110, August 2024

Original Article

Smart Sight: Intelligent Vision for Stroke Lesion Detection and Classification

M. Fathima Beevi[1], N. Santhi[2], N. Ramasamy[3]

[1,2]*Department of Electronics and Communication Engineering, Noorul Islam Centre for Higher Education, Tamilnadu, India.*
[3]*Department of Mechanical Engineering, Noorul Islam Centre for Higher Education, Tamilnadu, India.*

[1]*Corresponding Author : fathima.beevi.m@outlook.com*

Received: 04 June 2024 Revised: 18 July 2024 Accepted: 09 August 2024 Published: 31 August 2024

Abstract - *Stroke is the traumatic condition of nerve cells that block down the physical activity of the victims within a short span of time. It is the leading reason for disability and mortality among the older population. A timely diagnosis is a crucial step in case of stroke treatment. Contemporary diagnostic techniques, including Magnetic Resonance Imaging (MRI) and Computer-Aided Tomography (CT), have wide applications in detecting stroke lesions. This paper introduces a novel approach for addressing the critical task of accurately segmenting stroke lesions from medical images, which is vital for precise diagnosis and effective treatment planning. The proposed approach integrates DenseNet-201 and Capsule Network architectures to develop a hybrid deep learning model. DenseNet-201 serves as a feature extractor, facilitating enhanced feature propagation and gradient flow throughout the network. Meanwhile, Capsule Network introduces capsules to handle hierarchical relationships, improving the model's ability to capture intricate spatial hierarchies in the data. The dataset used for training and evaluation consists of brain CT images sourced from the Kaggle Repository, including both normal and stroke brain CT images. Through preprocessing and augmentation techniques, the dataset's quality and diversity are enhanced to ensure effective model training. Experiment results show how effective the suggested hybrid model is, achieving an accuracy of 93.45%, precision of 92.18%, recall of 92.56%, and F1-Score 92.36%. When compared to other approaches currently in use, the suggested method performs better in terms of robustness and segmentation accuracy. Overall, this hybrid deep learning model offers a promising solution to the challenges of stroke lesion detection and classification, with implications for improving patient care and treatment outcomes in clinical settings.*

Keywords - *Stroke lesions, CT imaging, Deep Learning models, Detection, Classification, Capsule network.*

1. Introduction

The nervous system's central processing unit, the human brain, controls a number of body processes. It comprises four main functional regions known as lobes: frontal, temporal, parietal, and occipital, each responsible for specific functions such as memory, sensory perception, emotions, speech, coordination, intelligence, muscular control, and consciousness. Brain stroke, a potentially life-threatening medical condition, occurs due to traumatic brain injury or blockage of blood vessels, leading to a lack of oxygen and subsequent cell death. Strokes disrupt blood flow to affected brain regions, resulting in loss of motor control, memory, and other functions. This condition has profound social, physical, and emotional implications for patients and their family members. Symptoms of stroke may manifest after its onset, including epileptic seizures, which can render individuals bedridden or paralyzed [1]. Transient Ischemic Attack (TIA) stroke is a temporary condition lasting less than one or two hours. Timely intervention is crucial in stroke treatment to prevent lasting tissue damage or death, with medical practitioners facing pressure to act swiftly. Effective stroke prevention programs targeting hypertension, hyperlipidemia, and tobacco use can help reduce mortality rates.

The accumulation of fatty deposits in the arteries providing blood to the brain is the main cause of strokes, as it reduces oxygen and blood flow to the brain. Additionally, blood clots originating elsewhere in the body can travel through the bloodstream, blocking crucial arteries and further denying nutrition and oxygen to the brain [2]. Figure 1 illustrates the brain stroke.Stroke, referred to as a cerebral vascular injury, is one of the main reasons of death worldwide, affecting over 20 million individuals, of whom 5 million will eventually pass away. Based on the degree of tissue damage, stroke cases are divided into two classes: reversible stroke and irreversible stroke. The CT and MRI imaging modalities are used to determine the tissue damage in the brain. In a healthy brain, blood passes through the thin walls of the arteries to supply neurons with oxygen and nutrients. Normal brain function is significantly impacted by a stroke [3].

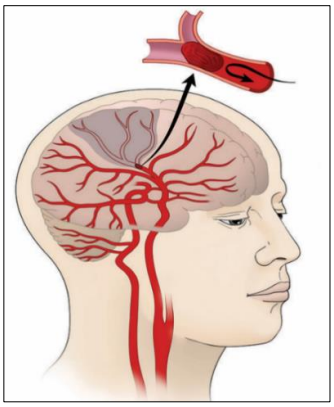

Fig. 1 Brain stroke

Haemorrhagic and ischemic strokes are the two forms that exist, as shown in Figure 2. Ischemic stroke is the most prevalent kind, resulting from brain cell death due to oxygen and nutrient deprivation, primarily caused by artery blockage. It constitutes 87% of all strokes and leads to brain tissue malfunction. Haemorrhagic stroke, less common but more severe, occurs when bleeding into the brain from a ruptured blood vessel disrupts neuronal function. About 12% of strokes are haemorrhagic, with high blood pressure often being the underlying cause. Intracerebral haemorrhage involves blood flow into brain tissue, while subarachnoid haemorrhage involves bleeding into the brain's surrounding space. Both types of strokes pose significant health risks and require prompt medical attention [4].

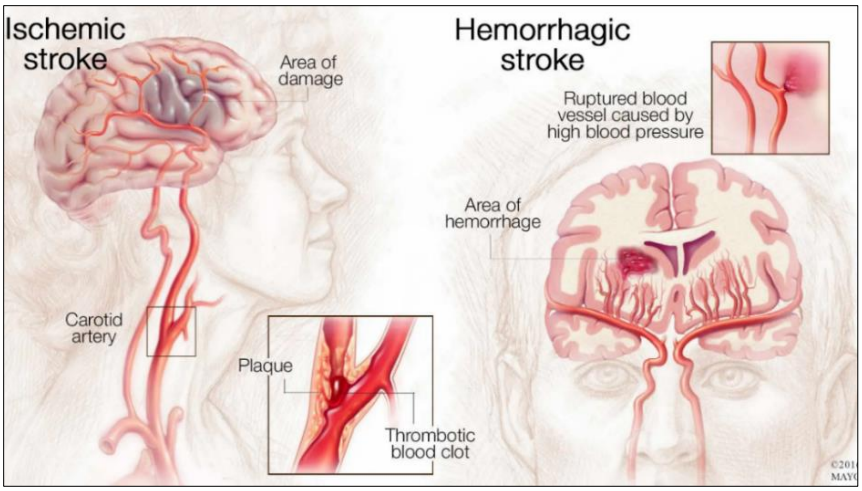

Fig. 2 Ischemic stroke and haemorrhagic stroke

Segmentation plays a crucial role in stroke diagnosis by making it easier to recognize and define the affected areas. MRI and CT are two often used modalities in the field of stroke segmentation. Both CT and MRI offer benefits and drawbacks, and the choice between the two relies on a variety of factors, including the state of the patient, accessibility, and certain clinical standards for stroke segmentation study. Artificial Intelligence (AI) has given medical imaging much attention, particularly with the rapid advancement of Deep Learning (DL) technology [5]. AI has gained prominence in the processing of medical imaging and has become a focus area.

The suggested work's notable contribution includes:

- A DL based intelligent system for the identification and categorization of stroke lesions.
- Improving the system's performance evaluation parameters.
- Evaluating the suggested system's performance in comparison to the current approaches.

The study is organized into several sections to ensure clarity and coherence when presenting the research findings. In Section 2, a thorough literature review is provided, offering insights into existing studies and methodologies related to the topic under investigation.

Section 3 elaborates on the proposed methodology, including experimental settings and procedural details to facilitate reproducibility and understanding. The results obtained through the implementation of the proposed model are discussed and compared with the present approaches in Section 4, enabling a comprehensive analysis of the model's efficiency. Finally, Section 5 serves as the conclusion of the study.

2. Related Works

Clerigues et al. [6] introduced a resilient DL approach for separating stroke lesions into acute and sub-acute categories using multimodal MRI data. Utilizing a UNet-based CNN architecture, the approach tackled class imbalance issues, incorporated a balanced sampling strategy for training patches, and implemented a dynamically weighted loss function. By leveraging pre-processing techniques based on brain hemisphere symmetry, the method enhanced feature extraction.

Evaluation of the 2015 ISLES challenge demonstrated superior performance, ranking first in online subtasks with DSC of 0.59 ± 0.31 and 0.84 ± 0.10 for SISS and SPES, respectively. The method exhibited generalizability to diverse tasks and imaging modalities, emphasizing broader clinical applicability.

Kumar et al. [7] introduced the Classifier-Segmentor Network (CSNet) to automate acute stroke lesion segmentation. When self-similar fractal networks were combined with U-Net architecture, CSNet addressed challenges like diverse lesion appearance, limited datasets, and the need for multiple MRI modalities. The hybrid training strategy enhanced parameter sharing and efficiency. The cascaded architecture refined accuracy through spatial and semantic information integration and a voting mechanism improved overall segmentation accuracy. Evaluation in the MICCAI (ISLES) challenge demonstrated CSNet's superiority in Accuracy, Dice-Coefficient, Recall, and Precision.

Nazari et al. [8] introduced a completely automated system designed for the localization and Acute Ischemic Stroke (AIS) segmentation. The technique compared images of stroke patients with those of healthy controls based on the Crawford-Howell t-test. Post-lesion segmentation, a classifier was developed to categorize the images into groups of non-strokes and strokes. The results demonstrated the system's significant potential to enhance the efficiency and accuracy in lesion segmentation of AIS in daily clinical practice. Notably, its seamless integration into the diagnostic workflow and minimal computational resource requirements further underscored its practicality.

Tomita et al. [9] employed deep residual neural networks for automatic segmentation of post-stroke lesions from T1-weighted MRI scans. Targeting accurate volumetric segmentation in chronic stroke patients, the analysis used 3D residual learning-based deep convolutional segmentation models. Analyzing 239 T1-weighted MRI scans, the model was evaluated against manual lesion tracing, achieving notable metrics such as DSC of 0.64 and median of 0.78, along with 20.4mm of Harsdorf Distance (HD) and 3.6 mm of average symmetric surface distance. This approach demonstrated segmentation accuracy and robustness, contributing to advancements in segmenting chronic ischemic stroke lesions in 3D.

Cui et al. [10] introduced DeepSym-3D-CNN, a symmetric 3D CNN for automated diagnosis of AIS utilizing images of DWI. Using images from 190 subjects, separating the right and left hemispheres from 3D DWI brain scans, the model was achieved, employing L2 normalization and feature subtraction. Comparative models were also constructed. The broad clinical applicability is suggested by the method's automatic identification of AIS from images of DWI and its capacity for extension to illnesses showing asymmetric lesions.

Wei et al. [11] employed DL models for AIS lesion segmentation, classification, and mapping using MRI images. They introduced a Semantic Segmentation Guided Detector Network, consisting of dual models for DWI segmentation and binary classification of lesion size and circulatory territory. The updated SGD-Net Plus recognized and registered AIS lesions in T1-weighted images, brain atlases, and DWI images automatically. The study emphasized the value of subject knowledge-oriented design in applications of AI, improving the usage of MRI in patient care and improving learning of patients' illnesses.

To predict the ultimate size and location of infarcts in acute stroke patients, Nazari et al. [12] developed a DCNN that was only trained with DWI. This approach eliminated the requirement for perfusion-weighted imaging. The Attention-Gated (AG) DCNN managed to compute volumes of infarcts 3–7 days following a stroke with a degree of accuracy that was comparable to the frameworks considering both PWI and DWI by utilizing DWI and apparent diffusion coefficient maps. This implied that a DWI-trained DCNN could reduce complexity and cost in acute stroke diagnostic procedures, eliminate the requirement for PWI, and streamline treatment decisions with shorter stroke imaging protocols.

Praveen et al. [13] proposed an autonomous advanced learning method for segmenting ischemic stroke lesions with a framework of layered sparse autoencoders. Unlike current techniques relying on manually created features, this approach utilized unsupervised learning. Sparse Auto Encoder (SAE) layers built a deep architecture, and a Support Vector Machine (SVM) classified patches as normal or lesions. Tested on the ISLES 2015 dataset, the method achieved high precision (0.908), average dice's coefficient (0.903), recall (0.904), and accuracy (0.904). Compared to advanced techniques, the proposed strategy performed significantly better in precision, dice coefficient, and recall, with improvements of 25.71%, 36.67%, and 16.96%, respectively. The SSAE framework's ability to learn unsupervised features allowed for effective training on large datasets, surpassing hand-crafted features.

Zhang et al. [14] proposed an automatic technique for extracting AIS from DWIs using deep 3D CNNs. Their approach utilized 3D contextual data, allowing for efficient detection of discernible characteristics. To address the class imbalance, they employed a Dice objective function during model training. Tested on a dataset of 242 participants, their model outperformed other CNN techniques significantly, achieving high precision (90.67%), dice similarity coefficient (79.13%), and F1 lesion-wise score (89.25%). Generalization tests on the ISLES2015-SSIS dataset also yielded extremely competitive results, showcasing the method's potential for quick and accurate application in clinical practices.

Zhang et al. [15] presented a framework for AIS lesion segmentation in DW MRI. Their Detection and Segmentation Network (DSN) used a triple-branch design to extract separate plane features in order to address data imbalances. With a significant 62.2% dice coefficient and 71.7% sensitivity, the Multi-Plane Fusion Network (MPFN) improved segmentation accuracy and showed supremacy on the ISLES2015 SSIS DWI sequence dataset.

Clerigues et al. [16] addressed challenges in detecting ischemic stroke in acute CT images by introducing an automated DL solution for segmenting lesions. Their approach included improvements like symmetric modality enhancement and filtering of uncertainty, as well as a more regularized network training strategy. It was tested on the ISLES 2018 dataset. In blind testing, the suggested approach outperformed other cutting-edge techniques, achieving a Dice similarity coefficient of 49%. This tool had clinical potential for estimating lesion core size and location, eliminating the need for time-consuming magnetic resonance imaging.

Soltanpour et al. [17] introduced a prediction algorithm for the efficient identification of Computed Tomography Perfusion (CTP) scans exhibiting ischemic stroke lesions. The algorithm utilized a sophisticated model with four parallel 2D U-Nets, each dedicated to extracting information from different CTP maps. This multichannel approach captured diverse lesion characteristics more accurately than simple predefined features. The algorithm combined probability maps from parallel U-Nets, utilizing pixel-level and neighbourhood information. Analysis using the ISLES 2018 dataset revealed improved results with a DSC of 71.3%, Recall of 73.6%, and Volume Similarity (VS) of 82.1%.

Shi and Liu [18] enhanced ischemic stroke lesion segmentation by modifying the U-Net architecture, addressing the challenges of training data scarcity and overfitting. Using the ISLES 2018 dataset, their modified U-Net incorporated shortcut connections as residual blocks and employed element wise-sum and concatenation techniques to mitigate overfitting. Evaluation based on the dice coefficient and Jaccard index showed a significant improvement over the standard U-Net, achieving a dice coefficient of approximately 0.77 for ischemic stroke segmentation.

Omarov et al. [19] addressed challenges in ischemic stroke segmentation using machine learning, proposing a modified 3D UNet architecture for improved performance on 3D CT images. Leveraging the 2018 ISLES dataset, their model achieved a Dice/F1 score similarity coefficient of 58%, surpassing the standard 3D UNet. Efficient averaging, data augmentation, and regularization techniques prevented overfitting in the limited dataset. The application of the loss function for intersection over union added value to zone recognition shapes. Object extraction was emphasized for increased segmentation accuracy.

Brain stroke detection and classification are critical due to the time-sensitive nature of treatment, where delays can severely impact efficacy. Diagnostic delays stem from various factors, including the skill of medical personnel, accessibility and quality of imaging technologies, and the variability in symptoms among patients. Stroke symptoms can vary widely depending on location and severity, complicating accurate diagnosis and classification. Limited accessibility to imaging tools like CT and MRI scans, particularly in resource-constrained settings, further exacerbates diagnostic challenges. Additionally, these tests lack sensitivity to detect milder strokes, leading to misdiagnosis or delayed treatment. This is especially problematic for younger individual's typical stroke symptoms. Furthermore, selecting the appropriate kernel function is crucial for SVM performance, as choosing incorrectly can lead to overfitting or subpar outcomes. Space complexity also poses challenges, particularly in distinguishing between mine stroke and actual stroke.

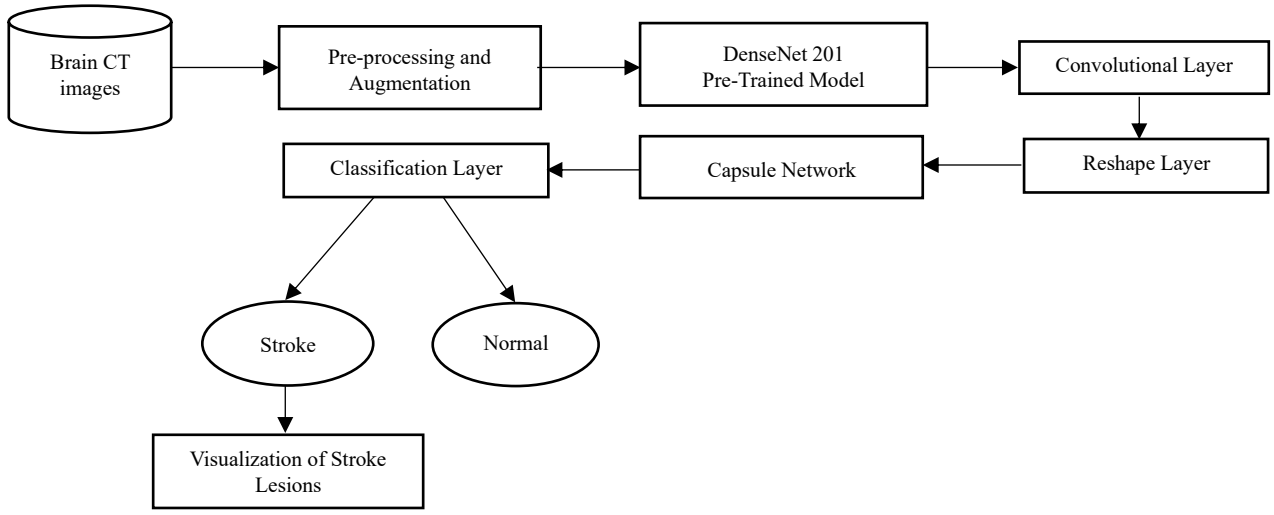

Fig. 3 Schematic block diagram of the proposed methodology

3. Materials and Methods

The detection and classification of stroke lesions is crucial. So, in this research, a deep learning based intelligent framework is developed for stroke lesion detection. Figure 3 shows the block diagram representation of the proposed methodology.

3.1. Brain CT Image Dataset

The brain CT images have been collected from the Kaggle, https://www.kaggle.com/code/rounakislamraisa/brain-stroke-prediction/input. The dataset consists of both normal and stroke brain CT images, as shown in Figure 4.

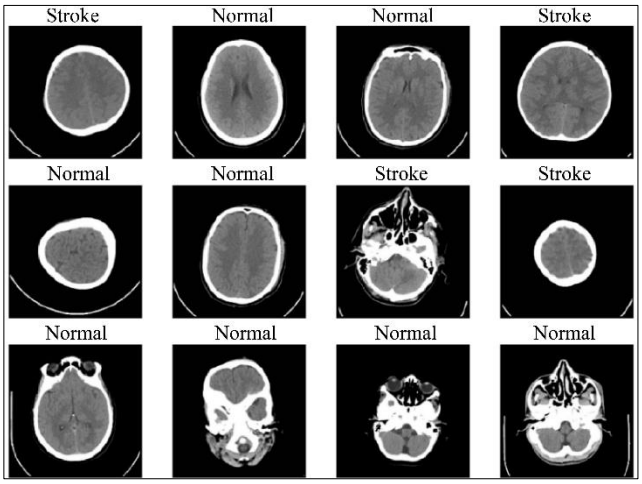

Fig. 4 Sample images in the dataset

3.2. Data Pre-processing and Augmentation

In the proposed study, pre-processing involves several steps aimed at standardizing and optimizing the input data to ensure effective model training and performance. Initially, pre-processing likely includes image normalization to correct for variations in brightness, contrast, and resolution among the brain CT images collected from different sources. This normalization process ensures consistency in image characteristics, thereby facilitating more accurate feature extraction during model training. Additionally, techniques like image resizing are performed to achieve uniform dimensions across the dataset, which streamlines computational processes and improves model efficiency.

Furthermore, data augmentation techniques are employed to augment the dataset's diversity and robustness, thereby enhancing the model's ability to generalize to unseen data and variations in input images. Data augmentation encompasses a variety of transformations applied to the original images, including rotation, translation, scaling, flipping, and adding noise. These transformations generate new synthetic samples from existing data, effectively increasing the dataset size and introducing variability in image appearance. For brain stroke detection, specific augmentation strategies are tailored to mimic potential variations in stroke lesions, such as altering lesion size, shape, and location, as well as simulating different imaging conditions and artefacts commonly encountered in clinical settings. By augmenting the dataset with diverse samples representing various stroke manifestations, the model becomes more robust and adaptable, ultimately improving its performance in accurately detecting stroke lesions in unseen medical images. Together, these pre-processing and augmentation techniques contribute to the development of a more effective and reliable deep learning model for accurately detecting and classifying stroke lesions in brain CT images, thus advancing diagnostic capabilities.

3.3. Proposed Model Architecture

3.3.1. DenseNet 201

DenseNet-201 is an extension of the DenseNet architecture, which stands for Densely Connected Convolutional Networks, which is a deep learning architecture. It differs from traditional Convolutional Neural Networks (CNNs) by establishing dense connections between layers, enabling enhanced feature propagation and gradient flow throughout the network. DenseNet is particularly efficient in terms of parameter usage and alleviates the vanishing gradient problem. Specifically, DenseNet-201 comprises 201 layers, which include activation, batch normalization, pooling, and convolutional layers, as shown in Figure 5.

The input layer receives the image data input, which is usually represented by pixel values. Multiple dense blocks, each with numerous convolutional layers, constitute DenseNet-201. Every layer in a dense block gets input from every layer that came before it, and every layer that comes after it gets its feature maps as input. By promoting feature reuse and facilitating gradient flow, this dense connectivity structure enhances information flow across the network. Transition layers use pooling and convolutional processes to lower the dimensionality of feature maps. A global average pooling layer is deployed at the network's end to minimize the feature maps' spatial dimensions to a vector of features [20].

The input feature maps of layer l are denoted byx_l. In a DenseNet, the output of each layer is concatenated with the feature maps of all subsequent layers, as given in Equation (1).

$$x\_l=H\_l\ ([x\_0,x\_1,....x\_(l-1)\ ]) \quad (1)$$

Where H_l represents the non-linear transformation applied by the l-th layer, and [.] denotes concatenation. By enabling direct access to all previous layers' feature maps, this concatenation technique encourages feature reuse and speeds up the learning process for each layer. The DenseNet architecture is typically organized into dense blocks and transition layers. The architecture comprises dense blocks, transition layers, global average pooling, and a fully connected layer for classification. The layered architecture of DenseNet 201 is shown in Figure 6.

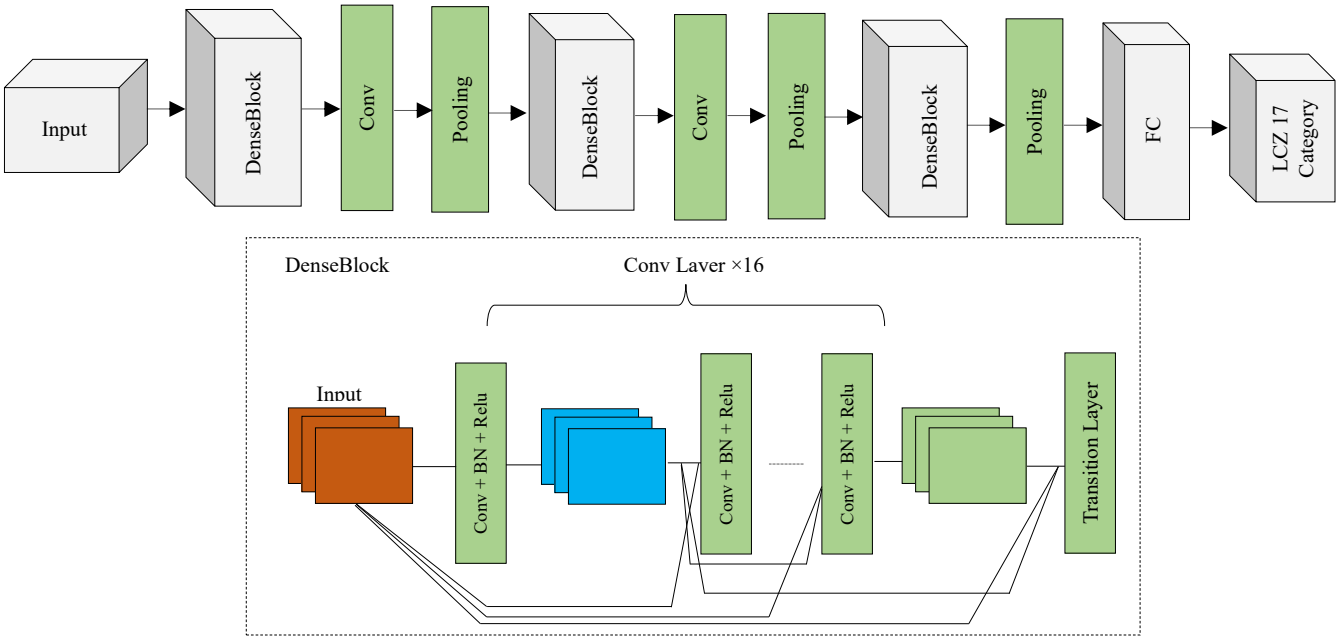

Fig. 5 Basic DenseNet architecture

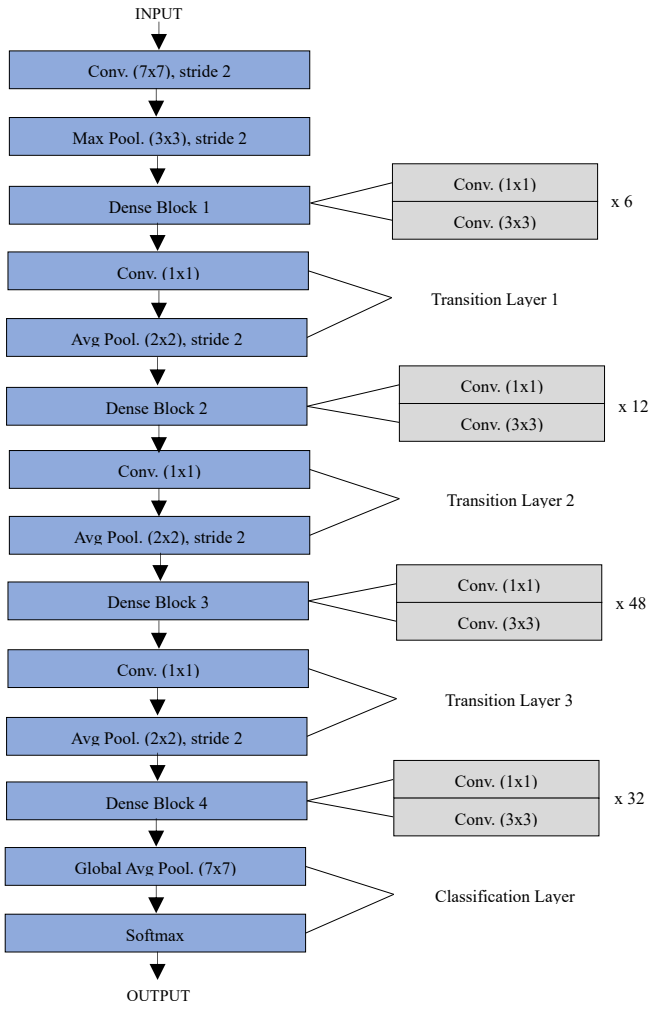

Fig. 6 Layered DenseNet 201 architecture

A dense block consists of multiple convolutional layers, each with a fixed number of output feature maps. Within a dense block, the output of each layer is concatenated with the feature maps of all preceding layers. The number of convolutional layers in a dense block is L, and the growth rate, which determines the number of feature maps produced by each convolutional layer within the dense block, as k. The output of the l^{th}layer in a dense block is calculated as Equation (2).

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]) = H_l([x_{l-1}, F_l(x_{l-1})]) \quad (2)$$

Where, F_l represents the composite function implemented by the l^{th} convolutional layer, and $F_l(x_{l-1})$ produces k feature maps. Thus, the output dimension of each layer within the dense block is $k \times l$, where l represents the layer index within the dense block. Transition layers, composed of batch normalization, 1×1 convolutional, and pooling layers, reduce spatial dimensions and regulate feature map numbers between dense blocks in DenseNet.

3.3.2. Capsule Network

Inspired by the organizing principles of organic neural structures, a capsule neural network, or CapsNet, is a type of Artificial Neural Network (ANN) used in machine learning that mimics hierarchical relationships. CapsNets aim to imitate the hierarchical structure of biological brain circuits.

CapsNets introduce capsules as fundamental components known as digit capsules, each representing a different class to address the limitations of conventional neural networks. Unlike standard neurons, capsules are adept at handling hierarchical structures and pose variations by encapsulating activation information and spatial relationships.

Each capsule outputs pose parameters alongside activation, representing specific entities or object parts. Through dynamic routing, the architecture of the capsule network consists of various elements, such as the encoder and decoder network. The encoder network of the capsule is shown in Figure 7.

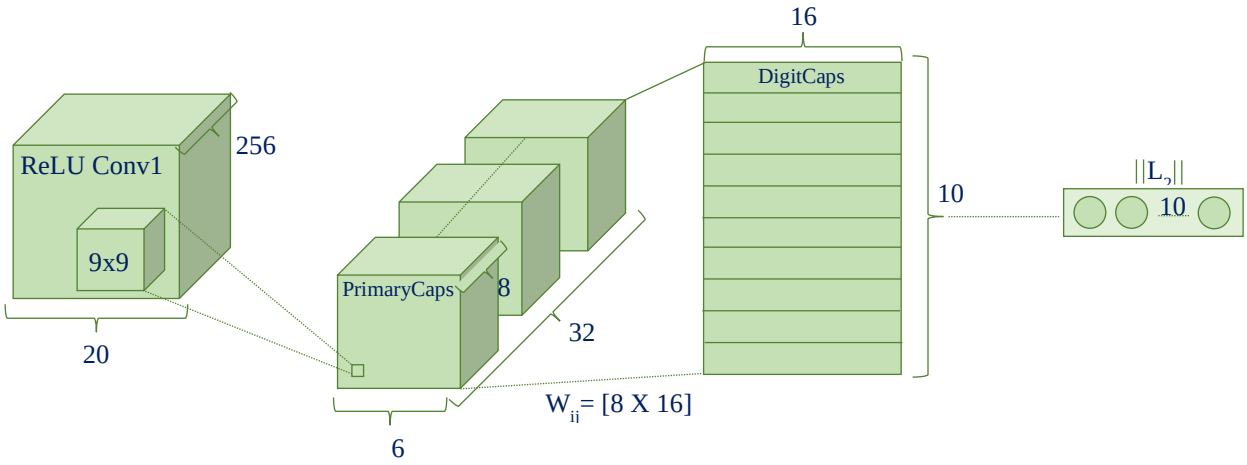

Fig. 7 Encoder network of CapsNet

CapsNets iteratively refine coupling coefficients based on pose parameter agreement, enabling better recognition of intricate data patterns and capturing complex spatial hierarchies, ultimately enhancing generalization capabilities. Image inputs are converted into vectors by the CapsNet architecture, which contains parameterization parameters that are essential for producing the image. An encoder is employed, which consists of a convolutional layer and further layers such as PrimaryCaps and DigitCaps.

The convolutional layer extracts low-level features from the input images to start the process. In order to capture important patterns, the PrimaryCaps layer uses clustering of neurons called capsules. Each capsule reflects an instantiation characteristic, such as posture. The DigitCaps layer is the foundation of CapsNets. A capsule represents a particular entity type, such as a class of digits, and encodes the chance that the entity exists, as well as its instantiation parameters. In the end, a reconstruction loss is used to encapsulate the instantiation parameters [21]. This involves computing the loss of each training example for each output class by Equation (3).

$$L_k = T_k max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k)max(0, \|v_k\| - m^-)^2 \quad (3)$$

Where margin loss for the k^{th} digit capsule is denoted by L_k, indicator T_k is binary, the k^{th}digit capsule's activity vector is denoted as v_k .The activity vector's length is denoted by $\|v_k\|$. The positive and negative margin represents as m^+ and m^-. A down-weighting factor for the loss resulting from erroneous digit capsules is represented by λ. Capsule networks utilize dynamic routing by agreement to update the coupling coefficients between lower-level and higher-level capsules. The routing mechanism aims to increase the agreement between predictions of lower-level capsules and the input

vectors of higher-level capsules. The coupling coefficients are iteratively updated by Equation (4).

$$c_{ij} = \frac{exp(b_{ij})}{\sum_k exp(b_{ik})} \quad (4)$$

Where b_{ij} represents the log prior probabilities of the coupling coefficients. In Capsule Networks, the Decoder Network is crucial for reconstructing input images from the data stored in the DigitCapsules. Utilizing the instantiation properties of selected DigitCapsules, such as pose and viewpoint, the decoder rebuilds the input data following the dynamic routing process. Fully-connected layers within the decoder facilitate this reconstruction, converting instantiation parameters back to the original input space to ensure accurate image recovery. Typically, reconstruction loss is computed using Euclidean distance between the reconstructed image and the original input to promote faithful image reconstruction. This reconstruction process not only enhances classification accuracy but also contributes to meaningful image reconstruction, aligning with the overall training objective of Capsule Networks. The decoder network is shown in Figure 8.

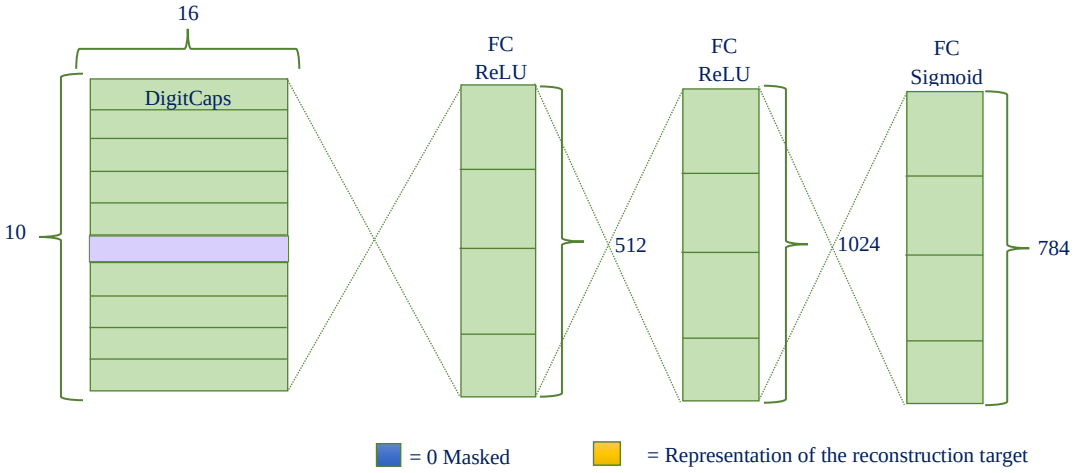

Fig. 8 Decoder network of CapsNet

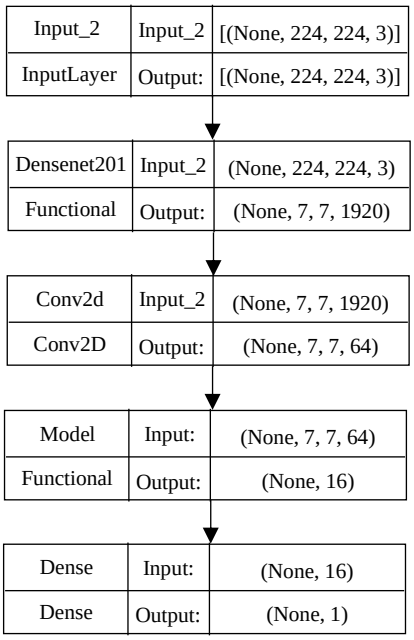

Fig. 9 Proposed network architecture

3.3.3. Proposed Hybrid Model

The suggested hybrid deep learning model combines a capsule network with DenseNet 201. The input layer takes 224 x 224 image inputs. After that, a DenseNet 201 architecture, which functions as a feature extractor, is fed to these inputs. After passing through a convolutional layer with 64 1x1

filters, the DenseNet model's output is then subjected to a ReLU activation function.

When input to the Capsule Network, this extra convolutional layer aids in additional feature extraction and refining. The prior layer output has spatial dimensions of 7×7 with 64 feature maps, as indicated by the CapsNet architecture's definition of an input form of 7×7 feature maps with 64 channels. The 32 capsules in the CapsNet are set up with an 8-dimensionality each.

The objective of CapsNet is to ascertain the spatial arrangements and hierarchical relationships among the items in the input images. The CapsNet output is then transmitted via a dense layer, which consists of a single neuron and a sigmoid activation function. This layer maps the outputs of the capsules to a binary classification decision, generating a probability that indicates the chance that the input image belongs to a particular class. The proposed model architecture is given in Figure 9.

3.4. Hardware and Software Setup

Thecomputational setup for this research utilized a machine with robust specifications featuring an Intel Core i7 processor. 32GB of RAM, and the formidable NVIDIA GeForce GTX 1080Ti GPU. Model implementation was seamlessly carried out through the Keras library, functioning as a prototype built upon the Tensorflow framework and executed using the versatile Python language. Keras, known for its user-friendly interface and powerful capabilities, proved instrumental in crafting intricate Neural Network architectures.

This framework ensures efficient utilization of computing resources, seamlessly accommodating CPU, GPU, and TPU environments. To leverage extensive computational capabilities and streamline model training, the deployment was orchestrated on Google Colab. This cloud-based Python notebook environment not only provides complimentary access to robust computational resources but also facilitates collaborative development, making it an optimal choice for training models.

Hyperparameters are essential configuration settings that define the behaviour and characteristics of a machine learning framework throughout the training process. Unlike the parameters of the model, which are learned from the data itself, hyperparameters are set by the user before training begins. The neural network model uses the Adam optimizer. The binary cross-entropy loss function guides the training process. During training, the model processes input data in batches of 32 samples per iteration.

The training is carried out over 20 epochs, signifying the number of times the model processes the entire training dataset. These hyperparameter choices, such as the optimizer, loss function, batch size, and number of epochs, collectively define the configuration for training the neural network model, aiming to optimize its performance on the proposed stroke detection. The model configuration of the suggested approach is tabulated in Table 1.

Table 1. Model configurations

Hyperparameter	Values
Optimizer	Adam
Loss Function	Binary Cross Entropy
No. of Epochs	20
Batch Size	32
Activation Function	ReLu, Sigmoid

4. Results and Discussion

The accuracy and loss plots play a vital role in understanding the performance and learning patterns of the proposed model. The accuracy plot provides a visual representation of how accurately the model predicts the labels of the data during training iterations on both the training and validation datasets.

It tracks the alignment between the model's predictions and the actual labels, serving as a key indicator of the model's performance throughout the training process.

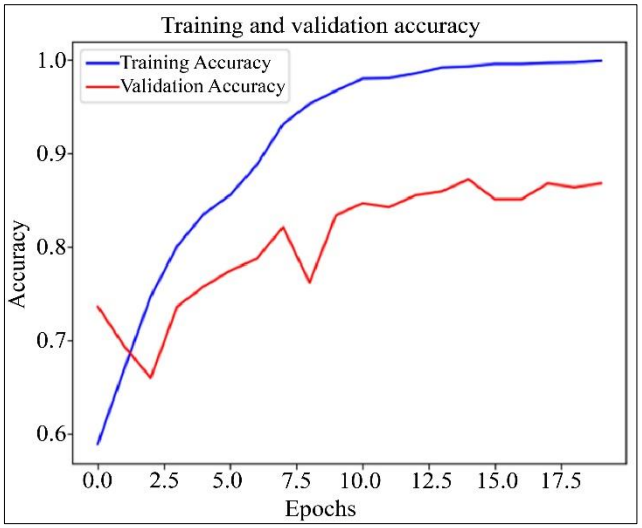

Fig. 10 Accuracy plot

The accuracy plot illustrates the model's effectiveness in distinguishing between CT images with and without signs of stroke over the course of training. Ideally, in the initial epochs, both training and validation accuracies increase concurrently, indicating the model's capability to generalize its knowledge beyond the training dataset.

This trend, depicted in Figure 10, suggests that the model is learning underlying patterns rather than simply memorizing the examples provided in the training set.

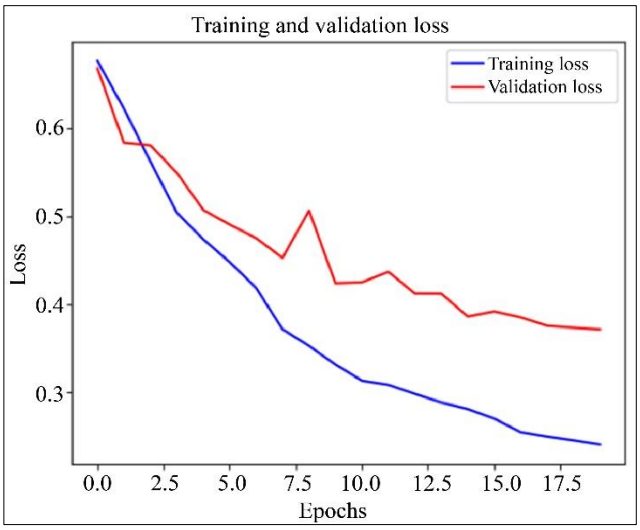

Fig. 11 Loss plot

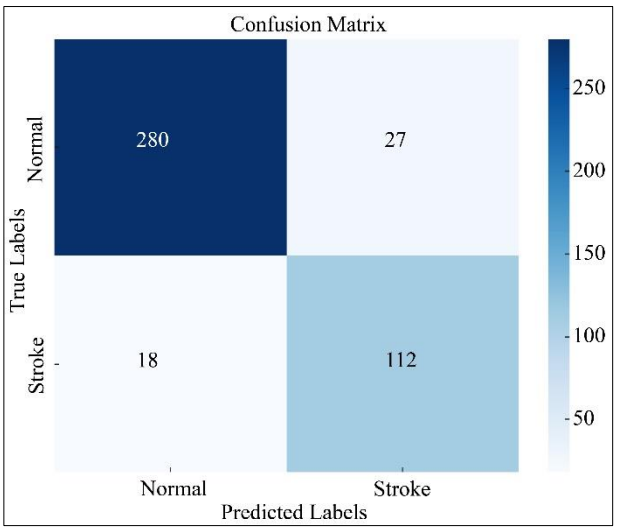

Fig. 12 Confusion matrix

The disparity between the predicted results and the true labels of the dataset is quantified numerically as the model's loss, which is visualized in the loss plot. Throughout the training process, the goal is for the loss to diminish progressively, indicating that the model is refining its predictions and minimizing errors over successive iterations, as illustrated in Figure 11.

A valuable method for evaluating the effectiveness of the proposed model in detecting stroke from CT images is through the utilization of a confusion matrix. This matrix provides a structured overview of the model's performance by comparing its predictions with the actual labels across different classes.

Essentially, it organizes the outcomes into a table format, where the rows represent the true labels and the columns represent the predicted labels, as shown in Figure 12. Each cell within the matrix contains the count of instances where the model's predictions align with the true labels or diverge from them. The confusion matrix is divided into four quadrants, with the diagonal components indicating accurate predictions and the off-diagonal elements indicating instances of misclassification. Through this visual representation, the effectiveness of the proposed model in accurately identifying stroke cases can be thoroughly assessed.

Performance metrics derived from the confusion matrix offer a thorough evaluation of the proposed model's efficacy in detecting stroke. In order to thoroughly evaluate the efficacy and operational efficiency of the proposed model, the F1-score, accuracy, precision, and recall are the four primary metrics utilized.

These measures, which are based on the concepts of False Positive (FP), False Negative (FN), True Negative (TN), and True Positive (TP), are essential for assessing the model's performance. These performance parameters have mathematical formulations that are shown in Equations (5) - (8).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1-score = 2 \times \frac{precision \times Recall}{Precision+Recall} \quad (8)$$

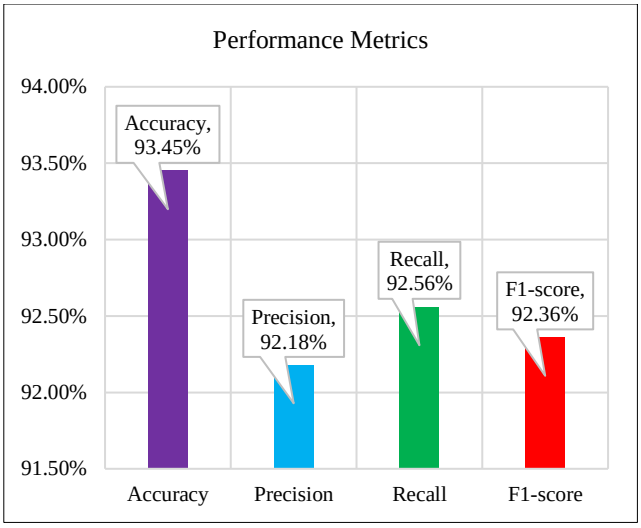

Fig. 13 Performance metrics

Figure 13 highlights the model's impressive performance, with an overall accuracy of 93.45%, indicating its ability to classify the majority of cases within the dataset effectively. Moreover, the precision and recall scores of 92.18% and 92.56%, respectively, underscore the model's proficiency in identifying positive cases while minimizing false positives.

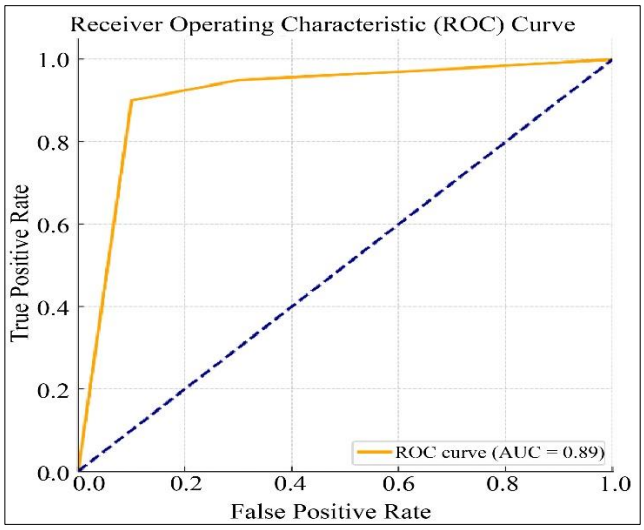

Fig. 14 ROC curve of the proposed method

These metrics highlight the model's reliability in accurately distinguishing between stroke and non-stroke cases. Furthermore, the F1-score, which combines recall and precision, reinforces the model's robustness with a value of 92.36%. Overall, these results affirm the efficacy of the model in accurately detecting and categorizing stroke lesions in brain CT images, showcasing its potential for enhancing diagnostic accuracy and patient care in clinical settings.

The Receiver Operating Characteristic (ROC) curve is a widely used graphical tool for evaluating the performance of classification models in binary classification tasks. It illustrates the trade-off between the true positive rate and the false positive rate across different threshold values. By plotting these rates against each other, the ROC curve provides insight into the classifier's ability to distinguish between positive and negative instances.

A key metric derived from the ROC curve is the Area Under the Curve (AUC-ROC), which quantifies the overall performance of the classifier. The AUC-ROC value ranges from 0 to 1, with higher values indicating superior discrimination ability. Figure 14 visualizes the ROC curve, offering a clear representation of the classifier's performance across various threshold levels.

Grad-CAM, short for Gradient-weighted Class Activation Mapping, stands as a pivotal visualization technique utilized to unravel the decision-making mechanisms of the proposed model in the realm of stroke detection. This technique adeptly illuminates the key regions within an input image that wields the most influence over the network's predictive outcome. By generating a heat map, Grad-CAM accentuates the spatial zones within the input image that bear the utmost relevance to the predicted stroke classification.

In essence, Grad-CAM serves as a potent tool offering profound insights into the intricate decision-making processes of the model, thereby facilitating model interpretation and aiding in the debugging process. Figure 15, presented below, showcases a vivid illustration of the Grad-CAM Heat map Visualization, illuminating the areas of utmost significance within the input image for stroke detection.

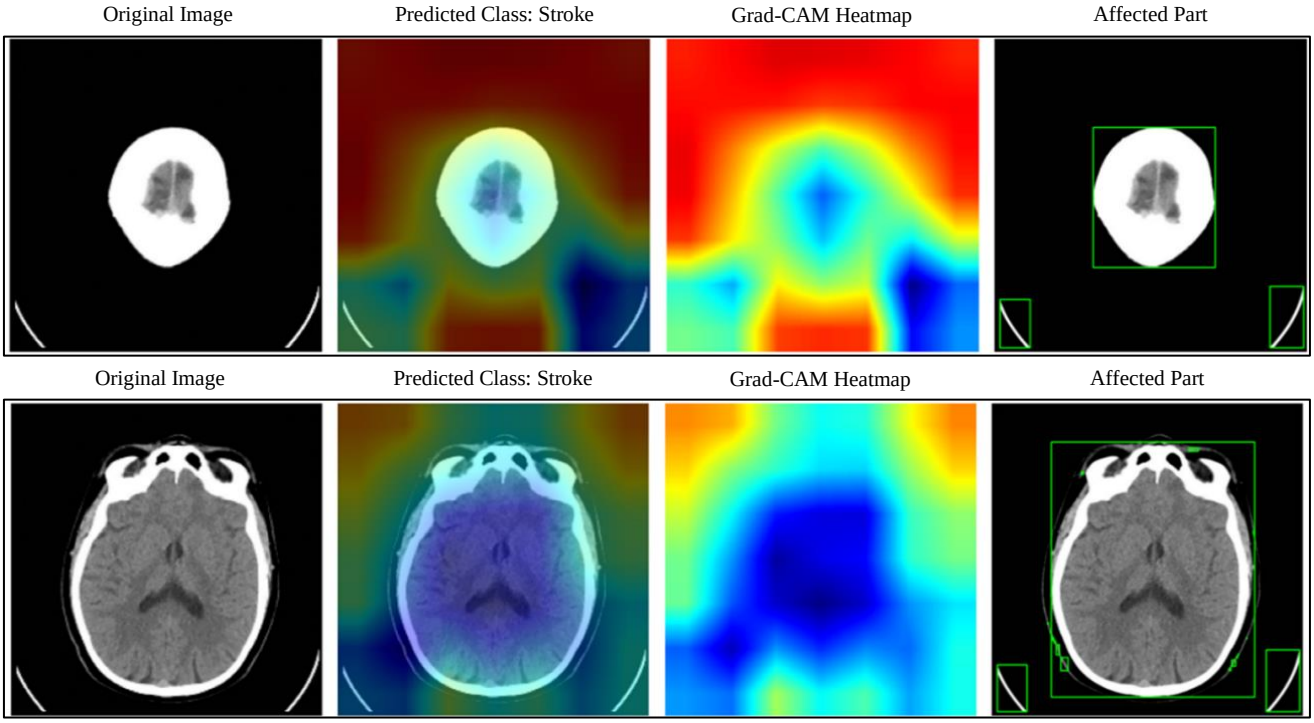

Fig. 15 Visualization using Grad- CAM approach

Table 2. Comparison of the proposed hybrid method with existing methods

Author	Methodology	Dataset	Results
Clèrigues et al. [6]	UNet-based CNN architecture,	2015 ISLES challenge	DSC: 0.59 ± 0.31 for SISS, 0.84 ± 0.10 for SPES
Tomita et al. [9]	Deep residual neural networks	239 T1-weighted MRI scans	DSC: 0.64, Median: 0.78, HD: 20.4mm, Average symmetric surface distance: 3.6mm
Praveen et al. [13]	Layered sparse autoencoders,	ISLES 2015 dataset	precision (0.968), average dice coefficient (0.948), recall (0.924), accuracy (0.904)
Zhang et al. [14]	Deep 3D CNNs	Dataset of 242 participants	Precision (92.67%), dice similarity coefficient (79.13%), F1 lesion-wise score (89.25%)
Soltanpour et al. [17]	Prediction algorithm for identifying CTP scans with ischemic stroke lesions, four parallel 2D U-Nets	ISLES 2018 dataset	DSC: 71.3%, Recall: 73.6%, Volume Similarity (VS): 82.1%
Shi and Liu [18]	Modified U-Net architecture	ISLES 2018 dataset	Dice coefficient: ~0.77 for ischemic stroke segmentation
Omarov et al. [19]	Modified 3D UNet architecture	ISLES 2018 dataset	Dice/F1 score similarity coefficient: 58%
Proposed method	**Hybrid deep learning model: DenseNet and Capsule Network**	**Brain CT images**	**Accuracy 93.45%, precision 92.18%, recall 92.56%, and F1-Score 92.36%.**

The performance evaluation of the proposed hybrid network, in contrast to existing methods predominantly reliant on machine learning and deep learning, constitutes a crucial aspect of the study. Table 2 presents a comparative analysis, showcasing the efficacy of the hybrid model through a meticulous assessment of outcomes garnered from established approaches. This evaluation scrutinizes various metrics and parameters to gauge the effectiveness and robustness of the proposed method against conventional methodologies utilized in stroke detection.

5. Conclusion

The proposed study introduced a novel hybrid deep learning approach by combining DenseNet-201 and Capsule Network (CapsNet) to enhance stroke lesion detection and classification. The primary goal is to elevate key performance metrics such as F1-score, recall, accuracy, and precision, which are crucial for accurate stroke diagnosis. Utilizing a dataset sourced from the Kaggle repository containing brain CT images of both normal and stroke patients, extensive data pre-processing and augmentation were conducted to ensure dataset quality and diversity, thus facilitating robust model training. The hybrid model architecture capitalizes on the strengths of both DenseNet-201 and CapsNet: DenseNet-201 facilitates effective gradient flow and feature propagation, while CapsNet excels in capturing intricate spatial hierarchies within the data. Experimental results demonstrate the efficacy of the proposed hybrid model, achieving impressive performance metrics, including an accuracy of 93.45%, precision of 92.18%, recall of 92.56%, and F1-Score of 92.36%. Comprehensive visualization techniques such as ROC curves, Grad-CAM heat maps, confusion matrices, and accuracy-loss plots provide insights into the model's robustness and interpretability. Comparative analysis against existing approaches substantiates the superiority of the hybrid model, positioning it as a promising tool for enhancing stroke diagnosis and treatment outcomes. In conclusion, the proposed framework presents a significant advancement in stroke detection accuracy, thereby contributing to the progression of medical imaging and AI applications in healthcare.

Acknowledgments

The author expresses profound appreciation to the supervisor for providing guidance and unwavering support throughout this study.

References

[1] Maarten G. Lansberg et al., "MRI Profile and Response to Endovascular Reperfusion After Stroke (DEFUSE 2): A Prospective Cohort Study," *The Lancet Neurology,* vol. 11, no. 10, pp. 860-867, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[2] Senthil Kumar Thiyagarajan, and Kalpana Murugan "A Systematic Review on Techniques Adapted for Segmentation and Classification of Ischemic Stroke Lesions from Brain MR Images," *Wireless Personal Communications,* vol. 118, pp. 1225-1244, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] A. Gregory Sorensen et al., "Hyperacute Stroke: Simultaneous Measurement of Relative Cerebral Blood Volume, Relative Cerebral Blood Flow, and Mean Tissue Transit Time," *Radiology,* vol. 210, no. 2, pp. 519-527, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[4] Max Wintermark et al., "Acute Stroke Imaging Research Roadmap II," *Stroke,* vol. 44, no. 9, pp. 2628-2639, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[5] M. Koenig et al., "Perfusion CT of the Brain: Diagnostic Approach for Early Detection of Ischemic Stroke," *Radiology,* vol. 209, no. 1, pp. 85-93, 1998. [CrossRef] [Google Scholar] [Publisher Link]

[6] Albert Clèrigues et al., "Acute and Sub-Acute Stroke Lesion Segmentation from Multimodal MRI," *Computer Methods and Programs in Biomedicine,* vol. 194, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Amish Kumar et al., "CSNet: A New DeepNet Framework for Ischemic Stroke Lesion Segmentation," *Computer Methods and Programs in Biomedicine,* vol. 193, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Sanaz Nazari-Farsani et al., "Automated Segmentation of Acute Stroke Lesions Using a Data-Driven Anomaly Detection on Diffusion Weighted MRI," *Journal of Neuroscience Methods,* vol. 333, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Naofumi Tomita et al., "Automatic Post-Stroke Lesion Segmentation on MR Images Using 3D Residual Convolutional Neural Network," *NeuroImage: Clinical,* vol. 27, pp. 1-7, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Liyuana Cui et al., "Deep Symmetric Threedimensional Convolutional Neural Networks for Identifying Acute Ischemic Stroke Via Diffusion-Weighted Images," *Journal of X-Ray Science and Technology,* vol. 29, no. 4, pp. 551-566, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Yi-Chia Wei et al., "Semantic Segmentation Guided Detector for Segmentation, Classification, and Lesion Mapping of Acute Ischemic Stroke in MRI Images," *NeuroImage: Clinical,* vol. 35, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Sanaz Nazari-Farsani et al., "Predicting Final Ischemic Stroke Lesions from Initial Diffusion-Weighted Images Using a Deep Neural Network," *NeuroImage: Clinical,* vol. 37, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] G.B. Praveen et al., "Ischemic Stroke Lesion Segmentation Using Stacked Sparse Autoencoder," *Computers in Biology and Medicine,* vol. 99, pp. 38-52, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[14] Rongzhao Zhang et al., "Automatic Segmentation of Acute Ischemic Stroke from DWI Using 3-D Fully Convolutional DenseNets," *IEEE Transactions on Medical Imaging,* vol. 37, no. 9, pp. 2149-2160, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] Long Zhang et al., "Ischemic Stroke Lesion Segmentation Using Multi-Plane Information Fusion," *IEEE Access,* vol. 8, pp. 45715-45725, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Albert Clèrigues et al., "Acute Ischemic Stroke Lesion Core Segmentation in CT Perfusion Images Using Fully Convolutional Neural Networks," *Computers in Biology and Medicine,* vol. 115, pp. 1-7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Mohsen Soltanpour et al., "Ischemic Stroke Lesion Prediction in CT Perfusion Scans Using Multiple Parallel U-Nets Following by a Pixel-Level Classifier," *IEEE 19th International Conference on Bioinformatics and Bioengineering,* Athens, Greece, pp. 957-963, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Wenyue Shi, and Heng Liu, "Modified U-Net Architecture for Ischemic Stroke Lesion Segmentation and Detection," *IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference,* Chengdu, China, pp. 1068-1071, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Batyrkhan Omarov et al., "Modified UNet Model for Brain Stroke Lesion Segmentation on Computed Tomography Images," *Computers, Materials & Continua,* vol. 71, no. 3, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Aayush Jaiswal et al., "Classification of the COVID-19 Infected Patients Using DenseNet201 Based Deep Transfer Learning," *Journal of Biomolecular Structure and Dynamics,* vol. 39, no. 15, pp. 5682-5689, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Guoyan Li et al., "Compound Fault Diagnosis of Planetary Gearbox Based on Improved LTSS-BoW Model and Capsule Network," *Sensors,* vol. 24, no. 3, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]