A Hybrid Model for Ischemic Stroke Brain Segmentation from MRI Images using CBAM and ResNet50-UNet

Fathia ABOUDI[1], Cyrine DRISSI[2], Tarek KRAIEM[3]
Laboratory of Biophysics and Medical Technologies, Higher Institute of Medical Technologies of Tunis,
University of Tuins El Manar, Tunisia[1]
Department of Neuroradiology, National Institute of Neurology, Mongi Ben Hmida, Tunisia[2]
Laboratory of Biophysics and Medical Technologies, Faculty of Medicine of Tunis, Tunisia[3]

Abstract*—Ischemic stroke is the most prevalent type of stroke and a leading cause of mortality and long-term impairment globally. Timely identification, precise localization, and early detection of ischemic stroke lesions brain are critical in healthcare. Various modalities are employed for detection, and magnetic resonance imaging stands out as the most effective. Different magnetic resonance imaging techniques have been proposed for the detection of ischemic stroke lesion tumors, allowing for image uploading and visualization. Automated segmentation of ischemic stroke lesions from magnetic resonance imaging images has an important role in the analysis, prognostic, diagnosis, and clinical treatment planning of some neurological diseases. Recently, computer-aided diagnosis systems based on deep learning techniques have demonstrated significant promise in medical image analysis, particularly in multi-modality medical image segmentation. Automated segmentation is a difficult task due to the enormous quantity of data provided by magnetic resonance imaging and the variation in the location and size of the lesion. In this study, we develop an automated computer-aided diagnosis system for the automatic segmentation of ischemic stroke lesions from magnetic resonance imaging images using a Convolution Block Attention Module (CBAM) and hybrid UNet-ResNet50 model. The UNet model is integrated into the architecture, and the ResNet50 backbone is pre-trained to enhance feature extraction. CBAM block is a model applied in this approach to extract the most effective feature maps. The proposed approach is evaluated on the public Ischemic Stroke Lesion Segmentation Challenge 2015 dataset, arranged into weighted-T1(T1), weighted-T2(T2), FLAIR, and DWI sequences. Experimental results demonstrate the efficacy of our approach, achieving an impressive accuracy value of 99.56%, a precision value of 97.12%, and a DC of 79.6%. Notably, our approach outperforms other state-of-the-art methods, particularly in terms of accuracy values, highlighting its potential as a robust tool for automated ischemic stroke lesion segmentation in magnetic resonance imaging.

***Keywords*—Medical image segmentation; ischemic stroke disease; UNet; ResNet50; convolution block attention module; magnetic resonance imaging; transfer learning**

I. Introduction

Stroke, characterized by the sudden onset of cerebral dysfunction, is a global health concern with significant mortality rates and long-term disability [1]. Recognized as a pandemic by the World Health Organization (WHO), the projected increase in stroke cases underscores its growing impact, with estimates reaching 23 million by 2030 [2], [3], [4]. This surge in incidence, now at 12 million cases annually, demands heightened attention from the medical community [5].

In the United States, approximately 800,000 people suffer from strokes each year, a number expected to significantly increase in the future due to the aging population [6]. In Tunisia, stroke stands as the leading cause of physical disability in adults and the third cause of death, with an annual incidence of around 10,000 new cases [7]. The aftermath of stroke, affecting around 70% of survivors, are often left with severe cognitive problems, requiring intensive and specialized care over a long period to facilitate recovery [3].

Stroke is categorized into two primary types: ischemic and hemorrhagic. In this study, we focuses on Ischemic Stroke Lesions (ISL) due to their widespread prevalence and the imperative for early intervention, including thrombolysis or thrombectomy. The automatic segmentation of ischemic stroke lesions from Magnetic Resonance Imaging (MRI) images plays a pivotal role in improving diagnosis, prediction, and treatment planning. Advanced neuroimaging modalities, particularly MRI, have proven indispensable in enhancing the efficiency and accuracy of stroke interventions.

In our research, we focused on Ischemic Stroke Lesions (ISL) because they affect a lot of people and require early intervention and treatment which will be acts of thrombolysis or thrombectomy. Additionally, the assessment of ischemic stroke lesions is a critical endpoint in clinical trials. It can help to improve how we diagnose, predict, find, and treat this condition. The evaluation of ischemic stroke lesions is a pivotal endpoint in clinical trials, offering insights that can enhance diagnosis, prediction, localization, and treatment. The advancements in medical imaging technology have significantly improved the intervention and clinical treatment of strokes, making them more efficient and accurate. Advanced neuroimaging modalities including Magnetic Resonance Imaging (MRI), Computed Tomography(CT scan), and Magnetic Resonance Angiography (MRA), etc, are used for ISL diagnosis. CT imaging can diagnose stroke patients who have tumors well, but it is not good at showing other parts of the brain or early signs of damage. Angiography means putting an injection of a contrast agent into the patient, which can affect their health and cause implications for the patient's body [8]. MRI emerges as the most effective tool for assessing patients with ischemic stroke. It is a non-invasive tool, more sensitive and it can accurately analyze the diseases. It can help to track the disease and predict the outcome. Furthermore, her sequence acquisitions (Diffusion-Weighted-Images (DWI), T1, T2, and Fluid-Attenuated-Inversion-Recovery (FLAIR)), can provide specific information about the extent of the lesion and her localization which represents the main clinical detail in detection processing. Its multi-parametric nature, encompassing various contrasts and sequences, positions MRI as an indispensable and highly effective method for the comprehensive assessment of ischemic stroke patients, thereby influencing critical aspects of their care and recovery. The ability to automatically detect infarct lesions proves invaluable for medical diagnosis, facilitating timely intervention and treatment planning. MRI stands as a cornerstone in post-stroke issue resolution, offering unparalleled advantages such as detailed disease monitoring, early lesion visualization, tissue characterization, and outcome prediction.

Ischemic strokes have a complex structure that makes their segmentation in MRI images difficult. Therefore, automatic stroke segmentation has been achieved by using Artificial Intelligence (AI) algorithms, which are very important for the medical research field. AI techniques, such as deep learning and machine learning, are very popular in the medical field. They can handle multidimensional medical data as well as a trained expert. Moreover, it is widely used and adopted in the process of medical imaging processing, especially in segmentation, classification, diagnosis, detection, and prognosis of stroke ischemia. It offers reliable, accurate, and consistent outcomes. This would reduce the testing time and enable the neuroradiologists to examine and interpret more data for their patients, which would be more cost-efficient.

Currently, various Computer-Aided Diagnosis (CAD) systems based on deep learning models are commonly developed and applied in ISL segmentation. Specifically, Convolutional neural networks (CNNs) are the most employed models for image classification and segmentation tasks in neurodegenerative diseases. CNN has demonstrated enormous potential in analyzing and characterizing medical images, including ischemic brain stroke segmentation. Furthermore, these models have leveraged the power of CNNs for image analysis and the advantages of statistical methods for data processing. This makes many patterns have been developed by researchers around the world.

However, CNNs often suffer from limitations when segmenting medical images, due to their limited spatial awareness and difficulty in processing diverse anatomical structures. To overcome these difficulties, the adoption of the Convolutional Block Attention Module (CBAM) is proving advantageous. CBAM remedies the rigidity of CNN by incorporating attention mechanisms that enhance spatial awareness, enabling the model to capture long-range dependencies in medical images. In addition, the dynamic importance of CBAM features facilitates adaptation to complex structures, enabling the network to focus on relevant regions and improve segmentation accuracy. This transition to CBAM represents a promising advance in medical image analysis, offering a more robust and flexible approach to image segmentation tasks.

In this paper, we propose a CAD system for ischemic stroke brain segmentation based on CBAM and a hybrid ResNet50-Unet model from MRI sequences to overcome these issues. We enhance the ResNet50-UNet architecture performance by integrating the CBAM block. The ResNet50-UNet framework combines the deep feature extraction capabilities of ResNet50 with the precise segmentation abilities of the UNet architecture. To introduce attention mechanisms and enrich feature representations, we meticulously inserted CBAM block after each convolutional block within the network. At each convolution step, the CBAM module dynamically computes channel-wise and spatial-wise attention, allowing the model to focus on relevant features and significant regions in the medical images. This customized integration gives our model greater spatial awareness and adaptability to complex anatomical structures, increasing segmentation accuracy compared to the conventional ResNet50-UNet architecture. The CBAM-enhanced ResNet50-UNet not only leverages the strength of both architectures but also capitalizes on the attention mechanisms to achieve more refined, context-sensitive segmentation of medical images.

In this study, we propose a systematic approach for the segmentation of ischemic stroke brain lesions using a Computer-Aided Diagnosis (CAD) system. Our method employs a hybrid ResNet50-Unet model, and we enhance its capabilities by integrating the Convolutional Block Attention Module (CBAM) into the processing of MRI sequences. The methodology unfolds through several key steps. We begin by integrating a CBAM block into the ResNet50-UNet architecture, a step designed to significantly improve the model's performance. This integration capitalizes on the deep feature extraction prowess of ResNet50 while harnessing the precise segmentation abilities inherent in the UNet architecture. Subsequently, attention mechanisms are introduced by strategically placing a CBAM block after each convolutional block within the network. This meticulous insertion allows for the dynamic computation of channel-wise and spatial-wise attention at every convolutional step. As a result, the model becomes adept at focusing on relevant features and significant regions in the medical images. To further refine the approach, we customize the integration, providing our model with heightened spatial awareness and adaptability to complex anatomical structures. This customization proves instrumental in increasing segmentation accuracy compared to the conventional ResNet50-UNet architecture. The culmination of these efforts results in a CBAM-enhanced ResNet50-UNet model that not only leverages the strengths of both architectures but also effectively utilizes attention mechanisms to achieve a more refined and context-sensitive segmentation of medical images. This approach holds promise for advancing the field of ischemic stroke diagnosis and treatment planning.

The remaining parts of this research paper are organized as follows: In section two, we review the current reviews state of the art techniques segmentation for ISL. In Section Three, we present the dataset description and preprocessing and, eventually, discuss the proposed method based on mechanism attention and ResNet50-UNet (CBAM ResnNet50-UNet). Section four reports the experimental results and compares our method with the results of the state-of-the-art multimodal MRI dataset Ischemic Stroke Lesion Segmentation Challenge (ISLES) 2015. We also address some important challenges and drawbacks of the methods we propose in the same section. Finally, the paper is concluded in Section Five.

II. Related Work

Machine learning (ML) and deep learning (DL) have transformed the landscape of medical imaging significantly, providing unparalleled capabilities in analyzing intricate datasets. Classical machine learning methods heavily depended on features manually designed for the analysis of brain images. Identifying abnormalities, particularly in cases such as brain ischemia, presented challenges owing to irregular shapes and ambiguous boundaries. Deep learning models, specifically CNNs, autonomously acquire both local and global features, proving crucial for early diagnosis. In the context of brain ischemia, various methods leveraging CNNs have been developed for automatic early detection and segmentation. This section provides an in-depth overview of leading methods for brain ischemia segmentation from multimodal MRI sequences. Utilizing the ISLES 2015 dataset, which aligns with my proposed approach, our analysis aims to capture insights accumulated between 2017 and May 2023.

Kamnitsas et al. [9] proposed an architecture with a dual pathway for a brain tumor and ischemic stroke segmentation, which processes a 3D Fully connected Conditional Random Field (CRF) to remove false positives. To overcome the computational load of processing 3D medical scans, they created a dense training scheme that is effective and efficient for processing 3D medical scans. They merge the processing of nearby image patches into one network pass and automatically adapt to the natural class imbalance in the data. They obtain an average dice coefficient equal to 69%.

Additionally, Liang et al. [10] introduced a framework to segment stroke lesions with DWI sequences that are based on two CNNs. One is a combination of two DeconvNets, which is the EDD Net; the other CNN is the multi-scale convolutional label evaluation net (MUSCLE Net), which aims to assess the lesions detected by the EDD Net and eliminate possible false positives. They tested their method on a large dataset including 741 subjects from DWI. The mean accuracy achieved is 67% in total. The mean Dice scores for subjects with only small and large lesions are 61% and 83%, respectively.

Zhiyang et al. [11] suggested a residual structured fully convolutional network (Res-FCN) based on 2D slices from DWI, ADC, and T2 WI. The suggested Res-FCN is trained and tested on ISLES 2015-SISS with 212 clinically acquired MRIs, which achieves a mean dice coefficient of 64.5% with a mean number of false negative lesions of 1.515 per subject.

Furthermore, Rongzhao et al. [12] used 3D contextual information and automatically learned features to propose an end-to-end model. To alleviate the hardness of training deep 3D CNN, they equipped the network with dense connectivity to allow the unimpeded propagation of information and gradients throughout the network. The Dice objective function was used to train the model to deal with the severe class imbalance problem in data. The model was built up with a DWI dataset with 242 subjects regrouped as 90 for training, 62 for validation, and 90 for testing. It achieved a Dice similarity coefficient, lesion-wise precision, and lesion-wise F1 score equal to 79.13%, 92.67%, and 89.25% respectively.

In 2019, Liangliang et al. [13] proposed a multi-kernel DCNN (MK_DCNN) composed of two symmetrical deep sub-networks, in which dense block are used to reduce the over-fitting problem of deep networks such as the extraction of effective features from sparse pixels. A multi-kernel and the dropout regularization method were used to split the network into two sub-networks for getting more sensory fields, and the dropout regularization method to achieve an effective feature mapping respectively. Then, they applied median filtering to reduce noise and preserve image edge detail. The developed architecture provided a dice coefficient of 57%, a symmetric surface distance value average equal to 2.01mm, and a Hausdroff distance equal to 2.38 mm with the SISS challenge dataset. With the same challenge, Amish et al. [14] created a modified U-Net model and a multi-path network. The model obtained an average dice coefficient of 70.07%, a sensitivity value of 49.28%, a specificity equal to 99.78%, and a precision of 98.72%.

In the same year, to perform segmentation, a model for the spatial arrangement of pixels preserved by learning the local characteristics of an image was proposed by Karthik et al. [15]. This research work presented a supervised fully convolutional network (FCN). The remarkable point of this research is the application of Leaky Rectified Linear Unit activation in the last two layers of the network for precise reconstruction of the ischemic lesion. This allows the network to learn from additional features that are not considered in the existing U-Net architecture. An average segmentation dice coefficient of 70% was obtained with ADAM optimizer based on experiments carried out on the ISLES 2015 dataset only on axial plane slices [15].

In 2020, Liangliang et al. [16] presented a Res-CNN based on a U-shaped structure and integrated the residential unit dual in the network to alleviate the degradation problem. Fusion methods and data augmentation were used before training the model to expand their dataset. The presented model obtained an average dice coefficient equal to 74.20% and a Hausdroff distance equal to 2.33mm. It's the same for Amish et al. [17] have also continued research on this topic. They proposed a new architecture based on the Classifier-Segmenter (CS-Net), which involves a hybrid learning strategy with a self-similar U-Net model, explicitly designed to perform the segmentation task. The advantage is to develop a cascade architecture, which improves the precision while removing redundant parts of the Segmenter's input. With the ISLES SISS-2015 dataset, they achieved a dice coefficient, a precision, and a recall of 63%, 74%, and 62% respectively. With ISLES 2017, they obtained a dice coefficient, a precision value, and a recall of 28%, 37%, and 45%, respectively.

Zhang et al. [18] presented a framework to quickly and automatically segment stroke lesions on DWI. First, they designed a detection and segmentation (DSN) to address data imbalance. Second, they proposed a triple-branch DSN architecture, which was used to extract the different features. Third, they developed a multi-plane fusion network (MPFN), which aimed to make the final prediction more accurate. The authors tested their methods on the ISLES 2015 SSIS DWI sequence dataset. Experimentally, they obtained dice coefficient and sensitivity values of 62.2% and 71.7%, respectively.

Liangliang et al. [19] have continued research on this contribution, now with a third approach after the two mentioned previously. In this point, a new network neural convolutional deep residual attention (DRANet) was proposed to accurately

TABLE I. SUMMARY OF EXISTING METHODS FOR MRI IMAGES SEGMENTATION

Authors	Methods	Dice Coefficient(%)	Accurracy(%)	Precision (%)
Kamnitsas et al. [9] (2017)	CRF	69	–	–
Liang et al.[10] (2017)	MUSCLE Net	61	67	–
Zhiyang et al.[11] (2018) (	Res-FCN	64.5	–	–
Rongzhao et al. [12] (2018)	3D CNN	79.13	–	92.67
Liangliang et al. [13] (2019)	MK-DCNN	57	–	–
Amish et al. [14] (2019)	UNet with multi-patchnetwork	70.07	–	98.72
Karthik et al.[15] (2019)	FCN	70	–	
Liangliang et al. [16] (2020)	Res-CNN	74.2	–	–
Amish et al. [17](2020)	CS-Net	63	–	74
Zhang et al. [18] (2020)	Multi-plane fusion network	62.2	–	–
Liangliang et al. [19] (2020)	DRANet	76	–	–
Aboudi et al.[20] (2022)	Unet	55.77	99.96	–
Aboudi et al.[21] (2022)	Hybrid ResNet50-Unet	64.14	99.43	–

and simultaneously segment and quantify lesions of stroke and white matter hyperintensity (WMH) in MRI images. Their solution's key architectural features are the use of the residual block and the Dice loss function to make the network training effective, as well as the use of the attention modules to produce a high-quality representation of the input images inside the network. Their proposed DRANET model obtains high-quality features from the input images. DRANet was trained and evaluated on 742 2D MRI images which are generated from (SISS) challenge and their approach achieves a dice coefficient of 76%.

Recently, Aboudi et al. [20], [21] developed two contributions to this idea. The first consists of developing a deep convolutional neural network (CNN) method inspired by the U-Net architecture. They applied a finetuning technique to adapt the U-Net architecture to our objectives. They evaluated her method on the public dataset ISLES 2015. Their model achieved a Dice Coefficient (DC) and accuracy equal to 55.77%, and 99.96% respectively. The second research consists of combining the UNet model with a pre-trained ResNet50 architecture to form a hybrid framework. They apply data augmentation techniques to improve the model's accuracy. They trained and tested their method on the ISLES 2015 dataset. The experimental results show the effectiveness of our method, which achieves a 99.43% average accuracy, and a 64.14% Dice Coefficient(DC).

The assessment of existing surveys underscores the rapid evolution of approaches in medical imaging, leveraging the capabilities of machine learning (ML) and deep learning (DL). While DL models, particularly CNNs, have shown substantial promise, persistent challenges and limitations have been identified. This observation prompts the introduction of attention mechanisms, such as CBAM, in our approach. Existing studies, including models like U-Net, have demonstrated success in brain ischemic lesion segmentation. However, issues like class balance, false positives, and effective 3D data processing persist. Noteworthy models like Res-FCN, Res-CNN, and innovative approaches like DRANet with attention modules

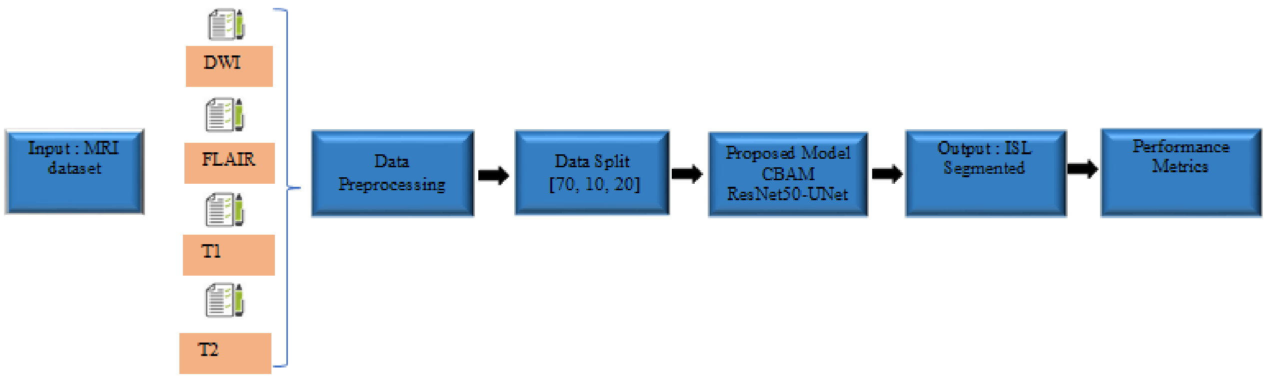

Fig. 1. Flowchart of the proposed approach.

have showcased impressive results. The summary Table I of existing methods highlights diverse approaches and metrics like Dice coefficient, precision, and accuracy. Despite these advancements, limitations in terms of precision and sensitivity justify the need for our CBAM-based approach and attention mechanism to address persistent challenges in accurate brain ischemic lesion segmentation. In essence, the analysis of past works underscores progress but emphasizes the ongoing need for improvements, justifying the introduction of attention mechanisms in our proposed approach.

III. Motivations and Contributions

In this research, we propose an attention model to enhance the performance of ISL segmentation. Only a handful of publications have explored the integration of attention block in segmentation models. Many researchers focus on improving the segmentation outcome without considering task efficiency. Thus, the most crucial aspect in any ML or DL models is to extract minimal yet valuable features. To address this issue, we will apply an attention-based mechanism to select relevant features from the entire MRI and use them for segmentation. Additionally, to minimize the algorithmic and computational complexity of the task, we will employ transfer learning instead of training a complete neural network from scratch, pre-trained with ResNet-50. This technique helps us improve segmentation performance while maintaining task accuracy. Therefore, we summarise our main contributions in this research paper as follows:

- An advanced CAD system for ischemic stroke brain segmentation based on CBAM and a hybrid ResNet50-Unet model from MRI sequences (DWI, T1, T2, and Flair) was developed.
- CBAM block was integrated after each convolution block into the ResNet50-UNet architecture to enhance the model performance and to enhance feature representation by emphasizing relevant channels and adjusting spatial perception.
- Significance evaluation of predicted ground truth for ISL segmentation in MRI sequences.

IV. Proposed Method

In this section, we provide a detailed explanation of the proposed CAD system for ischemic stroke brain segmentation, which relies on CBAM and a hybrid ResNet50-Unet model applied to MRI sequences (DWI, T1, T2, and Flair). Firstly, we present the dataset description and preprocessing steps. Subsequently, we delve into the detailed information about the architecture of the proposed CBAM ResNet50-Unet model. We also elaborate on the reasons for selecting this specific method. Fig. 1 illustrates an overview of our approach.

A. Dataset Description

The multimodal Ischemic Stroke Lesion Segmentation Challenge (ISLES) 2015, provided by the University Medical Center Schleswig Holstein (UKSH) Germany and the departments of Neuroradiology Hospital Rechts der Isar in Munich was used to perform this research study. The data acquisition parameters encompass a slice thickness of 5 mm, an echo time of 87 ms, and a repetition time of 3200 ms. ISLES 2015 dataset is reorganized as follows: SISS, which was employed in our work and involved segmenting sub-acute ischemic stroke lesions, and SPES, which involves estimating the stroke penumbra. This work uses a training dataset supplied by SISS that consists of 28 subject cases and a manually segmented and annotated ground truth. This dataset was acquired from 3T Philips systems and arranged into weighted-T1(T1), weighted-T2(T2), FLAIR, and DWI(b=1000) sequences with 57 to 154 slices. We have obtained a total of 4312 images. Fig. 2 illustrates an example of the four methods of MRI images. Each subject's data is represented in NIFTI(.nii) format, featuring an image shape of 240x240x155x3.

B. Preprocessing

The raw MRI images underwent preprocessing before being employed as input in our proposed approach. Each pixel from the original 3D images is converted into a series of 2D images. To enhance the dataset, non-relevant black images are removed, as they do not fall within the brain tumor category, and utilizing the entire image is unnecessary for medical image analysis. During the preprocessing step, images with black slices lacking information are excluded. Our focus centers on eight slices, commencing from slice 22, where

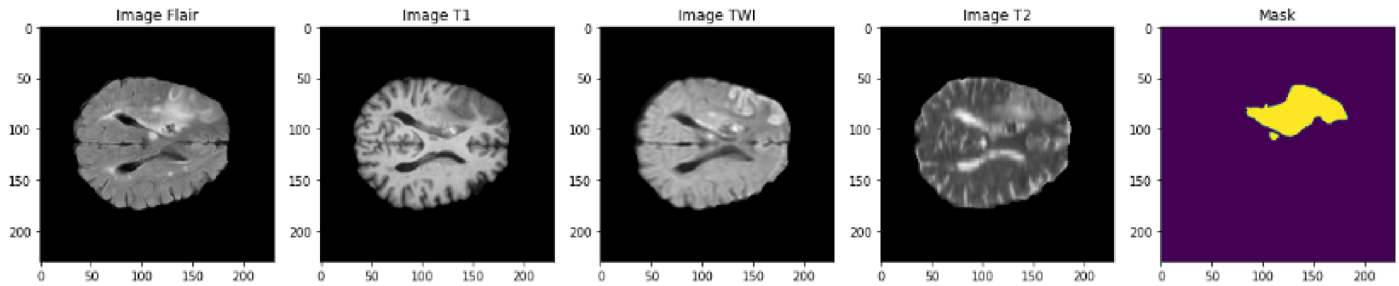

Fig. 2. The four MR image modalities are displayed, including Flair, DWI, T1, T2, and ground truth examples.

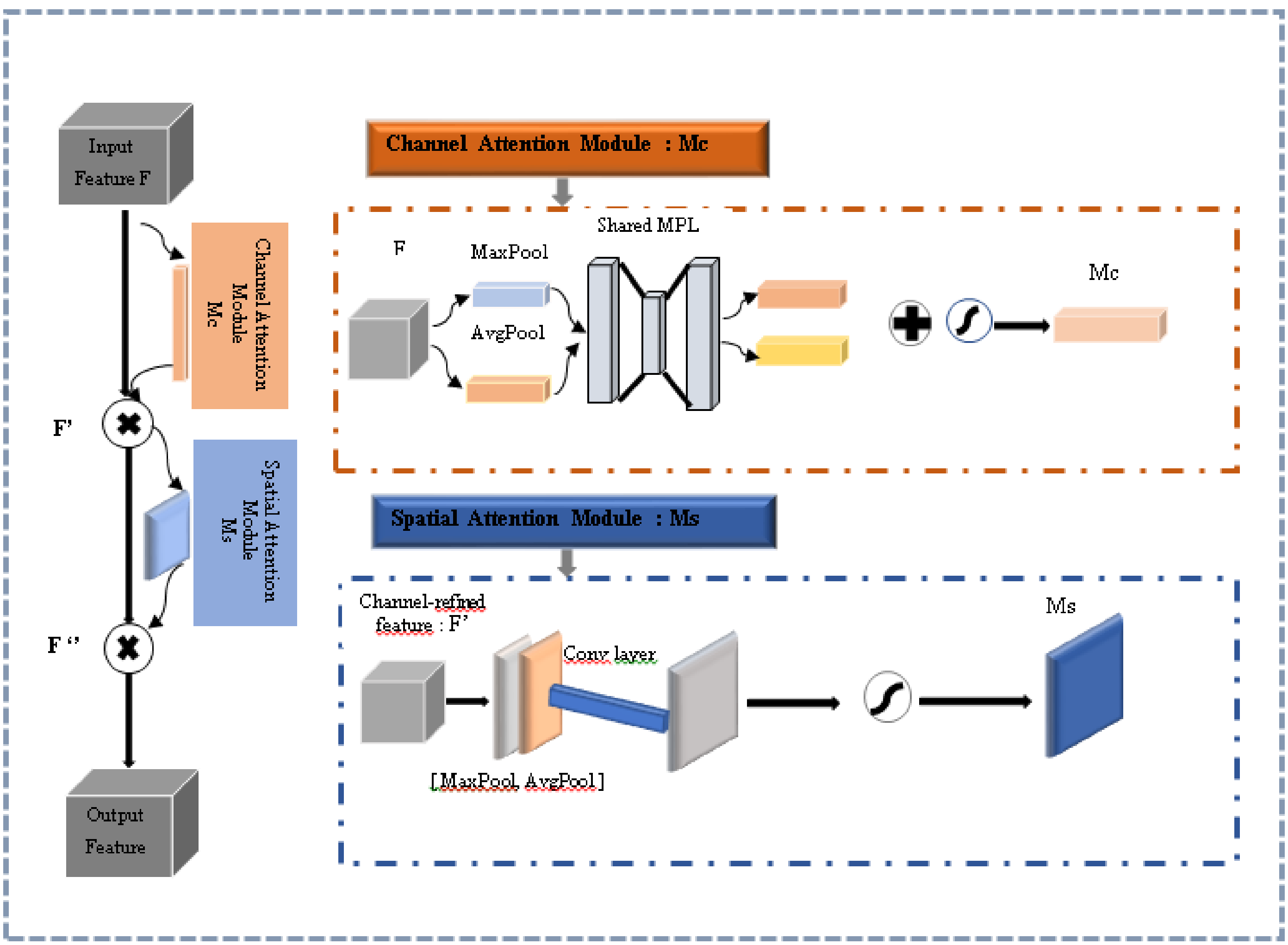

Fig. 3. Architecture of the CBAM block.

we define our Regions of Interest (ROIs). Additionally, all images undergo resizing to 128x128x3, ensuring uniformity in dimensionality. The dataset is partitioned into distinct subsets: 70% for training, 20% for validation, and 10% for testing.

C. CBAM Model

CBAM, introduced by Woo et al., serves as an attention module designed to facilitate channel size and spatial operations within CNN architectures. This module comprises two blocks(modules): the channel attention module and the spatial attention module. Fig. 3 describes the CBAM block architecture.

1) Channel Attention Module: Channel attention focuses on discerning the meaningful content within an input image, determining 'what' features are significant. This is achieved by analyzing inter-channel relationships to enhance feature selection and extraction while minimizing loss values. The calculation of channel attention involves compressing the spatial dimension of the input feature map. To aggregate spatial information, Global Average Pooling (GAP) and max-pooling layers are employed. GAP obtained the aggregate information(calculating the average for each patch of the feature map) when the max-pooling layer reached the differences of features(calculates the maximum). The fusion of the GAP layer and max-pooling layer performed better than using a single layer to produce a weighted channel descriptor (Fig. 5). Channel attention is calculated as follows:
$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$

$$= \sigma(W_1(W_0(F_c^{avp})) + W_1(W_0(F_c^{max})))$$

where F_c^{avp} and F_c^{max} denote the squeezed feature maps of the two pooling layers, σ represents the sigmoid function, $W_0 \in R^{r/C*C}$ and $W_1 \in R^{C*C/r}$ represent the weights of MLP.

2) Spatial Attention Module: It is based on 'where' is a relevant part. This module used the inter-spatial relationship between features to calculate the spatial attention map. Firstly, GAP and max-pooling functions are applied on the feature maps generated during the previous step (channel attention module) and squeezed into 2D features maps: $F_{avg}^s \in R^{1*H*W}$ and $F_{max}^s \in R^{1*H*W}$. After that, a convolution layer is added on the recombined two 2D feature maps to generate an efficiency feature maps. Then, a sigmoid function is used to calcule the spatial attention map $M_s(F) \in R^{H*W}$. Spatial attention is calculated as follows:
$M_s(F) = \sigma(f^{7x7}([AvpPool(F); MaxPool(F)]))$

$$= \sigma(f^{7x7}([F_{avg}^s; F_{max}^s]))$$

when σ is a sigmoid function and f^{7x7} denotes une convolution function with filter size of $7x7$.

In Spatial Attention Module, GAP and max pooling functions are used to generate spatial feature descriptor then a $7x7$ convolution filter used to emphasize a spatial information.

D. Pre-trained ResNet50 Model

ResNet50 stands as a prominent member of the Residual Network family architecture and is currently among the most widely utilized models in the field of image recognition. It was introduced by Kaiming He et al.[22] in their influential paper "Deep Residual Learning for Image Recognition. ResNet50 has demonstrated exceptional performance across diverse vision applications, encompassing tasks such as classification, object detection, and semantic segmentation. Fig. 4 illustrates the ResNet50 architecture. It comprises 48 convolutional layers organized into four stages, featuring residual blocks with shortcut connections to address the challenges of training very deep. ResNet50 effectively addresses the vanishing gradient problem. The architecture's bottleneck design optimizes computational efficiency, incorporating 1x1 convolutions for feature map dimensionality reduction. With Global Average Pooling (GAP) as a concluding layer and pre-trained weights often obtained from datasets like ImageNet, ResNet50 stands out for its ability to provide compact representations and serve as a robust foundation for transfer learning. This model's advantages encompass its proficiency in mitigating training challenges, its adaptability to diverse computer vision tasks, and its consistent delivery of state-of-the-art performance in image classification and segmentation.

The pre-trained ResNet50 model comprises four key blocks, each referred to as a stage. The initial block involves data preprocessing and feature extraction, followed by three residual blocks within different stages. These residual blocks, incorporating multiple convolutional layers, contribute to the network's depth and capacity to capture intricate features. This architecture, known for its efficacy in image recognition, has achieved state-of-the-art results in various visual tasks due to its depth, skip connections, and the ability to train deep networks effectively.

E. Unet Model

UNet model stands as the common model choice for segmentation tasks in the medical imaging domain, such as segmenting organs or tumors in medical scans. It was developed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-Net is characterized by a U-shaped architecture, featuring an encoder pathway and a symmetric decoder pathway. It adeptly gathers low-level features that encapsulate information about the shapes of different classes, alongside high-level features that leverage this information to discern the class to which each position belongs. The architecture seamlessly integrates both low-level and high-level features to reconstruct the original input's spatial dimensions, ensuring precise classification for each location. UNet architecture is comprised of two fundamental blocks: the encoder and the decoder. The encoder captures contextual information through downsampling, while the decoder reconstructs the segmented image with fine-grained details using upsampling. The unique aspect of U-Net is the incorporation of skip connections that connect corresponding layers between the encoder and decoder, aiding in the precise localization of segmented objects. U-Net's ability to handle limited labeled data and its effectiveness in preserving spatial information makes it a popular choice for image segmentation tasks.

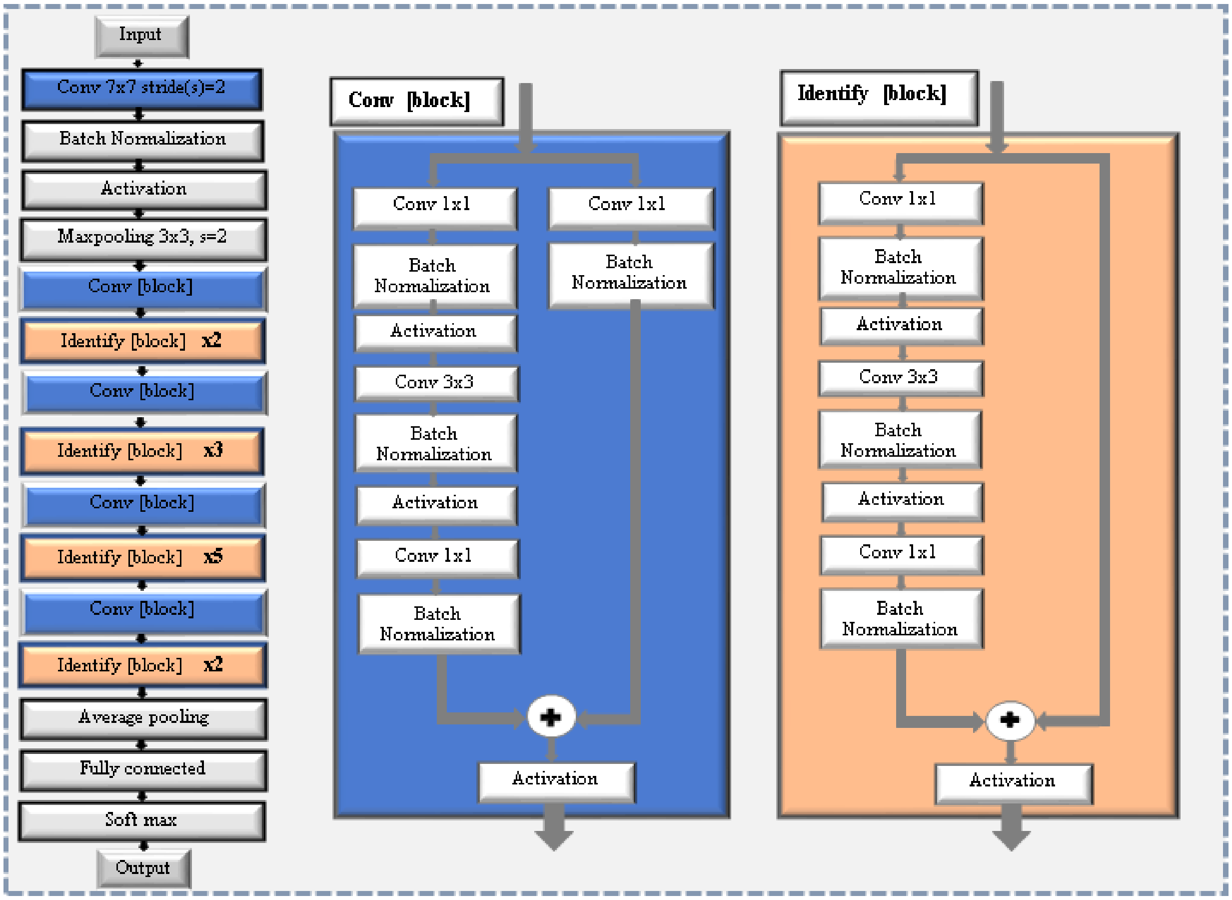

Fig. 4. The architecture of the pre-trained ResNet50 model.

F. *Proposed Hybrid CBAM ResNet50-Unet Model*

The UNet model architecture is organized into two primary blocks: the encoder and the decoder. In the encoder network, the input is processed through successive convolutional blocks, each succeeded by max-pooling layers for downsampling. These blocks progressively extract hierarchical features, capturing both low-level and high-level representations of the input. Subsequently, the decoder network begins with upsampling operations using transposed convolutions (Conv2DTranspose). Each upsampling step is accompanied by concatenation with the corresponding feature maps from the encoder, creating skip connections. Convolutional blocks in the decoder then process this concatenated information, gradually restoring the spatial dimensions of the original input. The final layer, utilizes a 1x1 convolution with softmax activation, producing a pixel-wise classification map. This architecture seamlessly integrates the encoding of input features and the decoding of spatial information, making it well-suited for segmentation tasks.

The backbone of the ResNet50 model is added into the encoder bloc only. We freeze the encoder layer using fine-tuning and transfer learning in the pre-trained ResNet50 model, entrain the absence of an update of weighted layer during the execution of training data. Instead, the weight of the convolutional layer off ResNet50 will be used. First, we modified the architecture of the ResNet50 to be similar to UNet, adding an expanding layer composed of multiple up-sampling layers and convolution layers at the end of her structure. Second, this is carried out up until the model's overall architecture is symmetric and takes the form shape of a U-Net. Such as this combination the trainable parameters models will be reduced. After, we train the input data MRI using the two proposed hybrid models with a transfer learning method.

In this approach, CBAM block has been thoughtfully incorporated after each convolutional block in the model. Each convolutional block consists of a sequence of convolutional layers, followed by a dropout operation, and finally, batch normalization. The novelty lies in adding the CBAM module after this layer sequence, designed to enhance feature representation by emphasizing relevant channels and adjusting spatial perception. This approach strengthens the model's ability to capture meaningful information while preserving spatial coherence. The addition of the CBAM module after each convolutional block contributes to enhancing the model's performance, especially in image segmentation tasks where the accuracy of representations is crucial. The method allows for the selective integration of attention mechanisms throughout the network, which can be beneficial for the model's prediction quality.

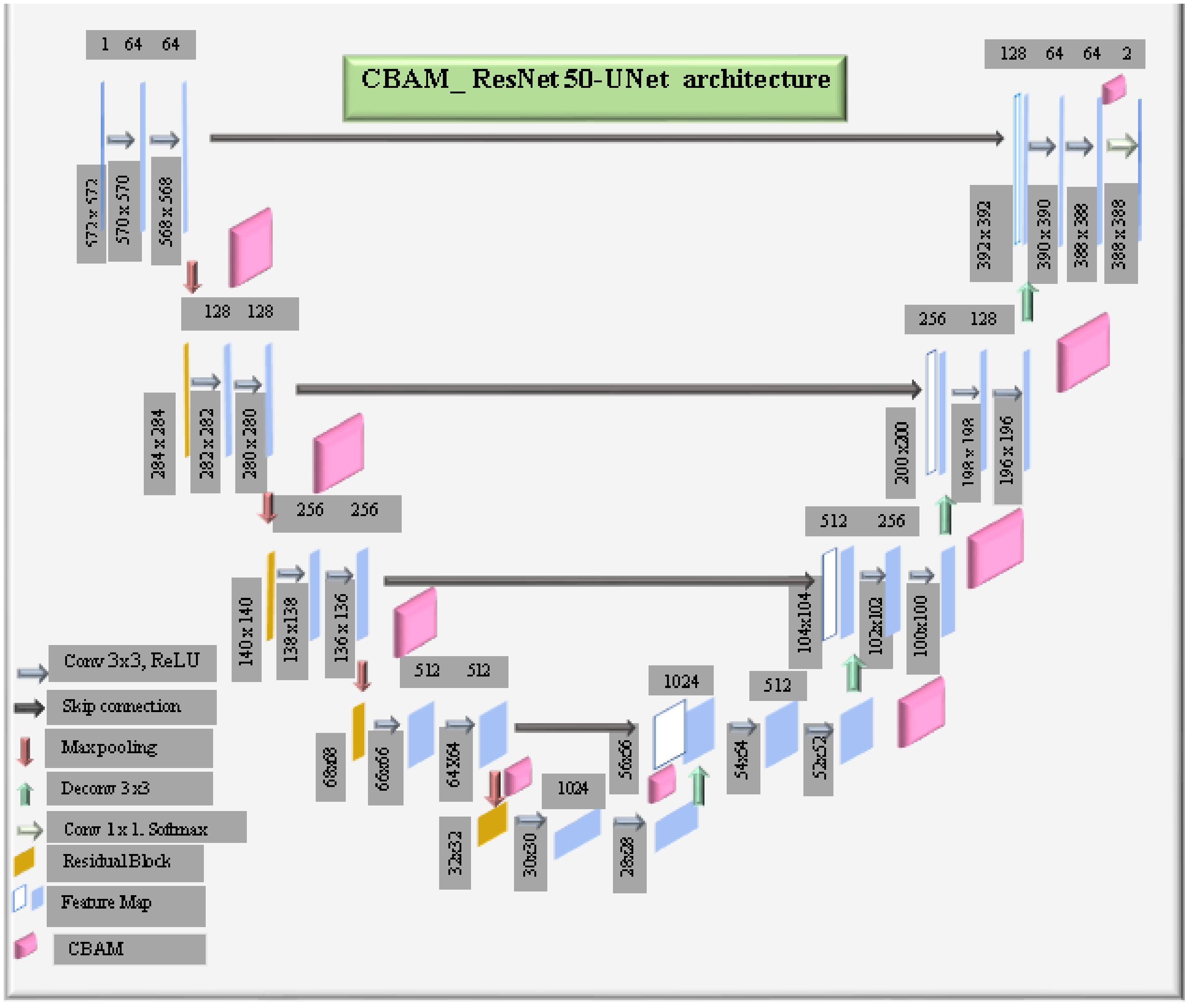

Fig. 5. The detailed architecture of the CBAM ResNet50-Unet model.

V. Experimental Results and Discussion

This section highlights the obtained experimental findings using our model based on CBAM ResNet50-Unet to segment the Ischemic Stroke Lesion tested on the ISLES 2015 MRI images.

A. Implementation Details

The experiments were conducted using the Python language in the Ubuntu environment on a Dell computer equipped with an Intel Core i5 8th generation processor and 8 GB of RAM. The system also featured an NVIDIA GeForce GTX 1050 graphics card with NVIDIA Driver Version 440.59, and the Kaggle framework was utilized. Keras and TensorFlow served as the primary frameworks for implementing the architecture.

To create the optimal model, various variables have been tested and the most effective parameters are chosen. The proposed CBAM ResNet50-UNet architecture was trained with the following parameters: a batch size of 32, 100 epochs, stochastic gradient descent (SGD) optimizer with a learning rate of 0.00001, momentum set to 0.9, and number of epochs of 100. The binary cross-entropy loss function was plotted against the epoch number, and a categorical cross-entropy loss function was also employed.

B. Evaluation Metrics

To assess the efficacy of our approach, the Dice Coefficient (DC) was employed as the primary metric, as well as specificity, sensitivity, precision, and recall. DC quantifies the spatial overlap between the automatically generated segmentation output and the ground truth. Specificity gauges the network's proficiency in predicting healthy tissues, while sensitivity evaluates its ability to identify lesions accurately. Precision, also known as positive predictive value, measures the proportion of relevant outcomes among the predicted positive instances. Accuracy serves as a measure of the overall effectiveness of the proposed approach, capturing the ratio of correct predictions to the total instances. Additionally, recall assesses the model's capability to correctly classify the total

TABLE II. A COMPARISON OF OUR APPROACH AND STATE-OF-THE-ART METHODS TESTED ON THE ISLES 2015 DATASET

Authors	Methods	Dice Coefficient(%)	Accurracy(%)	Precision (%)
Kamnitsas et al. [9] (2017)	CRF	69	–	–
Rongzhao et al. [12] (2018)	3D CNN	79.13	–	92.67
Liangliang et al. [13] (2019)	MK-DCNN	57	–	–
Amish et al. [14] (2019)	UNet with multi-patchnetwork	70.07	–	98.72
Zhang et al. [18] (2020)	Multi-plane fusion network	62.2	–	–
Liangliang et al. [19] (2020)	DRANet	76	–	–
Aboudi et al.[20] (2022)	Unet	55.77	99.96	–
Aboudi et al.[21] (2022)	Hybrid ResNet50-Unet	64.14	99.43	–
Our approach	**CBAM ResNet50-Unet**	**79.6**	**99.56**	**97.12**

relevant results. This comprehensive set of metrics provides a thorough evaluation of the performance of the proposed approach across various aspects of segmentation quality and predictive accuracy. These metrics are defined as [23]:

$DC = \frac{2TP}{2TP+FP+FN}$,

$Accuracy = \frac{TP+FP}{TP+FP+TN+FN}$,

$Precision = \frac{TP}{TP+FP}$,

where : TP = True Positive , FP = False Positive, FN = False Negative, and TN = True Negative

C. Experimental Results

The proposed hybrid model, CBAM ResNet50-UNet, leverages the advantages of transfer learning with a ResNet50 model to enhance performance, particularly in overcoming the challenges associated with a small dataset. The use of a pre-trained ResNet50 model allows for the transfer of knowledge from a large dataset to our specific problem domain, aiding in faster convergence and improved accuracy. Moreover, to further boost the model's capabilities, CBAM is incorporated after each convolutional block. CBAM enhances feature discriminability by capturing both spatial and channel-wise attention. This addition helps the model focus on important features, promoting better segmentation results. UNet being a complex architecture, typically demands a substantial amount of training time, and its performance can be influenced by computer specifications. The integration of CBAM aims to address these challenges, making the proposed hybrid model more efficient in handling segmentation tasks. During experimentation, various scenarios were tested to identify the optimal model based on loss and accuracy metrics. Additionally, we calculated DC, precision, and accuracy values to comprehensively compare the segmentation results of the model with the ground truth values, demonstrating the effectiveness of the proposed approach.

Our proposed approach, CBAM ResNet50-UNet is tested on the ISLES 2015 dataset. Fig. 7 and 8 illustrated an example of prediction outputs and ground truth comparisons for the CBAM ResNet50-Unet model applied to the DWI and T2 sequences. These figures provide a comprehensive visual representation of the model's performance in accurately delineating lesions in each sequence, highlighting the efficacy of our segmentation approach.

Fig. 6 visually presents the performance metrics, including accuracy and loss, to provide a comprehensive evaluation of the proposed method. These curves offer a detailed insight into

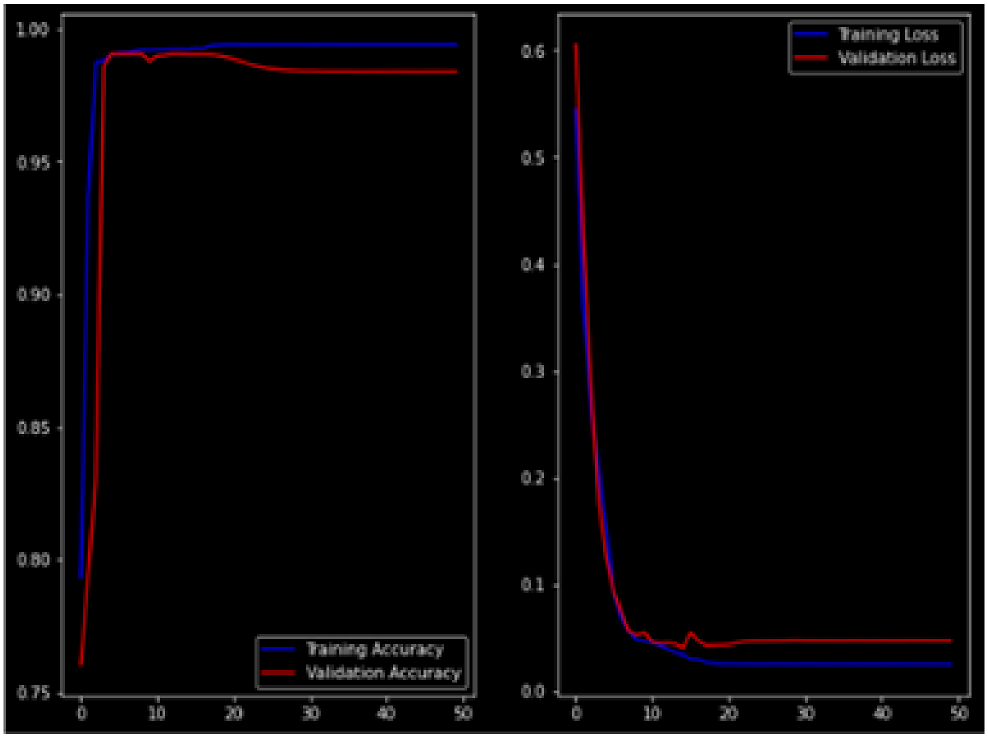

Fig. 6. Training and validation curves of accuracy and loss metrics for the CBAM ResNet50-Unet model.

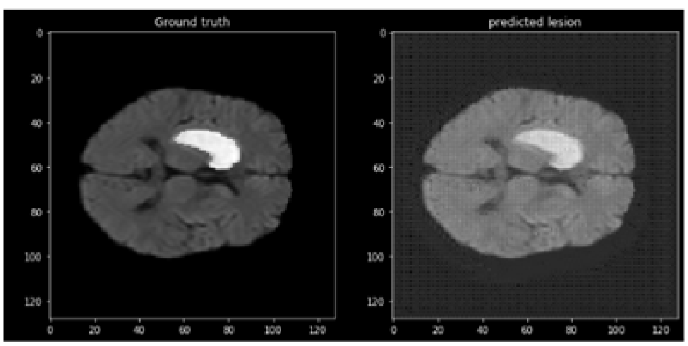

Fig. 7. Prediction outputs and ground truth comparisons for the CBAM ResNet50-Unet model applied to the DWI sequence.

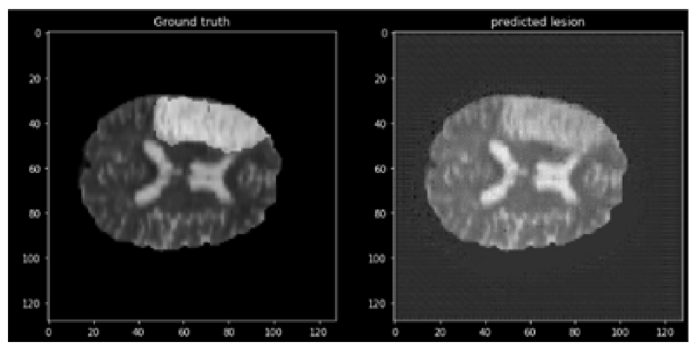

Fig. 8. Prediction outputs and ground truth comparisons for the CBAM ResNet50-Unet model applied to the T2 sequence.

how the model evolves over iterations, illustrating its capability to achieve accurate predictions, minimize loss, and optimize segmentation quality, thereby supporting the effectiveness and robustness of our proposed approach.

D. Comparison with State of the Art Studies

Table II compares the proposed model with state-of-the-art models that were trained on the ISLES 2015 dataset. Our CBAM ResNet50-Unet model stands out with a Dice Coefficient of 79.6%, an accuracy of 99.56%, and a precision of 97.12%. These results surpass the performance of numerous state-of-the-art approaches, highlighting the effectiveness of our model in predicting ISL MRI sequence outputs.

E. Discussion

This study presents our new approach CBAM ResNet50-UNet model for the segmentation of medical images, particularly focusing on lesions in the context of ISL segmentation. Through extensive experimentation and evaluation on the ISLES 2015 dataset, our proposed model has demonstrated remarkable performance, outperforming four state-of-the-art models in terms of accuracy, loss, and DC.

The integration of transfer learning with a pre-trained ResNet50 model, coupled with the incorporation of CBAM after each convolution block, has proven to be pivotal in enhancing the model's ability to discern intricate features and optimize segmentation accuracy. The visual representation in Fig. 10 and 9 showcases the model's efficacy in accurately delineating lesions across multiple sequences.

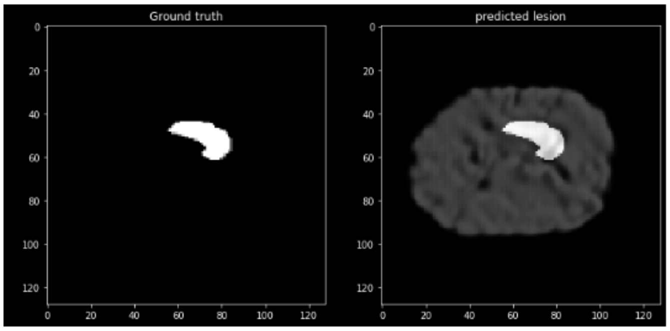

Fig. 9. Example 1 of prediction outputs and ground truth comparisons for the CBAM ResNet50-Unet model.

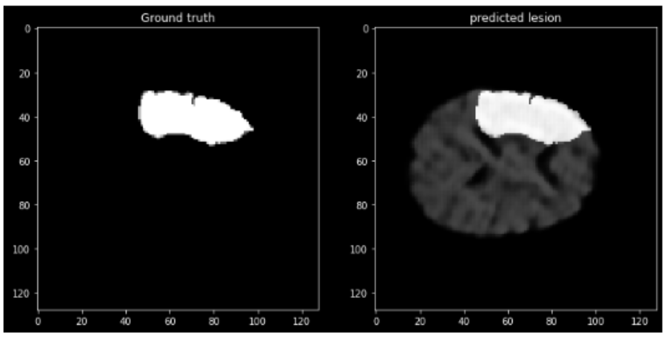

Fig. 10. Example 2 of prediction outputs and ground truth comparisons for the CBAM ResNet50-Unet model.

Moreover, the comprehensive analysis presented in Table II underscores the superiority of our proposed model in comparison to existing approaches. The achieved results not only validate the effectiveness of our methodology but also signify its potential for real-world applications in ischemic stroke lesions detection.

As illustrated in Fig. 11, our proposed CBAM Resnet50-UNet model, when evaluated on the ISLES 2015 dataset, demonstrated notable performance metrics. Specifically, it achieved a Dice coefficient of 79.6 %, an accuracy of 99.6 %, and a precision value of 97.1 %. Comparing these results to existing methods incorporating CRF, 3D CNN, MK-DCNN, UNet with multi-patch network, DRANet, Unet, and Hybrid ResNet50-Unet, our model outperformed them with Dice coefficients of 69 %, 79.1 %, 57 %, 70.7 %, 62.2 %, 76 %, 55.8 %, and 64.1 %, respectively. In terms of accuracy, our model achieved a significantly higher value of 99.9 %, surpassing all previous related works. Notably, Aboudi et al. reported an accuracy of 99.9 % in 2022, whereas their earlier works in the same year recorded values of 99.4 % and 99.7 %. Additionally, while the precision in model [14] reached 98.7 %, our CBAM ResNet50-Unet demonstrated a commendable precision value of 97.1 %

VI. Conclusion

In conclusion, this research paper introduces a novel CAD system for ischemic stroke brain segmentation in MRI sequences, integrating CBAM and a hybrid ResNet50-Unet

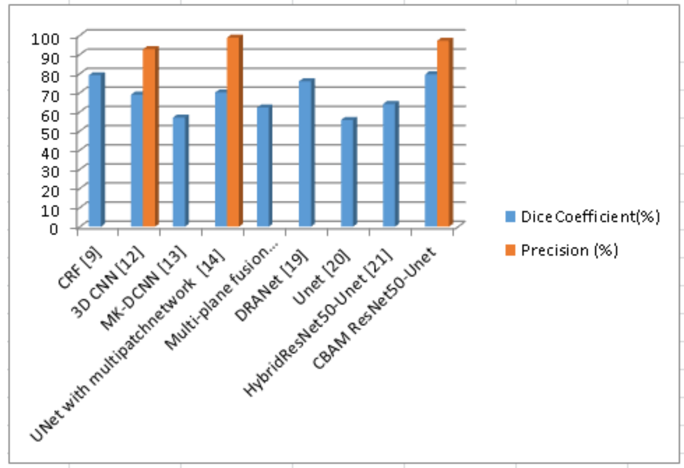

Fig. 11. Proposed CBAM Resnet50-UNet model.

model. Our motivation stems from the need to enhance both segmentation outcomes and task efficiency. The developed CAD system demonstrates promising results on the multimodal ISLES 2015 dataset, achieving a Dice Coefficient of 79.6% and a precision value of 97.12%. The integration of the CBAM block into the ResNet50-Unet architecture gives our model greater spatial awareness and adaptability to complex anatomical structures, enhances feature selection and extraction, and improves feature representation and segmentation accuracy. These findings contribute significantly to the field of ischemic stroke brain segmentation, laying the foundation for future work in unsupervised segmentation models and further refinement of attention mechanisms within our system.

References

[1] C. O. Johnson, M. Nguyen, G. A. Roth, E. Nichols, T. Alam, D. Abate, F. Abd-Allah, A. Abdelalim, H. N. Abraha, N. M. Abu-Rmeileh, et al., Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the global burden of disease study 2016, The Lancet Neurology 18 (5) (2019) 439–458.

[2] S. Doyle, F. Forbes, A. Jaillard, O. Heck, O. Detante, M. Dojat, Subacute and chronic ischemic stroke lesion mri segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3, Springer, 2018, pp. 111–122.

[3] M. Hommel, S. T. Miguel, B. Naegele, N. Gonnet, A. Jaillard, Cognitive determinants of social functioning after a first ever mild to moderate stroke at vocational age, Journal of Neurology, Neurosurgery & Psychiatry 80 (8) (2009) 876–880.

[4] N. Gupta, A. Mittal, Brain ischemic stroke segmentation: a survey, Journal of Multi Disciplinary Engineering Technologies 8 (1) (2014) 1.

[5] V. L. Feigin, Anthology of stroke epidemiology in the 20th and 21st centuries: Assessing the past, the present, and envisioning the future, International journal of Stroke 14 (3) (2019) 223–237.

[6] T. van der Zijden, A. Mondelaers, L. Yperzeele, M. Voormolen, P. M. Parizel, Current concepts in imaging and endovascular treatment of acute ischemic stroke: implications for the clinician, Insights into Imaging 10 (1) (2019) 1–10.

[7] N. Yengui, I. Rebai, H. Benrhouma, A. N. Ben, A. Rouissi, I. Kraoua, I. Turki, Accidents vasculaires cérébraux secondaires aux hémopathies chez l'enfant, revue neurologique 173 (2017) S175.

[8] M. Sheng, W. Xu, J. Yang, Z. Chen, Cross-attention and deep supervision unet for lesion segmentation of chronic stroke, Frontiers in Neuroscience 16 (2022) 836412.

[9] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation, Medical image analysis 36 (2017) 61–78.

[10] L. Chen, P. Bentley, D. Rueckert, Fully automatic acute ischemic lesion segmentation in dwi using convolutional neural networks, NeuroImage: Clinical 15 (2017) 633–643.

[11] Z. Liu, C. Cao, S. Ding, Z. Liu, T. Han, S. Liu, Towards clinical diagnosis: Automated stroke lesion segmentation on multi-spectral mr image using convolutional neural network, IEEE Access 6 (2018) 57006–57016.

[12] R. Zhang, L. Zhao, W. Lou, J. M. Abrigo, V. C. Mok, W. C. Chu, D. Wang, L. Shi, Automatic segmentation of acute ischemic stroke from dwi using 3-d fully convolutional densenets, IEEE transactions on medical imaging 37 (9) (2018) 2149–2160.

[13] L. Liu, F.-X. Wu, J. Wang, Efficient multi-kernel dcnn with pixel dropout for stroke mri segmentation, Neurocomputing 350 (2019) 117–127.

[14] A. Kumar, A. Debnath, T. Tejaswini, S. Gupta, B. Chakraborty, D. Nandi, Automatic detection of ischemic stroke lesion from multimodal mr image, in: 2019 Fifth International Conference on Image Information Processing (ICIIP), IEEE, 2019, pp. 68–73.

[15] R. Karthik, U. Gupta, A. Jha, R. Rajalakshmi, R. Menaka, A deep supervised approach for ischemic lesion segmentation from multimodal mri using fully convolutional network, Applied Soft Computing 84 (2019) 105685.

[16] L. Liu, S. Chen, F. Zhang, F.-X. Wu, Y. Pan, J. Wang, Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality mri, Neural Computing and Applications 32 (2020) 6545–6558.

[17] A. Kumar, N. Upadhyay, P. Ghosal, T. Chowdhury, D. Das, A. Mukherjee, D. Nandi, Csnet: A new deepnet framework for ischemic stroke lesion segmentation, Computer Methods and Programs in Biomedicine 193 (2020) 105524.

[18] L. Zhang, R. Song, Y. Wang, C. Zhu, J. Liu, J. Yang, L. Liu, Ischemic stroke lesion segmentation using multi-plane information fusion, IEEE Access 8 (2020) 45715–45725.

[19] L. Liu, L. Kurgan, F.-X. Wu, J. Wang, Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease, Medical Image Analysis 65 (2020) 101791.

[20] F. Aboudi, C. Drissi, T. Kraiem, Efficient u-net cnn with data augmentation for mri ischemic stroke brain segmentation, in: 2022 8th International Conference on Control, Decision and Information Technologies (CoDIT), Vol. 1, IEEE, 2022, pp. 724–728.

[21] F. Aboudi, C. Drissi, T. Kraiem, A hybrid resnet50-unet model for ischemic stroke brain segmentation from mri images, Vol. 22, 2022, p. 467.

[22] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[23] Y. Chen, L. Xing, L. Yu, H. P. Bagshaw, M. K. Buyyounouski, B. Han, Automatic intraprostatic lesion segmentation in multiparametric magnetic resonance images with proposed multiple branch unet, Medical physics 47 (2020) 6421–6429.