RESEARCH ARTICLE

MAGRes-UNet: Improved Medical Image Segmentation Through a Deep Learning Paradigm of Multi-Attention Gated Residual U-Net

TAHIR HUSSAIN AND HAYARU SHOUNO, (Member, IEEE)

Graduate School of Informatics and Engineering, Department of Informatics, The University of Electro-Communications, Tokyo 182-8585, Japan

Corresponding authors: Tahir Hussain (f2240014@gl.cc.uec.ac.jp) and Hayaru Shouno (shouno@uec.ac.jp)

ABSTRACT Precise segmentation is vital for successful diagnosis and treatment planning. Medical image segmentation has demonstrated remarkable advances with the introduction of deep convolutional neural networks, particularly encoder-decoder networks such as U-Net. Despite their excellent performances, these methods have some limitations. First, the structure is limited in its ability to combine information because feature maps to extract valid information from the final encoding stage are incompatible at the encoding and decoding levels. Second, the approach ignores significant semantic details and does not consider different types of small-scale contextual information when segmenting medical images. Lastly, most methods employing 3D architectures to process input medical images increase the computational complexity of the model without significantly improving the accuracy. To resolve these issues, we propose a segmentation network called Multi-Attention Gated Residual U-Net (MAGRes-UNet). This network incorporates four multi-attention gate (MAG) modules and residual blocks into a standard U-Net structure. The MAG module integrates the information from all encoding stages and focuses on small-scale tumors while disambiguating irrelevant and noisy feature responses, thereby promoting meaningful contextual information. The residual blocks simplify the network training and mitigate the problem of vanishing gradients. This improves the ability of the network to effectively learn intricate features and deep representations. Moreover, our network employs the Mish and ReLU activation functions (AFs), which utilize AdamW and Adam optimization strategies to achieve enhanced segmentation performance. The proposed MAGRes-UNet method was compared with the U-Net, Multi-Attention Gated-UNet (MAG-UNet), and Residual-UNet (ResUNet) models. In addition, a statistical *T-test* was performed to assess the difference in model significance between the approaches. The analysis revealed that MAGRes-UNet employing Mish and AdamW provides significant performance improvement over the ReLU AF and Adam optimizer on two benchmark datasets: Multi-Class BT T1-weighted Contrast-Enhanced Magnetic Resonance Imaging (T1-CE-MRI) and skin lesions HAM10000 (Human Against Machine with 10,000 training images). MAGRes-UNet using Mish and AdamW provides competitive performance over the representative medical image segmentation methods.

INDEX TERMS Attention mechanism, brain tumor segmentation, Mish activation function, residual block, U-Net.

I. INTRODUCTION

In recent years, the rapid development of medical imaging modalities, such as magnetic resonance imaging (MRI), computed tomography, ultrasound, and X-rays, has led to the generation of large image datasets, which have improved

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno M. Garcia.

the usability and accessibility of the imaging modalities in clinical scenarios [1]. Medical image analysis also contributes to quicker diagnosis and treatment procedures. MRI provides excellent contrast between soft and hard tissues for the assessment of brain tumors (BTs) without exposing the patient to radiation. According to the International Association of Cancer Registry (IARC) [2], more than 28,000 BT cases are annually reported in India. The precise surgical diagnosis and planning of BT are intricate processes that improve the survival rate of patients with BT [3]. The appropriate segmentation of BTs is essential for accurate surgical localization and diagnosis [4]. Because the BT regions vary in size, shape, position, and light dispersion, their accurate segmentation remains challenging [5]. MRI offers precise information about the tissue structure of BTs and helps in accurate segmentation in BT tasks [4]. To locate the BT regions, radiologists examine different T1-weighted contrast-enhanced (T1-W CE), T1-W, T2-W fluid-attenuated inversion recovery (FLAIR), and T2-W sequences simultaneously [6]. However, manual segmentation is time-consuming and highly dependent on the competence level of the radiologist [3]. Hence, automation is essential to make BT segmentation more reliable and efficient.

Recently, deep learning (DL) algorithms have demonstrated excellent performance in image segmentation [7], [8], [9], [10], [11], [12]. Convolutional neural networks (CNNs) are valuable for medical image analysis because they can automatically extract features at low to high levels [13]. In particular, deep convolutional neural networks (DCNNs) are increasingly being used in medical image segmentation. Furthermore, fully convolutional neural networks (FCNNs) [7], [14] and U-Net-like encoder-decoder architectures [8], [9], [15], [16], [17] have shown superior performance when compared with traditional approaches [18], [19], [20]. In these networks, the convolutional layers are combined with the encoder's downsampling layers and the decoder's upsampling layers to extract deep features and fuse high-resolution features to achieve accurate pixel-level semantic predictions. They are particularly suitable for medical image segmentation because they can extract features and learn complex patterns from medical images.

Walsh et al. [21] developed a U-Net-based lightweight segmentation method for BT. The application was tested on the BITE dataset with promising results. In their study, the intersection over union (IoU) achieved a success rate of 89%. According to Ronneberger et al. [8], a U-shaped segmentation successively produces convolutional operations. Their model integrates and reconstructs the encoder stage features in the decoder stage for segmentation. However, this approach may result in semantic gaps. Skip connections have been proposed to eliminate these gaps [1], [22]. The authors of [23] presented UNet++, a convolution-based segmentation model that incorporates dense skip connections for segmenting medical images. The skip connections in the proposed model outperformed those in the U-Net model.

However, it had too many parameters when compared with the U-Net model. Consequently, more straightforward skip connections were favored in the ResUNet+ architecture than the UNet++ model. Cao et al. [24] proposed the DenseUNet model to improve the performance of the U-Net segmentation model and solve the vanishing gradient issue. Instead of convolutional layers, dense blocks were employed in the U-Net-based model. The testing of their model on the ISBI 2012 EM dataset was performed without a pre-training phase or post-processing module. To enhance the feature map discrimination, the ResUNet+ model incorporates residual blocks and convolution layers. Li et al. [25] incorporated pretrained block structures in the segmentation model to add a node structure to the encoder and decoder, which provided superior features over U-shaped segmentation models.

Qamar et al. [26] used inception module components in a novel U-Net model for segmenting brain tissues. Zhang et al. [27] proposed a U-Net residual attention network (AresU-Net), and simultaneously deployed the attention technique and residual blocks to the U-Net model for the 2D BT segmentation network. Huang et al. focused on improving the features and gathered multiscale features for analysis, thereby reducing the network parameters and increasing the accuracy [28]. Zhang et al. [29] proposed a U-Net based on a swin transformer and used the spatial and channel information of the image to achieve more accurate semantic segmentation. Wang et al. [30] proposed repeating loop unit structures by fusing the encoder and decoder structures used in the U-Net architecture. The model developed by Zhuang [31] incorporates several pairs of encoder and decoder branches to capture intricate details.

Several approaches have been used to segment skin lesions according to the preference and experience of the researcher. The initial studies used histograms and thresholding methods [32], [33], [34], [35], [36]. The authors of [34] presented a new technique for skin lesion segmentation based on type-2 fuzzy logic. Compared with other studies, this method successfully detected uncertainties at the edges of the lesions. Çelebi et al. [35] processed a wide range of dermoscopic images using ensembles of thresholding techniques. They applied multiscale thresholding fusion to analyze the skin lesion images. However, the approach had a limited degree of effectiveness with a single threshold coefficient. Peruch et al. [36] used an impressive thresholding technique and proposed a stepwise process that resembled a dermatologist's actions. Other alternative methods include techniques such as preprocessing, dimensionality reduction, blurring, thresholding, and post-processing. Al-Masni et al. [37] presented a deep-resolution convolutional network for segmenting skin lesions. Unlike U-Net [8], this approach does not use upsampling or downsampling techniques to ensure precise resolution across feature maps. The RGB color images are immediately provided to the network using DL techniques, which require minimal preprocessing.

Recent studies have shown that increasing the number of input color channels improves the skin lesion segmentation. Three RGB channels, three HSV channels, and one L channel of the CIELAB color space were combined by Yuan and Lo [38] to create seven channels for their DNN model, resulting in improved segmentation. The authors of [39] also employed 7 channels in the DL method but adopted different strategies. The initial layer included three RGB channels, whereas the deeper intermediate layer contained four more channels (S of the HSV color space, I of the YIQ color space, B of the CBR color space, and Z of the XYZ color space). Xie et al. [40] developed a high-resolution feature block containing three branches: normal convolutional, spatial, and channel attention. ASCU-Net was proposed by Tong et al. [41], which employed three attention techniques: an attention gate [17], spatial attention module, and channel attention module. The segmentation of skin lesions was improved using a transfer learning approach. Kadry et al. [42] and Rajinikanth et al. [43] used a pretrained VGG network [44] to encode the key features of a skin lesion image, which were then resampled to generate a segmentation mask.

Many studies have addressed the flaws of the standard U-Net models [8]. However, the structure of convolutional operations can lead to the under- or over-segmentation of medical images. Therefore, the realization of an automated segmentation model remains challenging because of the following factors: (1) the need for labeled medical data can distort the assessment of model performance and cause potential overfitting when dealing with limited datasets. (2) To properly handle medical datasets that involve image processing tasks, it is essential to have background knowledge to guide the learning process of the model. (3) Medical images present difficulties owing to their unclear boundaries, movement distortion, varying intensities, and inconsistencies in the size and shape of the critical area. (4) Models struggle to perform well on new datasets because of differences in the equipment used for the same medical procedure and variations in the image-acquisition protocols. Considering these constraints, automated segmentation has become a popular research topic.

With the advancement of CNNs [45], several networks [7], [46] utilizing encoder-decoder architectures have been used for semantic segmentation. The encoder classifies and analyzes the local pixels to extract high-level semantic information from the low-level local pixels. The decoder simultaneously maps high-level semantic information onto specific pixels [47]. To be effective, these networks require large datasets with numerous samples, which is difficult in the medical field [48].

To solve these problems, the U-Net [8] model was proposed, which provides good medical image segmentation performance by using skip connections between the encoding and decoding stages. This mitigates the downsampling-induced information loss [49]. However,

semantic gaps may exist between the encoder and decoder levels [15], leading to invalid fusion and misclassification of the decoder results, which may affect the segmentation process [47]. Several novel U-Net networks, including UNet++ [23], U-Net3+ [28], and MultiResUNet, have been proposed to reduce the effect of semantic gaps and improve the segmentation performance [15]. U-Net++ and U-Net3+ utilize nested, dense, and full-scale skip connections, whereas MultiResUNet replaces the skip connections with an improved Res path.

However, these methods have limited feature extraction capabilities, and some information is diluted when certain layers cannot extract certain features. In addition, a skip connection introduces irrelevant clutter (e.g., invalid features caused by poor extraction capabilities in the encoding stage), as shown in Fig. 1(c-e). Some skip connections may negatively affect the segmentation performance because of their ineffectiveness [47].

Several methods have recently been proposed to solve these problems. The Context Pyramid Fusion Network (CPFNet) [50] uses multiple global pyramid guidance modules to provide information from various stages of encoding, and atrous convolutions are used to broaden the receptive field [51]. These modules provide information from different levels to the decoding stage and suppress low-level background noise and features. In contrast, CPFNet decoders obtain limited encoder information at low levels, preventing the effective use of the encoder's multiscale information. Moreover, CPFNet skips connections, which may result in data loss during downsampling [52]. A channel transformer module was designed using a transformer module in UCTransNet [47]. Integrating the encoder-level information into the decoders enabled the decoder features to efficiently carry multiscale information, thereby reducing ambiguity. However, the transformer requires a large amount of training data and performs poorly on datasets with fewer samples.

Our study proposes MAGRes-UNet (Multi-Attention Gated Residual U-Net) to solve the above problems by integrating two modules into an existing U-shaped framework. First, we employed MAG modules to enable seamless information flow between the encoding and decoding stages. These MAG modules target the reduction of semantic gaps and ensure efficient transfer of pertinent information across the network architecture, ultimately improving the segmentation performance. Second, the proposed network incorporates residual blocks to elevate the feature extraction capabilities and extend the potential dilution of information in specific network layers to multiple scales. This design enhancement is crucial for maintaining feature quality throughout the network, which contributes to more precise and accurate segmentation results. Our approach effectively mitigates issues with multiscale clutter and suboptimal skip connections, which are often observed in conventional U-Net models. Furthermore, integrating the

Mish AF and AdamW optimizer further enhanced the model's performance by improving the nonlinearity and optimizing the weight updates. MAGRes-UNet presents a promising solution for enhancing the segmentation accuracy even with limited data. This study makes the following contributions:

1) A U-shaped framework called MAGRes-UNet, shown in Fig. 2, which leverages two modules—multi-attention gates (MAGs) and residual blocks module—to improve the medical image segmentation performance is proposed.

2) The standard U-Net [8] model in our study incorporates four MAG modules to use high-level decoding features, thus facilitating the enhancement of lower-level encoding features. This improves the concatenation between the encoder and decoder when processing small-scale tumors. The inclusion of residual blocks deepens the network through identity connections, optimizing the training process. These connections enable the transmission of fine-grained details and reduce issues related to vanishing gradients, resulting in more accurate and robust segmentation of medical images between different classes.

3) This study used the Mish [53] activation function (AF) and AdamW [54] optimizer. Our MAGRes-UNet model demonstrated remarkable improvement in training and segmentation accuracy, as validated by the experimental results. Comparative analysis, including statistical tests *T-Test*, showcased the consistent superiority of our approach over MAG-UNet, ResUNet, and U-Net across various performance metrics. This superiority was consistently observed on two distinct medical imaging datasets, BT T1-weighted CE-MRI and skin cancer HAM10000, when compared with state-of-the-art (SOTA) DL models.

The remainder of this paper is organized as follows: Section II examines the proposed methods. Section III explains and analyzes the experiments, including data preprocessing and metric evaluation. Section IV presents results and discussion along with a performance comparison with the state-of-the-art (SOTA) models. Finally, section V presents the main conclusions of this study.

II. METHODS
This study proposes a MAGRes-UNet framework to precisely segment multiclass brain MRI tumors and skin lesions. Our architecture is based on U-Net because of its successful application in biomedical image segmentation. U-Net is an FCNN and is widely known for its effectiveness in semantic segmentation tasks. By incorporating the MAG and residual block modules, we extend the capabilities of U-Net to address the challenges inherent in medical image segmentation. U-Net and MAGRes-UNet are presented in detail, and their distinct features are outlined.

A. U-Net MODEL
According to Ronneberger et al. [8], the U-Net is a DL CNN architecture developed for biomedical image segmentation. U-Net has numerous applications, including consumer videos [55], [56], earth observations [57], and medical imaging [58], [59], [60]. This network consists of an encoder followed by a decoder with a bottleneck layer between them. Our architecture uses convolutional blocks as building blocks for the encoders and decoders. As part of this block, we performed 2D convolution, batch normalization (BN), dropouts, and Mish (discussed in Section II-B3) as the AF, and max-pooling for downsampling with a stride size of 2. Mish [53] is a newly proposed AF that has demonstrated superior performance over rectified linear units (ReLU) [61], leaky ReLU [62], exponential linear units (ELUs) [63], and Swish [64] in various DL [65] methods. The number of feature channels is doubled for each downsampling step (64, 128, 256, 512). The encoder stage is followed by a bottleneck layer containing 1024 feature channels.

The decoder network is based on upconvolution, followed by coupled convolution, BN, dropout, and Mish AF, and decreases the number of feature channels by half (512, 256, 128, 64). A skip connection is utilized to convey the spatial information from the encoder to the decoder to recover lost information in the encoder blocks without introducing new parameters or processing costs; this improves the processing speed and accuracy [66]. To transfer the features of the preceding layer to the desired number of classes, a 1×1 convolutional layer is adopted at the end of the network, followed by the softmax [67] function. U-Net has been extensively used for semantic segmentation, including biomedical image segmentation. Fig. 1(a) shows the visual representation of the standard U-Net architecture.

B. PROPOSED MULTI-ATTENTION GATED RESIDUAL-UNet (MAGRes-UNet) MODEL
The U-Net model is a popular choice for medical image segmentation; however, it has some limitations, such as a vanishing gradient with increasing network depth. As the network deepens, the gradient becomes almost zero in the lower layers [68]. Furthermore, when generating segmentation masks using the U-Net framework, the low- and high-level features that contain information on the boundary, edges, and location of the tumor region are equally important. Consequently, essential features may become less influential, whereas unimportant ones may become more valuable. This undermines the success of the segmentation. To improve the medical image segmentation performance metrics, we must address the gradient vanishing problem associated with the traditional U-Net design. To access more distinctive features, it is essential to utilize downsampling rather than relying solely on skip connections between the encoder and decoder. However, this approach may lead to a loss of local details from previous layers owing to convolutional operations.
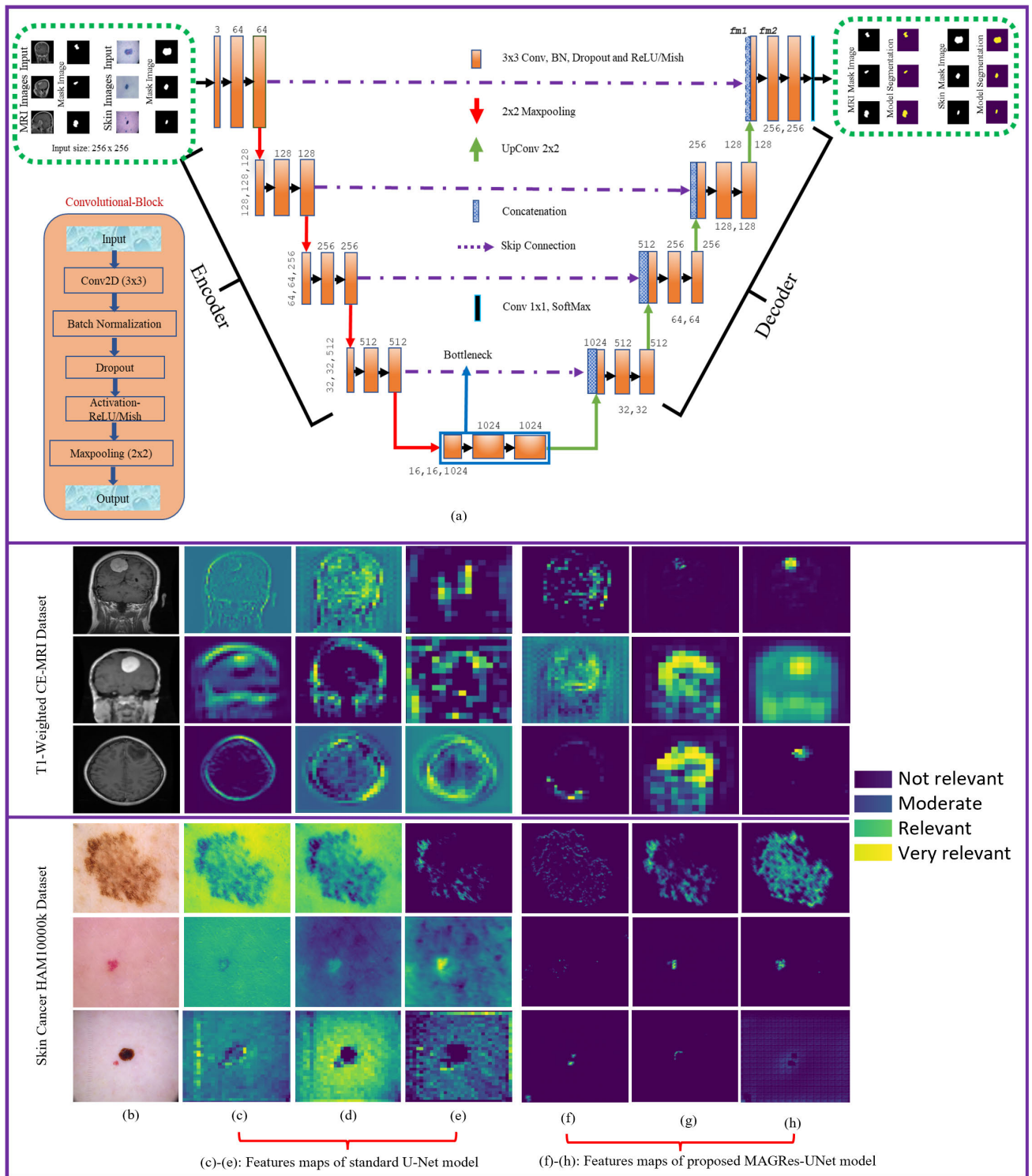
FIGURE 1. Comprehensive overview of clutter issues in the standard U-Net model, which are passed to the decoder via skip connections. (a) The standard U-Net model serves as the baseline architecture. (b) Original input medical images. (c), (d) Feature maps (fm1 and fm2) of the standard U-Net model in (a); the feature map characteristics are visualized without incorporating attention gates and residual blocks. (e) Results of combining the feature maps of (c) and (d) using a basic skip connection. These simple skip-connected features limit the ability to capture intricate patterns and salient features. (f), (g) Feature maps of the proposed MAGRes-UNet model shown in Fig. 2, which leverages MAGs and residual block modules. (h) Resulting features after concatenating (f) and (g) when utilizing MAG instead of a skip connection. The utilization of MAG and residual blocks enhances the model's ability to capture detailed and pertinent information within the region of interest while suppressing irrelevant information, thereby improving the feature quality and segmentation accuracy. By contrast, skip connections can introduce unwanted clutter, complicating the segmentation performance.
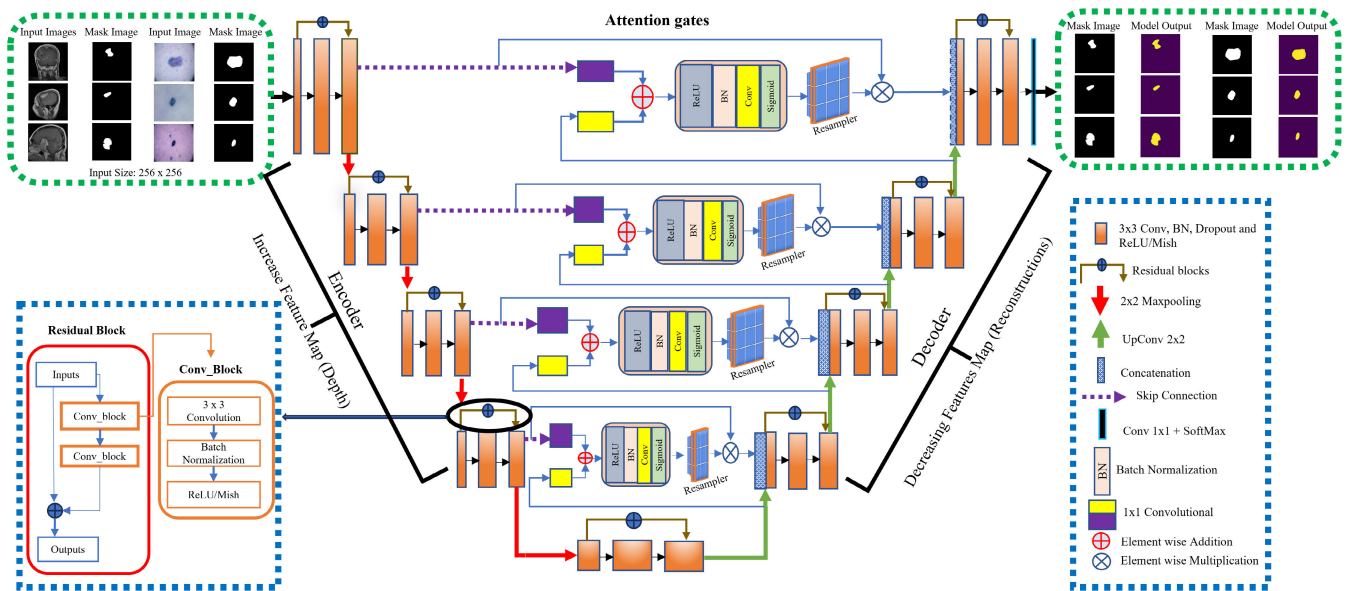
FIGURE 2. Illustration of the end-to-end proposed MAGRes-UNet model with two encoding-decoding modules. The MAGRes-UNet model combines residual blocks and attention gates using a single U-Net architecture. A series of MAG units are added to the skip connection to highlight important feature information while disambiguating undesirable and noisy feature responses. MAGRes-UNet not only extracts abundant semantic information to enhance feature learning, but also focuses on small-scale tumors, potentially resulting in improved performance in medical image segmentation tasks.
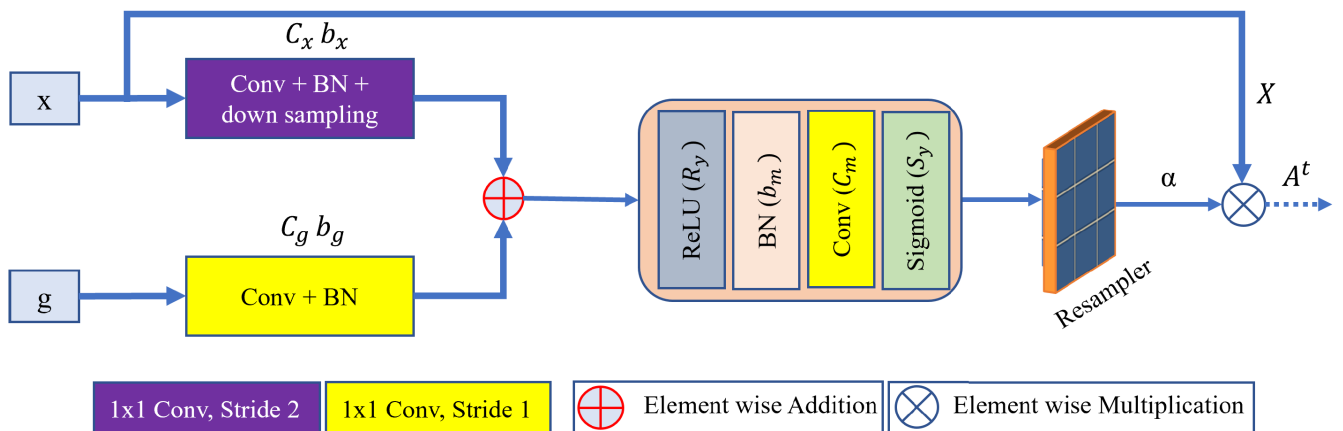

FIGURE 3. Illustration of MAG architecture. A feature map with features from various stages, *x* and *g*, serves as an input to the MAG. The result of the MAG is given by $A^t = X \cdot \alpha$, where α is the attention coefficient, and · denotes element-wise multiplication. The MAG change map was closer to the ground truths (GTs).

To address these challenges, the MAGRes-UNet model was introduced as a solution. This model incorporates residual blocks to address the gradient vanishing problem, which helps retain low-level features. Consequently, residual blocks are favored in the encoder structure of the proposed model. Moreover, the classical U-Net structure suffers from semantic gaps between the encoder and decoder. To remedy this, in contrast to other U-Net variations, new connection nodes, known as MAG nodes, are introduced between the encoder and decoder, which enhance the segmentation performance for both the BT and skin cancer datasets. Fig. 2 offers a comprehensive view of the proposed MAGRes-UNet model. We conducted an

in-depth investigation of the intricacies of the MAGRes-UNet model.

As shown in Fig. 2, MAGRes-UNet has an encoder-decoder architecture consisting of contracting (left encoder) and expanding (right decoder) paths. The network has the input dimensions of $256 \times 256 \times 3$, where each image measures 256×256, and the input comprises 3 channels. The standard U-Net comprises four plain blocks. The contracting path replaces them with four residual blocks. Each residual block has two convolutional units. The initial number of layers within the contracting path is $256 \times 256 \times 64$. The individual convolutional unit encompasses a BN and dropout layer and employs the Mish AF (discussed in Section II-B3).

Mish is a newly proposed AF with superior performance over the widely used ReLU employed in the original U-Net architecture [8]. Our model employs a convolutional layer with a 3×3 filter along with the Mish AF. For downsampling, we selected a 3×3 convolutional kernel with a stride of 2. The number of channels for the feature maps was doubled during this process; however, the map size was halved. The fourth residual block in the contracting path acts as a bridge connecting the two paths and has dimensions of $16 \times 16 \times 1024$. The expanding path consists of four MAG units and four residual blocks, differentiating it from the current segmentation models. This arrangement eliminates the vanishing gradient issue to optimize the training process, disambiguates undesirable and noisy feature outcomes, and improves the prominence of important feature information. Each layer in the residual block produces a gating signal vector along the expansion path. The feature maps at the bottom level serve as the initial gating signal for the expanding path. The gating signal and feature map are combined with an attention gate unit to enhance the segmentation performance for small-scale tumors with more accurate location information. The feature map is upsampled using bilinear interpolation before the individual residual blocks, after which the output of the attention gate and the 3×3 convolution are concatenated. The image size is doubled for each upsampling operation; however, the number of feature channels is reduced by half. At the end of the expanding path, 1×1 convolution and softmax AF are used to map the multichannel feature maps to the desired number of classes.

1) MULTI-ATTENTION GATES

The attention mechanisms were proposed by Bahdanau et al. [69] and Luong et al. [70]. They were subsequently applied to the U-Net model by Oktay textitet al. to segment the pancreas [17]. To the best of our knowledge, MAGRes-UNet is the first model to segment multiclass (glioma, meningioma, and pituitary) brain MRI tumor and skin cancer datasets. The main goal of the MAG is to reduce irrelevant feature responses in the feature maps and effectively focus on the intended area while learning significant features. Recent research has shown that training the DL models using MAG can improve the network performance [71], [72]. Fig. 3 illustrates the architecture of the MAG units used in this study.

The MAG module in Fig. 3 accepts two inputs and produces one output. The input feature x belongs to the encoder part, and the gating feature g belongs to the lower decoder part. The gating feature g allows more useful features to be extracted from the encoded feature x, whereas invalid features are ignored. Furthermore, the input feature x is downscaled in steps of 2 to be compatible with the gating feature g. The inputs x and g are combined pixel-by-pixel using 1×1 convolutional operations ($C_{x,g}$) and BN ($b_{x,g}$). A ReLU ($R_y = \max(0, y)$) AF is applied to activate the
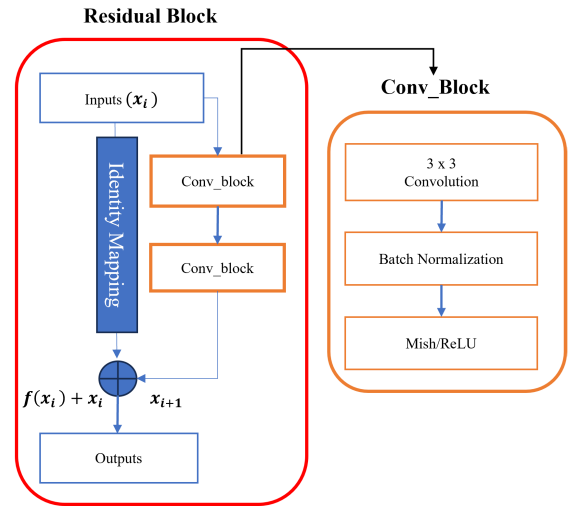

FIGURE 4. Residual blocks. The symbol (+) indicates the addition of elements.

results obtained in the previous step. Subsequently, both x and g are processed through 1×1 convolutional (C_m) and BN (b_m) operations to obtain the weights. The collective weights are passed through a sigmoid AF ($S_y = \frac{1}{1+e^{-y}}$), which converts the input values to a range between 0 and 1. Then, the collective weight signal is multiplied element-wise and rescaled to the same scale as the input signal X. The following formulae can be used to describe the feature selection procedure of MAG.

$$\beta = R_x \left[(C_g g + b_g) + (C_x x + b_x) \right] \qquad (1)$$

$$A = S_y(C_m \beta + b_m) \qquad (2)$$

$$A^t = X * \alpha \qquad (3)$$

R_x represents the ReLU AF and S_y represents the sigmoid AF. C_g, C_x, and C_m are the linear transformations calculated using $1 \times 1 \times 1$ convolution. A^t represents the MAG output, α represents the attention coefficients, and $*$ represents the element-wise multiplication of the original vector X.

2) RESIDUAL BLOCKS

Because there are currently only a few layers in the standard U-Net architecture, we deepened it considerably to improve its performance [73]. The U-Net model uses encoder and decoder architectures with skip connections to acquire high and low-level features. However, as the network deepens, a vanishing gradient occurs, which implies that the gradients diminish during backpropagation, resulting in ineffective learning [74]. To solve this problem, Hu et al. [75] introduced residual blocks to facilitate the gradient flow and ease the training process. Consequently, we implemented a residual block in the standard U-Net model, as shown in Fig. 4. In Fig. 4, each block has two 3×3 convolutional layers, followed by BN and Mish AF. The input to the residual block passes through a 1×1 convolutional layer, which undergoes BN and is then summed element-wise with the

output. The network can pass information more effectively owing to the residual blocks. It can use and access low-level feature details for accurate predictions because the original input data are preserved through a shortcut connection. Consequently, the features can be integrated better across different network layers, thereby enhancing the overall performance. According to [75], each residual block is defined as follows.

$$y_i = h(x_i) + F(x_i, W_i) \tag{4}$$

$$x_{i+1} = f(y_i) \tag{5}$$

The inputs and outputs of the ith residual unit are x_i and x_{i+1}, where $F(\cdot)$ is the residual function, $f(y_i)$ is the activation function, and $h(x_i)$ is the identity-mapping function. W_i is the weight vector of the feature map in the i-th residual unit.

3) MISH ACTIVATION FUNCTION

Neural networks incorporate nonlinearity through the AF. Therefore, it is essential to develop and test deep neural networks. The widely used AFs are the ReLU [61], leaky ReLU (L-ReLU) [62], and the recently introduced Swish [64]. Our method uses the most up-to-date AF Mish [53], outperforming ReLU and Swish on challenging datasets. In addition, Mish is easy to implement in neural networks because of its simplicity [53]. Mish is a smooth, non-monotonic AF for a neural network, written as

$$f(x) = \tanh(\omega(x)) \tag{6}$$

where $\omega(x)$ is a softplus activation function, given by $\ln(1 + e^x)$. Mish is an activation function that implements the self-gating property. This function replaces other point-wise activation functions, such as ReLU. By using the self-gating property, network parameters can be modified without providing any input to the gate. The Mish function in TensorFlow is defined as x multiplied by the hyperbolic tangent of the softplus function x ($x \cdot \tanh(\text{softplus}(x))$). Mish must be non-monotonic, smooth, and unbounded above and below to achieve optimal neural network results. The Mish activation plot is shown in Fig. 5.

III. EXPERIMENTS

This section introduces the dataset and preprocessing used to evaluate the model. Next, we discuss the implementation details and evaluation process followed in the study. The selected hyperparameters are listed in Table 1.

A. DATASET AND PREPROCESSING

The proposed model was evaluated on two publicly available medical image datasets. The first dataset, T1-weighted CE-MRI [76], was acquired from Nanfang Hospital in Guangzhou, China, and the General Hospital at Tianjing Medical University, China, from 2005 to 2010. The data pertained to 3064 slices collected from 233 patients (708 meningiomas, 1426 gliomas, and 930 pituitary tumors). The in-plane resolution of the images was 512×512. Each
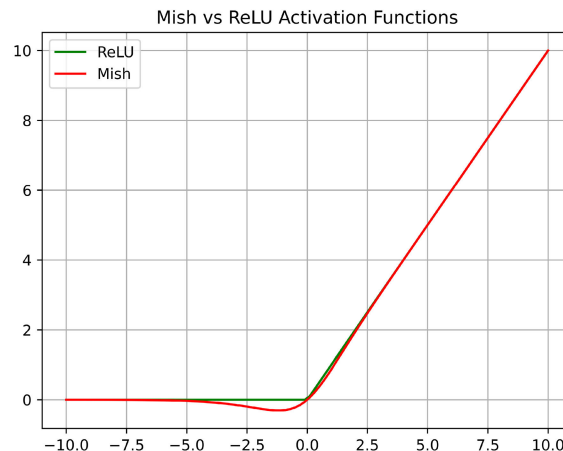

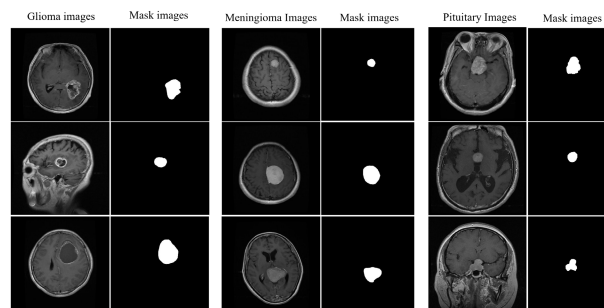**FIGURE 5.** Mish activation function.

FIGURE 6. Multi-class BT T1-weighted CE-MRI sample dataset.

tumor class was manually annotated by an expert radiologist. Fig. 6 shows the original tumor class and corresponding masked-image samples. Table 2 lists the details of the BT CE-MRI image dataset.

The second dataset comprised 10015 skin cancer HAM10000 dermoscopic images [77] of 450×600 pixels, with masked images representing seven types of lesions: actinic keratoses, basal cell carcinomas, benign keratoses, dermatofibroma, melanoma, nevi, and vascular lesions. The images were collected from different population groups by using different dermoscopic devices, resulting in diverse and representative datasets. Fig. 7 shows skin cancer samples from the HAM10000 dataset.

Medical images contain complex biological tissue structures and may have low imaging quality, which makes their segmentation challenging. Because of limited processing resources, both dataset images were downscaled to 256×256. Pixel intensities differ based on the manufacturer, acquisition conditions, and sequences; therefore, the input scans must be normalized. Furthermore, we employed Z-score normalization on the 2D images to reduce device noise, thus significantly increasing the preprocessing and decreasing overfitting. Each image is processed by calculating the Z-score based on its mean value and standard deviation, and
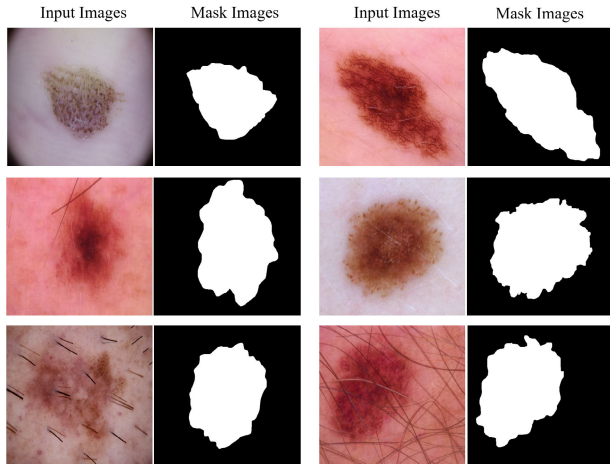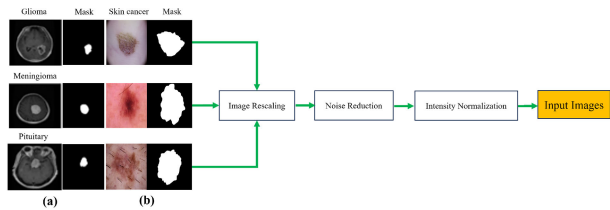
FIGURE 7. Skin cancer HAM10000 sample dataset.

FIGURE 8. Data preprocessing steps: (a) BT T1-weighted CE-MRI modalities (b) Skin cancer HAM10000 modalities.

TABLE 1. Hyperparameter values for the proposed model.

Parameters	Values
Optimizer	Adam/AdamW
Learning rate	0.0001
Epochs	40
IoU smoothing factor	0.00001
Dropout	0.25
Activation function	ReLU/Mish
Loss function	Categorical cross-entropy
Batch size	32
Dice score smoothing factor	0.00001
Patience	10
Monitor	Validation loss
Metrics	Accuracy, Loss, IoU, Dice score, Sensitivity, Specificity, and Precision

it can be calculated as

$$z' = \frac{z - \mu}{\delta} \qquad (7)$$

where z and z' represent the input and normalized images, respectively. The mean and standard deviations of the input images were μ and δ, respectively. A similar method was used for all the annotated masked images in the training set. Deep models require data preprocessing to improve the segmentation accuracy, even though they are primarily noise-

TABLE 2. Details of BT CE-MRI image dataset.

Tumor Class	Original MRI Images	Masked Images
Glioma Tumor	1426	1426
Meningioma Tumor	708	708
Pituitary Tumor	930	930
Total no images	**6128**	

resistant. Fig. 8 shows the data preprocessing used in this study to effectively train the DL model and mitigate the class imbalance, leading to a significant improvement in segmentation performance. The allocation strategy for both datasets involved dedicating 80% of the data for training, 10% for validation, and another 10% for testing purposes.

B. METRICS EVALUATION

The models were evaluated using several criteria to determine their efficiency. Our model was assessed quantitatively based on the accuracy, IoU, Dice score, specificity, sensitivity, and precision. For each evaluation criterion, higher scores indicated better segmentation. These metrics can be mathematically defined as follows.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \qquad (8)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \qquad (9)$$

$$\text{Dice-score} = \frac{2TP}{2TP + FN + FP} \qquad (10)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad (11)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad (12)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (13)$$

True Positive (TP) in the equations indicates that the actual tumor matches the suspected tumor. A *True Negative (TN)* indicates that the predicted and actual non-tumor areas match. *False Positive (FP)* indicates that the predicted tumor area is not the actual tumor. In contrast, *False Negative (FN)* indicates that the predicted non-tumor area is the actual tumor area.

IV. RESULTS AND DISCUSSION

The proposed method was evaluated using two datasets to assess its effectiveness and robustness. This will be discussed in the following subsections. A comprehensive performance evaluation is presented in the first subsection based on two datasets: BT CE-MRI and skin cancer HAM10000 dataset. In the second subsection, we explore the ablation study of the proposed method, where we discuss the effects of the critical components, namely, the residual blocks, MAG, Mish, ReLU, Adam, and AdamW, on the segmentation performance. In the third subsection, we compare our performance results with

TABLE 3. Proposed model performance on BT CE-MRI dataset.

Model	AF	Optimizer	Accuracy (%)	IoU (%)	Dice score (%)	Loss (%)
U-Net	ReLU	Adam	92.31	92.47	92.80	0.0183
	Mish	AdamW	92.60	93.91	92.79	0.0151
Residual-UNet	ReLU	Adam	92.53	93.25	92.15	0.0146
	Mish	AdamW	93.72	93.19	93.72	0.0143
MAG-UNet	ReLU	Adam	96.38	96.29	95.27	0.0210
	Mish	AdamW	98.20	97.63	96.41	0.0133
MAGRes-Net	ReLU	Adam	98.46	98.13	96.92	0.0113
	Mish	**AdamW**	**99.94**	**98.29**	**97.75**	**0.0101**

TABLE 4. Proposed model performance on skin cancer HAM10000 dataset.

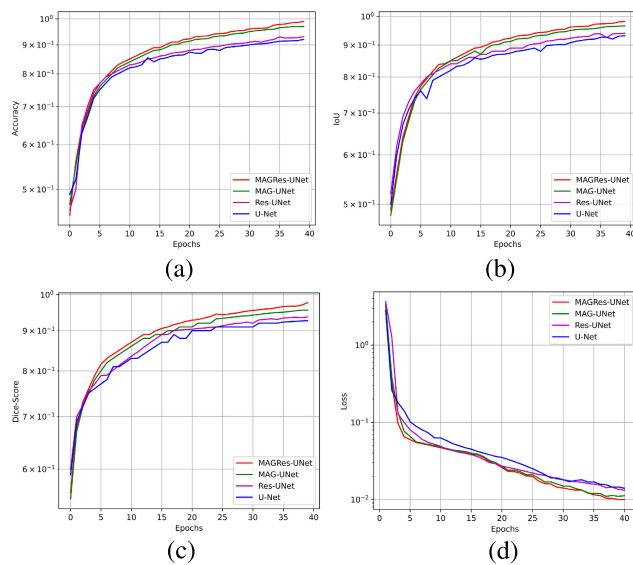| Model	AF	Optimizer	Accuracy (%)	IoU (%)	Dice score (%)	Loss (%)
U-Net	ReLU	Adam	92.03	91.68	89.17	0.0231
	Mish	AdamW	92.21	92.30	93.20	0.0142
Residual-UNet	ReLU	Adam	93.80	93.25	90.59	0.0130
	Mish	AdamW	94.31	93.91	93.62	0.0132
MAG-UNet	ReLU	Adam	95.36	96.50	96.04	0.0141
	Mish	AdamW	96.94	96.19	95.73	0.0112
MAGRes-Net	ReLU	Adam	98.49	96.46	96.58	0.0115
	Mish	**AdamW**	**99.71**	**97.83**	**97.36**	**0.0102**

FIGURE 9. Validation performance of the proposed model using the Mish AF and AdamW optimizer, assessed on the BT T1-weighted CE-MRI dataset: (a) Accuracy, (b) IoU, (c) Dice score, and (d) Loss.

FIGURE 10. Validation performance of the proposed model using the Mish AF and AdamW optimizer, evaluated on the skin cancer HAM10000 dataset: (a) Accuracy, (b) IoU, (c) Dice score, and (d) Loss.

those of the SOTA segmentation approaches by using the BT and skin cancer datasets.

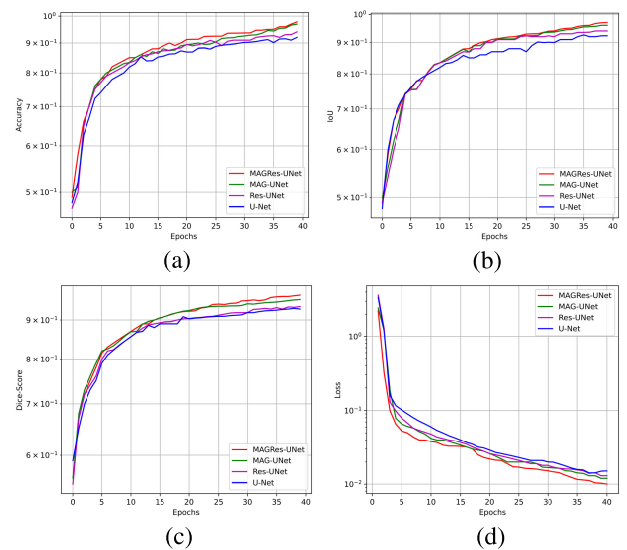We conducted experiments using Keras with a TensorFlow backend. A mechanism for early termination was employed in the validation set to prevent overfitting. We evaluated the results based on various metrics, including accuracy, Dice score, IoU, sensitivity, specificity, and precision. Furthermore, we conducted a statistical test to assess whether the proposed models exhibited significant differences in

performance. Table 5 lists the statistical performance of the proposed model. This analysis uses various evaluation metrics to compare the efficiency of Mish AF with the AdamW optimizer and ReLU AF with the Adam optimizer. The proposed MAGRes-UNet consistently outperformed the other architectures, including MAG-UNet, ResUNet, and U-Net, on both medical datasets, as shown in Table 5. This superiority is statistically significant, as shown in Fig. 11, Fig. 12, and Table 5, when using the Mish AF and AdamW optimizer.

A. OVERALL PERFORMANCE ANALYSIS

This study aimed to evaluate four distinct models (MAGRes-UNet, MAG-UNet, ResUNet, and U-Net) on challenging datasets: T1-weighted CE-MRI and skin cancer HAM10000. The proposed model's architecture draws inspiration from the U-Net [8] model, known for its success in medical image segmentation.

However, the U-Net model's continuous downsampling process reduces the segmentation accuracy for small-scale tumor images. To address this limitation and achieve better segmentation performance, we explored the integration of a MAG and a residual block into the standard U-Net architecture. The MAG highlights the salient features while suppressing irrelevant noise, and the residual block aids in training and mitigates the vanishing gradient problem. Additionally, we utilized Mish and ReLU AF along with the AdamW and Adam optimization strategies to improve the segmentation accuracy.

The experimental results in Tables 3 and 4 demonstrate that MAGRes-UNet with Mish and AdamW consistently outperformed the other architectures (MAG-UNet, ResUNet, and U-Net) on both datasets, demonstrating its exceptional capabilities. MAGRes-UNet exhibited remarkable accuracy, IoU, and Dice coefficient and significantly lower loss values than those of the other models, as shown in Fig. 9(a)-(d) and Fig. 10(a)-(d). Incorporating the MAG, residual modules, Mish AF, and AdamW optimizer significantly improved all the evaluation metrics. Specifically, on the T1-weighted CE-MRI dataset, MAGRes-UNet achieved an impressive accuracy of 99.94%, IoU of 98.29%, Dice score of 97.75%, and a remarkably low loss of 0.0101%. Similarly, on the skin cancer HAM10000 dataset, the model achieved an accuracy of 99.71%, IoU of 97.83%, Dice score of 97.36%, and a low loss of 0.0102%. Incorporating MAG and residual mechanisms using Mish and AdamW resulted in a smoother, faster, and more stable performance, reducing fluctuations in the evaluation metrics.

A comprehensive evaluation was conducted using boxplots of the testing dataset to assess the robustness of the models. The boxplots in Fig. 11(a)-(d) and Fig. 12(a)-(d) show the distribution and variability of the evaluation metrics across test instances for each model employing Mish AF with AdamW optimizer as well as ReLU AF with the standard Adam optimizer. The results demonstrate that MAGRes-UNet consistently surpasses the other models

(U-Net, ResUNet, and MAG-UNet) in performance across all evaluation metrics, including accuracy, IoU, Dice score, specificity, sensitivity, and precision, for both datasets. We conducted a statistical *T-test* to confirm the results, as presented in Table 5. By using the SciPy library in Python, the *T-test* enables a comparison of the following performance metrics: accuracy, IoU, Dice score, specificity, sensitivity, and precision for the MAGRes-UNet, MAG-UNet, ResUNet, and U-Net models employing both Mish with AdamW and ReLU AF with the Adam optimizer. This analysis aimed to determine the importance of the AF and optimizer pairings in the suggested models. Accordingly, we postulated the following hypotheses:

H_0 (n.s.s*): The proposed models utilizing the (Mish + AdamW) and (ReLU + Adam) pair do not significantly differ from each other.

H_1 (s.s**): The proposed models utilizing the (Mish + AdamW) and (ReLU + Adam) pair vary significantly from each other.

Table 5 presents the findings of this analysis. According to the null hypothesis H_0, there is no statistically significant (n.s.s*) difference between utilizing (Mish + AdamW) and (ReLU + Adam), whereas the alternative hypothesis H_1 suggests otherwise at a significance level of 5% ($\alpha = 0.05$).

Table 5 lists the *P-values* of the proposed models for accuracy, IoU, Dice score, specificity, sensitivity, and precision. The table shows that s.s** indicates statistical significance, whereas n.s.s* denotes no statistically significant difference. Notably, the proposed MAGRes-UNet model employing Mish AF with AdamW optimizer demonstrated statistical significance across all metrics for both medical datasets, as indicated by the corresponding *P-values* in Table 5. For instance, for the BT CE-MRI dataset, the *P-values* for the accuracy, IoU, Dice score, specificity, sensitivity, and precision of the MAGRes-UNet model were 0.025, 0.028, 0.023, 0.005, 0.011, and 0.042, respectively. Similarly, for the HAM10000 skin cancer dataset, the *P-values* were 0.003, 0.015, 0.043, 0.021, 0.174, and 0.000. *P-values* below 0.05 indicated significance.

There was a statistically significant difference between the proposed models that used the Mish + AdamW and ReLU + Adam approaches, as revealed by the 95% confidence level test of significance applied to the accuracy, IoU, Dice score, specificity, sensitivity, and precision. This proved the alternative hypothesis (H_1). Notably, MAGRes-UNet utilized Mish with the AdamW counterpart, whereas the other proposed models (U-Net, ResUNet, and MAG-UNet) employed ReLU with Adam. As Table 5 indicates, the MAGRes-UNet model using Mish and AdamW emerged as a reliable and statistically significant choice and is preferred for medical image segmentation tasks.

Fig. 13, Fig. 14, and Table 5 illustrate the effect of incorporating MAG and residual blocks in the proposed segmentation model, which led to a marked improvement in performance on the BT T1-weight CE-MRI and HAM10000 skin cancer datasets. MAG enhances feature extraction
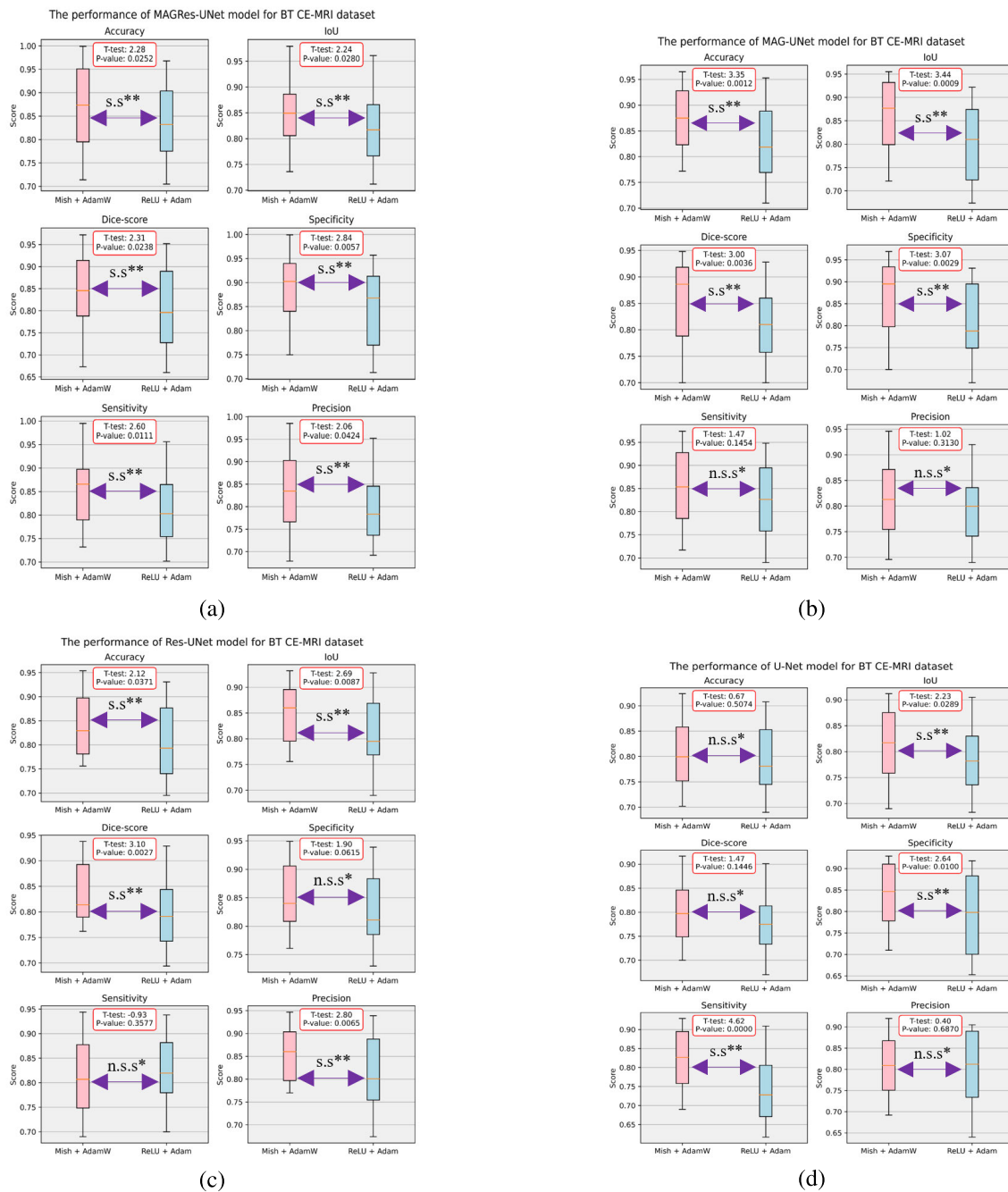
FIGURE 11. Comparison between (a) MAGRes-UNet, (b) MAG-UNet, (c) ResUNet, and (d) U-Net in multi-class BT T1-weighted CE-MRI dataset segmentation in terms of various metrics. (a) MAGRes-UNet consistently outperforms and is statistically significant when compared with (b) MAG-UNet, (c) ResUNet, and (d) U-Net for all metrics. By using MAG for skip connections in the U-Net model, (b) MAG-UNet yields higher segmentation performance than (c) ResUNet and (d) U-Net in most experimental configurations. The notations s.s** and n.s.s* on the arrows indicate the level of significance measured by the *P-value* (s.s** indicates statistical significance, whereas n.s.s* indicates no statistical significance) of the proposed model utilizing Mish with AdamW and ReLU AF with Adam optimizer.

by allowing the model to focus on the most pertinent elements within the input images. Furthermore, MAG aids in capturing long-range dependencies, which are essential for accurately delineating extensive structures and patterns in both datasets. In addition, residual blocks facilitate the training of deeper networks, which are vital for complex

image analysis and improve the resilience of the model to noisy data. These architectural elements contribute to improved generalization and ensure adaptability to diverse cases. The qualitative results are shown in Fig. 13(d) and Fig. 14(d). The quantitative results in Table 5 consistently demonstrate the superiority of the proposed models using
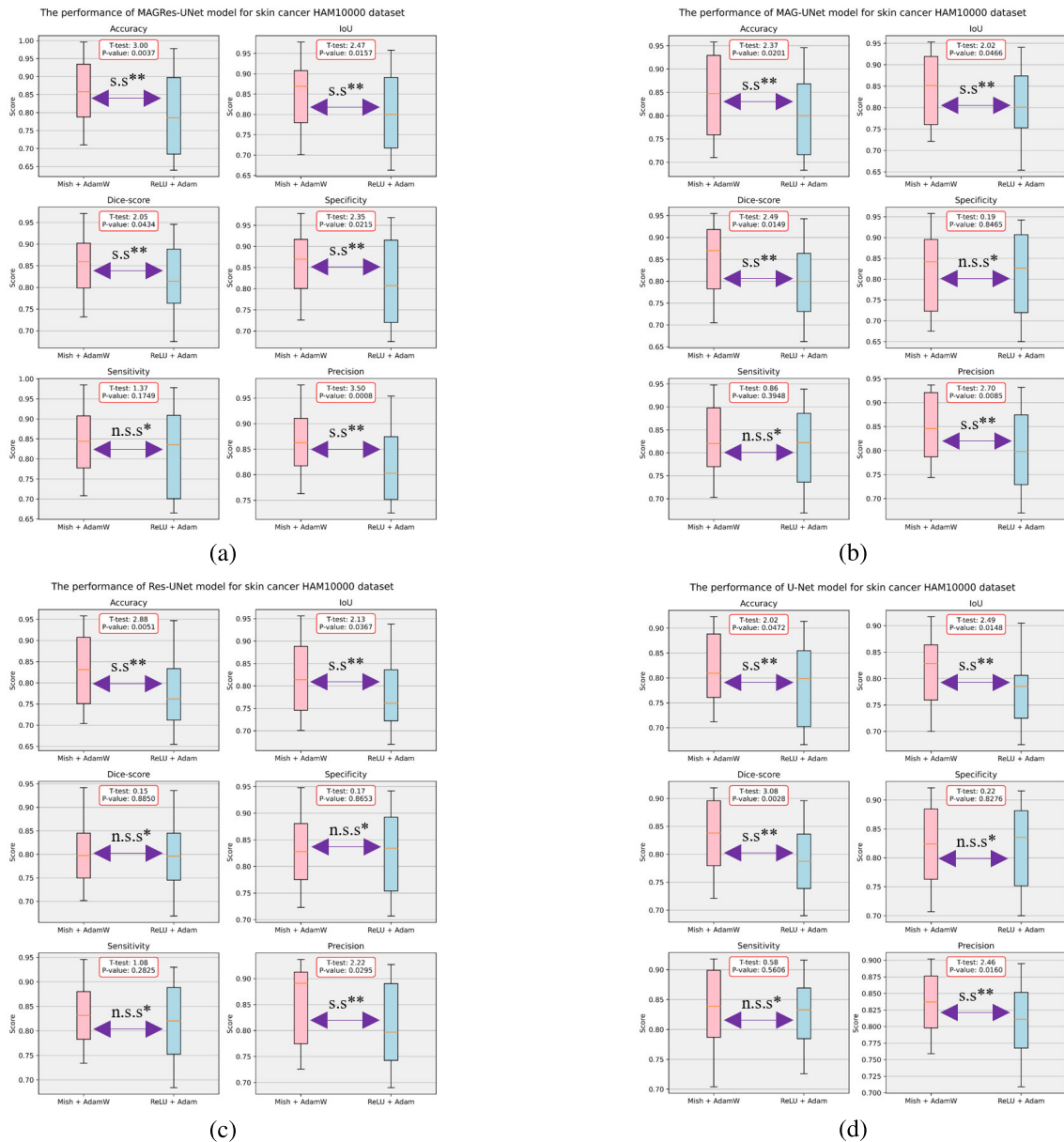
FIGURE 12. Comparison between (a) MAGRes-UNet, (b) MAG-UNet, (c) ResUNet, and (d) U-Net for the task of skin cancer HAM10000 dataset segmentation in terms of different metrics. (a) MAGRes-UNet consistently outperforms and is statistically significant when compared with (b) MAG-UNet, (c) ResUNet, and (d) U-Net for all metrics. By using MAG in the skip connections of the U-Net model, (b) MAG-UNet also yields higher segmentation performance than (c) ResUNet and (d) U-Net in most experimental configurations. The notations s.s** and n.s.s* on the arrows indicate the level of significance measured by the *P-value* (s.s** indicates statistical significance, whereas n.s.s* indicates no statistical significance) of the proposed model utilizing Mish with AdamW and ReLU AF with Adam optimizer.

MAG and residual blocks when compared with those without them in Fig. 13(a)-(c) and Fig. 14(a)-(c), emphasizing their crucial role in achieving precise and robust segmentation, especially in the fields of medical imaging diagnosis.

A close comparison of Figs. 13(a)-(d) and 14(a)-(d) reveals a noticeable trend. In particular, it is evident that in the images in Figs. 13(d) and 14(d), the outputs of the proposed MAGRes-UNet model are strikingly similar to their respective GT masks. This similarity is considerably more

pronounced when compared with the results presented in Figs. 13(a)-(c) and 14(a)-(c), where the U-Net, ResUNet, and MAG-UNet models are applied. This observation is substantiated by our comprehensive analysis, as detailed in Tables 3, 4, and 5. The in-depth evaluation of the models using various metrics confirms that the MAGRes-UNet model, which incorporates MAG, residual blocks enriched with the Mish AF, and AdamW optimizer, outperforms all the other models. This assertion holds for the multiclass brain MRI
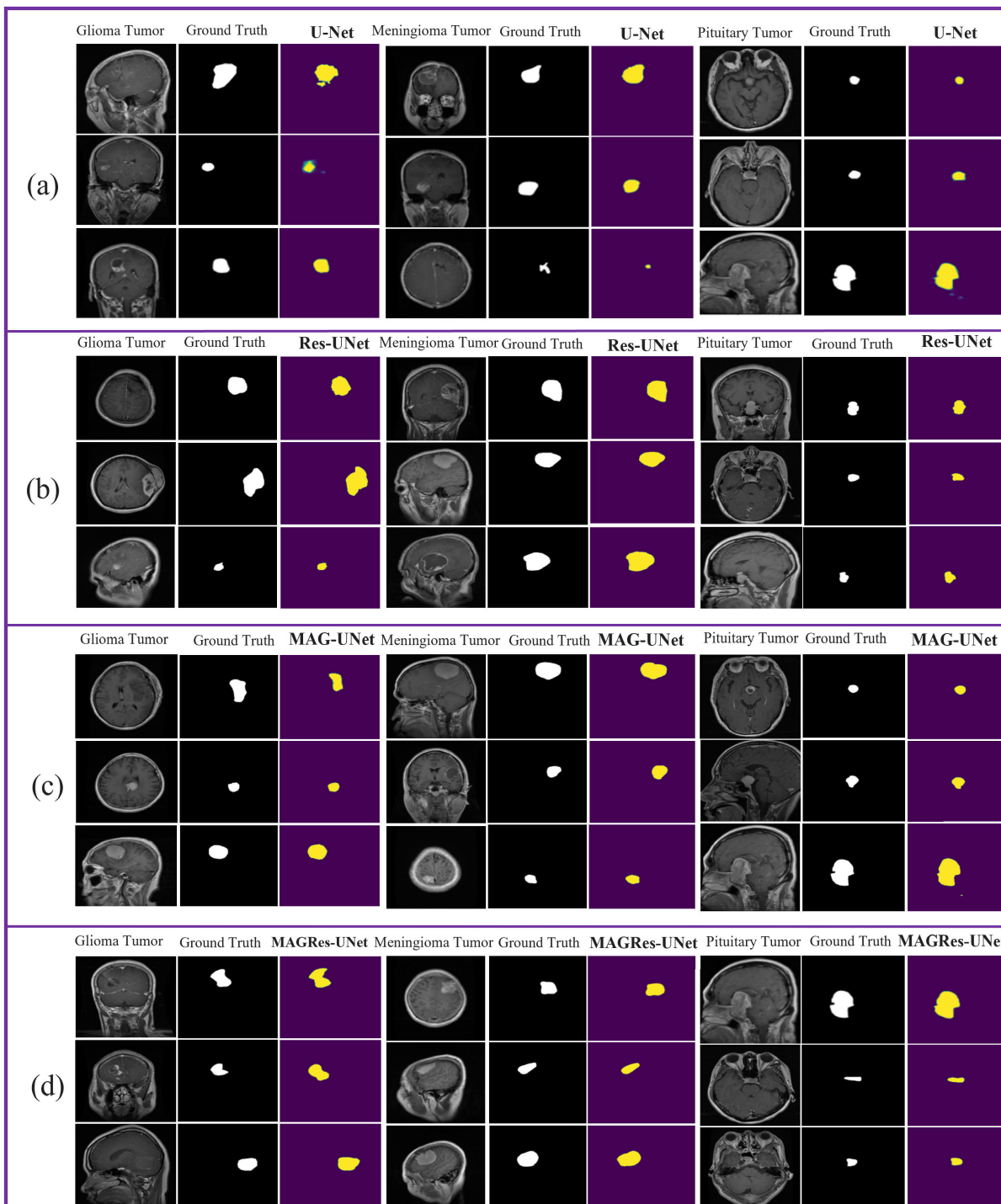
FIGURE 13. Segmentation results of the model on BT T1-weighted CE-MRI dataset: (a) U-Net (b) ResUNet (c) MAG-UNet (d) MAGRes-UNet.

tumor and skin cancer HAM10000 datasets. Fig. 15 presents a holistic view of the qualitative assessment by comparing the various models and datasets. The images in Fig. 15 are BT T1-weighted CE-MRI images displayed in different planes—sagittal, axial, and coronal. Upon closer inspection, it becomes evident that while some models suffer from drawbacks, such as over-segmentation and inaccurate bound-

ary delineation, the segmentation predictions yielded by the MAGRes-UNet model remain exceptionally accurate. The U-Net and ResUNet models exhibited under-segmentation issues for both datasets. Although the MAG-UNet model demonstrated commendable performance, MAGRes-UNet consistently outperformed the other models, as corroborated by the data presented in Tables 3, 4, and 5,
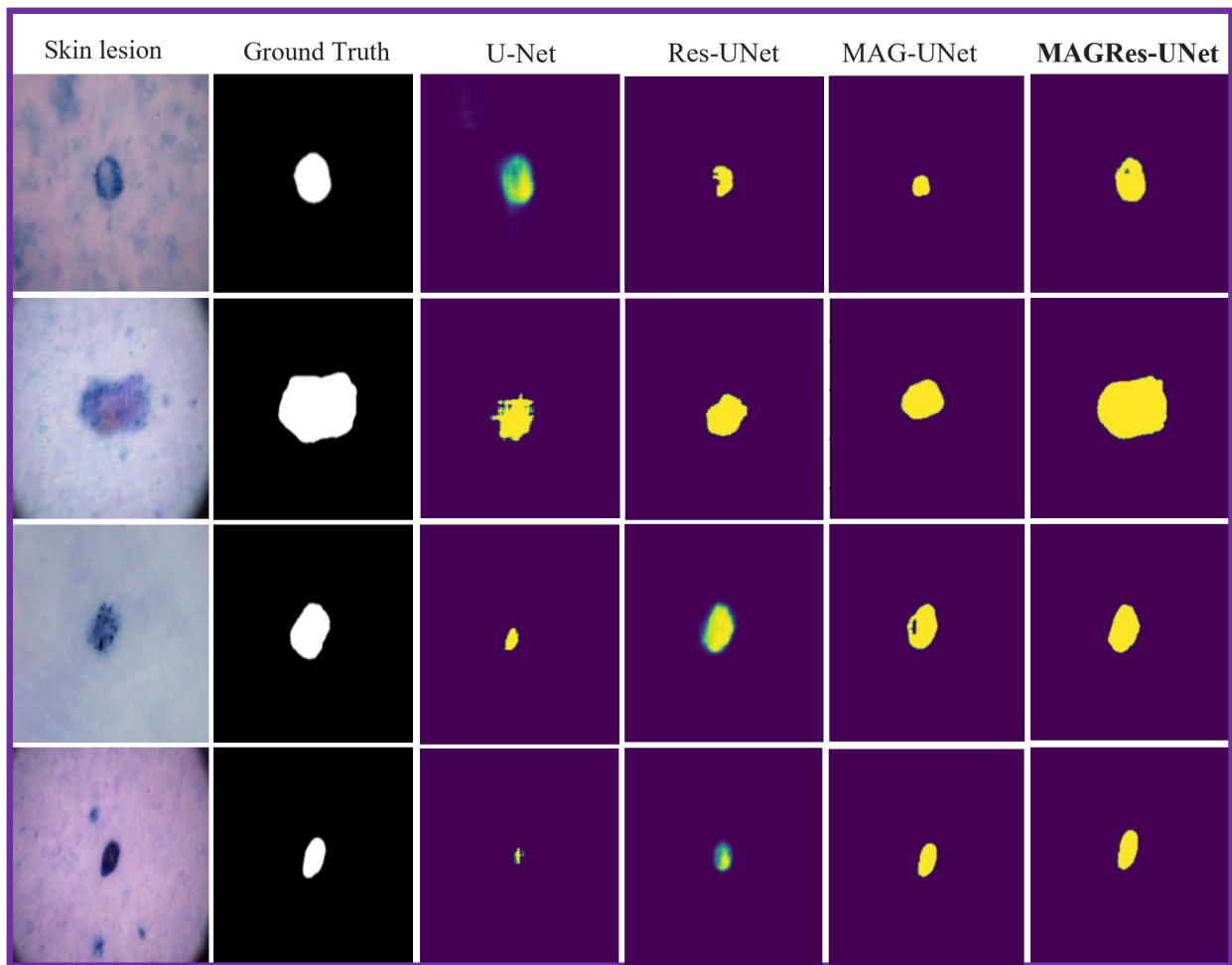
FIGURE 14. Segmentation results of the models (a) U-Net (b) ResUNet (c) MAG-UNet (d) MAGRes-UNet on skin cancer HAM10000 dataset.

along with the visual evidence provided in Figs. 13, 14, and 15.

B. ABLATION EXPERIMENT

Ablation experiments were conducted on two datasets (T1-weighted CE-MRI and skin cancer HAM10000) to evaluate the performance of the proposed model. These datasets were applied to configurations involving MAG mechanisms, residual blocks, AFs, and optimizers. The primary objective was to comprehensively evaluate the model performance based on key metrics (accuracy, IoU, Dice score, specificity, sensitivity, and precision) that accurately gauge their ability to segment BTs and skin cancer in medical images.

The models tested with the BT CE-MRI dataset included U-Net, ResUNet, MAG-UNet, and MAGRes-UNet. Table 5 presents the results of the baseline U-Net model equipped with the MAG modules and residual blocks and employing two distinct activation functions: ReLU and Mish. These were paired with specific optimizers, including Adam and AdamW. It is evident from the analysis that integrating Mish AF consistently enhances the model performance.

Remarkably, the MAGRes-UNet model, which employs Mish AF with the AdamW optimizer, shows exceptional results. This configuration yielded notably higher accuracy, IoU, Dice score, specificity, sensitivity, and precision values than the other configurations detailed in Table 5. This underscores the pivotal role of the combination of the MAG, residual blocks, Mish AF, and optimized weight decay (AdamW) in augmenting the precision of the model in BT segmentation.

A similar trend in performance improvement was evident for the skin cancer HAM10000 dataset. This strengthens the perception that using Mish AF improves the model's performance. Again, the MAGRes-UNet model utilizing Mish AF with AdamW optimizer emerged as the dominant configuration, outperforming the other models across all the metric values shown in Table 5.

Through this ablation analysis, we can observe that the combination of MAG, residual blocks, Mish AF, and AdamW optimizer contributes significantly to the improvements in the proposed MAGRes-UNet model for BT and skin cancer datasets. To confirm the robustness of the model, a statistical
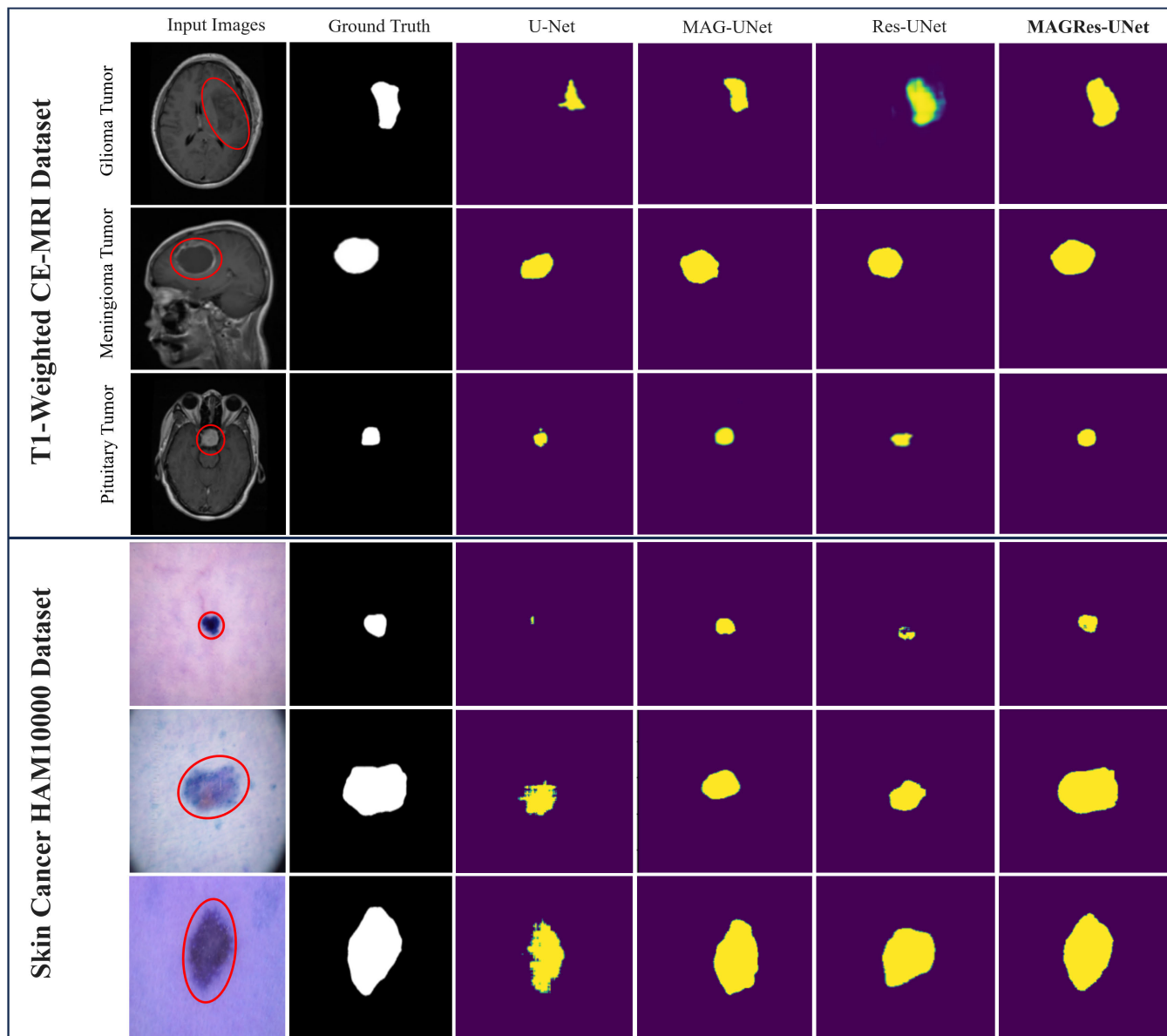
FIGURE 15. Comparison of segmentation results of U-Net, MAG-UNet, ResUNet, and MAGRes-UNet models on both BT T1-weighted CE-MRI and skin cancer HAM10000 datasets. The first column shows the input dataset, the second column shows the GT masked images, and the remaining columns show the model-predicted results.

T-test was performed to evaluate the performance variance between the different configurations. We compared the segmentation outcomes achieved by the MAGRes-UNet model using Mish with AdamW and ReLU with Adam optimizer. Based on a statistical *T-test* analysis, the performances of the two configurations differed significantly. According to all evaluation metrics for BT and skin cancer segmentation tasks, the MAGRes-UNet model, which employs the Mish AF and AdamW optimizer, consistently demonstrated superior segmentation performance. These results are listed in Table 5 and Fig. 15. The obtained *P-values* are well below the predefined significance level ($\alpha = 0.05$), further confirming the statistical significance of these differences, as shown in Table 5. In this table, (s.s**) represents statistical significance,

whereas (n.s.s*) represents no statistical significance. This study validated the feasibility of applying the proposed MAGRes-UNet model to medical imaging applications. Furthermore, it emphasized the importance of AF and optimizers in the development of precise and reliable segmentation methods.

C. COMPARISON WITH SOTA METHODS
This section compares the proposed method with SOTA models to demonstrate its superiority in segmenting the BT and skin cancer datasets. The results are listed in Tables 6 and 7. The experimental results indicate that MAGRes-UNet is more efficient than the SOTA models because it obtains the best results for the metrics. The

TABLE 5. Comprehensive performance comparison of the proposed model with Mish AF and AdamW optimizer and that with ReLU AF and standard Adam optimizer. The evaluation uses various metrics for BT CE-MRI and skin cancer HAM10000 datasets. The proposed MAGRes-UNet, utilizing the Mish AF and AdamW optimizer, outperforms all other models (U-Net, ResUNet, MAG-UNet). In particular, the MAGRes-UNet model is statistically significant in its superiority. The bold values signify the best results, and the red underlined values show the second-best results for both datasets (s.s** means statistically significant, and n.s.s* means no statistical significance).

Method	Dataset	AF and Optimizer	Accuracy (%)	IoU (%)	Dice Score (%)	Specificity (%)	Sensitivity (%)	Precision (%)
U-Net	BT CE-MRI	ReLU + Adam	90.8	90.51	90.11	91.8	90.93	90.57
		Mish + AdamW	92.47	91.2	91.73	92.6	92.81	92.05
T-test			0.67	2.23	1.47	2.64	4.62	0.4
P-value			0.507 n.s.s*	0.028 s.s**	0.144 n.s.s*	0.010 s.s**	0.000 s.s**	0.687 n.s.s*
ResUNet		ReLU + Adam	93.13	92.8	92.91	93.92	93.7	93.93
		Mish + AdamW	95.48	93.25	93.83	94.96	94.47	94.7
T-test			2.12	2.69	3.1	1.9	-0.93	2.8
P-value			0.037 s.s**	0.008 s.s**	0.002 s.s**	0.061 n.s.s*	0.357 n.s.s*	0.006 s.s**
MAG-UNet		ReLU + Adam	95.36	92.28	92.87	93.19	94.83	92.09
		Mish + AdamW	96.53	95.5	94.81	96.9	97.41	94.62
T-test			3.35	3.44	3	3.07	1.47	1.02
P-value			0.001 s.s**	0.000 s.s**	0.003 s.s**	0.002 s.s**	0.145 n.s.s*	0.313 n.s.s*
MAGRes-UNet		ReLU + Adam	96.89	96.1	95.21	95.7	95.61	95.21
		Mish + AdamW	**99.91**	**97.9**	**97.21**	**99.93**	**99.5**	**98.52**
T-test			2.08	2.24	3.31	2.84	2.6	2.06
P-value			0.025 s.s**	0.028 s.s**	0.023 s.s**	0.005 s.s**	0.011 s.s**	0.042 s.s**
U-Net	Skin Cancer HAM10000	ReLU + Adam	91.49	90.28	89.69	91.17	91.06	89.04
		Mish + AdamW	92.3	91.71	90.4	92.13	91.53	90.2
T-test			2.02	2.49	3.08	0.22	0.58	2.46
P-value			0.047 s.s**	0.014 s.s**	0.002 s.s**	0.827 n.s.s*	0.560 n.s.s*	0.016 s.s**
ResUNet		ReLU + Adam	94.7	93.81	93.63	94.21	93.09	92.71
		Mish + AdamW	95.81	94.9	94.26	94.8	94.61	93.7
T-test			2.88	2.13	0.15	0.17	1.08	2
P-value			0.005 s.s**	0.036 s.s**	0.885 n.s.s*	0.865 n.s.s*	0.282 n.s.s*	0.029 s.s**
MAG-UNet		ReLU + Adam	94.69	94.17	94.39	94.26	93.92	93.29
		Mish + AdamW	95.8	95.32	95.5	95.81	94.82	93.7
T-test			2.37	2.02	2.49	0.19	0.86	2.7
P-value			0.020 s.s**	0.046 s.s**	0.014 s.s**	0.846 n.s.s*	0.394 n.s.s*	0.008 s.s**
MAGRes-UNet		ReLU + Adam	97.81	95.89	94.68	96.83	97.81	95.47
		Mish + AdamW	**99.75**	**97.80**	**97.10**	**97.81**	**98.52**	**97.61**
T-test			3.00	2.47	2.05	2.35	1.37	3.50
P-value			0.003 s.s**	0.015 s.s**	0.043 s.s**	0.021 s.s**	0.174 n.s.s*	0.000 s.s**

TABLE 6. Comparison results of the proposed and SOTA models on the BT dataset. The bold values indicate high scores for the respective metrics.

Method	Metrics					
	Accuracy (%)	IoU (%)	Dice score (%)	Specificity (%)	Sensitivity (%)	Precision (%)
TAGU-Net [78]	**99.92**	96.35	97.67	**99.96**	97.76	98.51
O2U-Net [79]	98.04	-	80.83	-	77.39	-
DCNN [80]	98.00	91.00	90.00	99.00	96.00	**99.00**
Mask RCNN [81]	95.10	95.00	95.00	-	95.00	-
SE-ResNet [82]	97.83	97.60	97.40	-	-	-
U-Net	92.47	91.20	91.73	92.60	92.81	92.05
ResUNet	95.48	93.25	93.83	94.96	94.47	94.7
MAG-UNet	96.53	95.50	94.81	96.90	97.41	94.62
MAGRes-UNet*	99.91	**97.90**	**97.21**	99.93	**99.50**	98.52

TABLE 7. Comparison results of the proposed and SOTA models on skin cancer datasets. The bold values indicate high scores for the respective metrics.

Method	Metrics					
	Accuracy (%)	IoU (%)	Dice score (%)	Specificity (%)	Sensitivity (%)	Precision (%)
SASegNet [83]	96.60	97.00	96.57	97.40	97.80	97.52
MFSNet [84]	-	90.20	90.60	**99.90**	**99.90**	**-**
iFCN [39]	95.30	78.34	88.64	98.08	85.44	-
MAAU [85]	95.50	80.90	88.10	-	-	90.90
SE-ResNet [82]	97.83	97.6	97.4	-	-	-
U-Net [86]	93.14	93.14	84.76	92.63	94.79	-
ResU-Net [86]	93.67	93.67	85.67	93.38	94.54	-
RecU-Net [86]	93.80	93.80	85.92	93.95	93.34	-
R2U-Net [86]	94.24	94.21	86.16	94.25	94.14	-
U-Net	92.30	91.71	90.40	92.13	91.53	90.20
ResUNet	95.81	94.90	94.26	94.80	94.61	93.70
MAG-UNet	95.80	95.32	95.50	95.81	94.82	93.70
MAGRes-UNet*	**99.75**	**97.80**	**97.10**	99.93	98.52	**97.61**

prediction mask generated by MAGRes-UNet is highly consistent with the GT mask, as shown in Figs. 13, 14 and 15. The structure of MAGRes-UNet is similar to those of U-Net, ResUNet, and MAG-UNet. However, MAGRes-UNet can extract multiscale contextual information, which allows it to capture high-level features with more spatial context information.

The results in Tables 6 and 7 unequivocally demonstrate the superiority of MAGRes-UNet over the SOTA models. Tables 6 and 7 show that the results of the various methodologies are comparable. However, the proposed MAGRes-UNet showed superior performance for both medical datasets. Among the most cutting-edge techniques currently used for segmenting medical images [39], [78], [79], [80], [81], [82], [83], [84], [85], [86], the methods in [78], [80], and [84] performed slightly better for some metrics shown in Tables 6 and 7 than our suggested 2D model. Despite this, our model demonstrated enhanced segmentation of the tumor area compared to the SOTA model. Overall, the proposed 2D model is advantageous, particularly considering the low complexity of the model and the shorter time required for training.

V. CONCLUSION

This study investigated the effectiveness of MAG and residual blocks for medical image segmentation using the standard U-Net architecture. MAGRes-UNet adds a series of attention gates to the skip connection to highlight relevant feature information while preventing insignificant and noisy feature responses, which is helpful in small-scale tumor segmentation. In addition, the introduction of residual blocks alleviated the vanishing gradient problem. The proposed method was extensively evaluated using two benchmark datasets. The MAGRes-UNet model utilizing the Mish AF and AdamW optimizer delivered excellent accuracy (99.91%), IoU (97.90%), Dice score (97.21%), specificity (99.93%), sensitivity (99.50%), and precision (98.52%) on the BT T1-weighted CE-MRI dataset. For the skin cancer HAM10000 dataset, the proposed MAGRes-UNet model achieved an accuracy of 99.75%, IoU of 97.80%, Dice score of 97.10%, specificity of 97.81%, sensitivity of 98.52%, and precision of 97.61%. Based on these results, it can be concluded that the MAGRes-UNet model with Mish and AdamW is superior to the other models (MAG-UNet, ResUNet, and U-Net) with the ReLU AF and standard Adam optimizer. Furthermore, a statistical *T-test* analysis revealed a significant difference in the performance of the proposed method using Mish with AdamW when compared with ReLU with Adam for all metrics on both medical datasets, as shown in Table 5.

In conclusion, the proposed MAGRes-UNet can enrich semantic information, improve feature learning capability, and focus on small-scale tumor information. MAGRes-UNet requires a larger amount of contextual information and local details between different slices because the 2D U-Net model cannot utilize the 3D information from the MRI data. To enhance the segmentation performance of MAGRes-Net in the future, we will investigate 3D network architectures and apply the enhanced architecture to additional datasets to demonstrate its generalizability. However, our segmentation method and current MRI scanning systems need improvements. Moreover, BT segmentation remains challenging owing to the complexity of MRI brain images and the limitations of labels for DL models. One of our future projects will be to develop a clinical medical segmentation model.

REFERENCES

[1] S.-T. Tran, C.-H. Cheng, T.-T. Nguyen, M.-H. Le, and D.-G. Liu, "TMD-UNet: Triple-UNet with multi-scale input features and dense skip connection for medical image segmentation," *Healthcare*, vol. 9, no. 1, p. 54, Jan. 2021.

[2] *Nh Narayana Health, Health for All, All for Health, NH CARES Brain Tumour, Types, Risk Factors, Symptoms, and Surgery*. Accessed: Nov. 10, 2023. [Online]. Available: https://www.narayanahealth.org/diseases/brain-tumour

[3] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2018.

[4] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Phys. Med. Biol.*, vol. 58, no. 13, pp. 97–129, Jul. 2013.

[5] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[6] S. Iqbal, M. U. Ghani, T. Saba, and A. Rehman, "Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN)," *Microsc. Res. Technique*, vol. 81, no. 4, pp. 419–427, Apr. 2018.

[7] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, Munich, Germany, 2015, pp. 234–241.

[9] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-UNet: UNet-like pure transformer for medical image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Tel Aviv, Israel. Berlin, Germany: Springer, Oct. 2022, pp. 205–218.

[10] Z. Rauf, A. Sohail, S. H. Khan, A. Khan, J. Gwak, and M. Maqbool, "Attention-guided multi-scale deep object detection framework for lymphocyte analysis in IHC histological images," *Microscopy*, vol. 72, no. 1, pp. 27–42, Feb. 2023.

[11] M. L. Ali, Z. Rauf, A. R. Khan, and A. Khan, "Channel boosting based detection and segmentation for cancer analysis in histopathological images," in *Proc. 19th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Aug. 2022, pp. 1–6.

[12] S. H. Khan, A. Khan, Y. S. Lee, M. Hassan, and W. K. Jeong, "Segmentation of shoulder muscle MRI using a new region and edge based deep auto-encoder," *Multimedia Tools Appl.*, vol. 82, no. 10, pp. 14963–14984, Apr. 2023.

[13] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, Dec. 2020.

[14] T. Hussain and H. Shouno, "Explainable deep learning approach for multi-class brain magnetic resonance imaging tumor classification and localization using gradient-weighted class activation mapping," *Information*, vol. 14, no. 12, p. 642, Nov. 2023.

[15] N. Ibtehaz and M. S. Rahman, "MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.

[16] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[17] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[18] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2405–2414, Nov. 2006.

[19] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, "Automatic nuclei segmentation in H&E stained breast cancer histopathology images," *PLoS One*, vol. 8, no. 7, Jul. 2013, Art. no. e70221.

[20] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky, "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 137–154, Feb. 2003.

[21] J. Walsh, A. Othmani, M. Jain, and S. Dev, "Using U-Net network for efficient brain tumor segmentation in MRI images," *Healthcare Anal.*, vol. 2, Nov. 2022, Art. no. 100098.

[22] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "NAS-UNet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, pp. 44247–44257, 2019.

[23] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Deep Learn. Medical Image Anal. Multimodal Learn. Clin. Decision Support, 4th Int. Workshop, 8th Int. Workshop*, Granada, Spain. Berlin, Germany: Springer, 2018, pp. 3–11.

[24] Y. Cao, S. Liu, Y. Peng, and J. Li, "DenseUNet: Densely connected UNet for electron microscopy image segmentation," *IET Image Process.*, vol. 14, no. 12, pp. 2682–2689, Oct. 2020.

[25] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[26] S. Qamar, H. Jin, R. Zheng, P. Ahmad, and M. Usama, "A variant form of 3D-UNet for infant brain segmentation," *Future Gener. Comput. Syst.*, vol. 108, pp. 613–623, Jul. 2020.

[27] J. Zhang, X. Lv, H. Zhang, and B. Liu, "AResU-Net: Attention residual U-Net for brain tumor segmentation," *Symmetry*, vol. 12, no. 5, p. 721, May 2020.

[28] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059.

[29] J. Zhang, Q. Qin, Q. Ye, and T. Ruan, "ST-UNet: Swin Transformer boosted U-Net with cross-layer feature enhancement for medical image segmentation," *Comput. Biol. Med.*, vol. 153, Feb. 2023, Art. no. 106516.

[30] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation," *Frontiers Comput. Neurosci.*, vol. 13, p. 56, Aug. 2019.

[31] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," 2018, *arXiv:1810.07810*.

[32] K. Møllersen, H. M. Kirchesch, T. G. Schopf, and F. Godtliebsen, "Unsupervised segmentation for digital dermoscopic images," *Skin Res. Technol.*, vol. 16, no. 4, pp. 401–407, Nov. 2010.

[33] D. D. Gomez, C. Butakoff, B. K. Ersboll, and W. Stoecker, "Independent histogram pursuit for segmentation of skin lesions," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 157–161, Jan. 2008.

[34] M. E. Yüksel and M. Borlu, "Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 976–982, Aug. 2009.

[35] M. E. Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Res. Technol.*, vol. 19, no. 1, pp. 252–258, Feb. 2013.

[36] F. Peruch, F. Bogo, M. Bonazza, V.-M. Cappelleri, and E. Peserico, "Simpler, faster, more accurate melanocytic lesion segmentation through MEDS," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 557–565, Feb. 2014.

[37] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Comput. Methods Programs Biomed.*, vol. 162, pp. 221–231, Aug. 2018.

[38] Y. Yuan and Y.-C. Lo, "Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 519–526, Mar. 2019.

[39] Ş. Öztürk and U. Özkaya, "Skin lesion segmentation with improved convolutional neural network," *J. Digit. Imag.*, vol. 33, no. 4, pp. 958–970, Aug. 2020.

[40] F. Xie, J. Yang, J. Liu, Z. Jiang, Y. Zheng, and Y. Wang, "Skin lesion segmentation using high-resolution convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 186, Apr. 2020, Art. no. 105241.

[41] X. Tong, J. Wei, B. Sun, S. Su, Z. Zuo, and P. Wu, "ASCU-Net: Attention gate, spatial and channel attention U-Net for skin lesion segmentation," *Diagnostics*, vol. 11, no. 3, p. 501, Mar. 2021.

[42] S. Kadry, D. Taniar, R. Damaševicius, V. Rajinikanth, and I. A. Lawal, "Extraction of abnormal skin lesion from dermoscopy image using VGG-SegNet," in *Proc. 7th Int. Conf. Bio Signals, Images, Instrum. (ICBSII)*, Mar. 2021, pp. 1–5.

[43] V. Rajinikanth, S. Kadry, R. Damaševicius, D. Sankaran, M. A. Mohammed, and S. Chander, "Skin melanoma segmentation using VGG-UNet with Adam/SGD optimizer: A study," in *Proc. 3rd Int. Conf. Intell. Comput. Instrum. Control Technol. (ICICICT)*, Aug. 2022, pp. 982–986.

[44] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[45] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[46] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[47] H. Wang, P. Cao, J. Wang, and O. R. Zaiane, "UCTransNet: Rethinking the skip connections in U-Net from a channel-wise perspective with transformer," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 3, pp. 2441–2449.

[48] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[49] W. Chi, L. Ma, J. Wu, M. Chen, W. Lu, and X. Gu, "Deep learning-based medical image segmentation with limited labels," *Phys. Med. Biol.*, vol. 65, no. 23, Dec. 2020, Art. no. 235001.

[50] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "CPFNet: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020.

[51] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[52] X. Zhao, M. Huang, L. Li, X. S. Qi, and S. Tan, "Multi-to-binary network (MTBNet) for automated multi-organ segmentation on multi-sequence abdominal MRI images," *Phys. Med. Biol.*, vol. 65, no. 16, Aug. 2020, Art. no. 165013.

[53] D. Misra, "Mish: A self regularized non-monotonic activation function," 2019, *arXiv:1908.08681*.

[54] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.

[55] S. Dev, H. Javidnia, M. Hossari, M. Nicholson, K. McCabe, A. Nautiyal, C. Conran, J. Tang, W. Xu, and F. Pitié, "Identifying candidate spaces for advert implantation," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Oct. 2019, pp. 503–507.

[56] S. Dev, M. Hossari, M. Nicholson, K. McCabe, A. Nautiyal, C. Conran, J. Tang, W. Xu, and F. Pitié, "Localizing adverts in outdoor scenes," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2019, pp. 591–594.

[57] S. Dev, S. Manandhar, Y. H. Lee, and S. Winkler, "Multi-label cloud segmentation using a deep network," in *Proc. USNC-URSI Radio Sci. Meeting*, Jul. 2019, pp. 113–114.

[58] O. Ali, H. Ali, S. A. A. Shah, and A. Shahzad, "Implementation of a modified U-Net for medical image segmentation on edge devices," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 11, pp. 4593–4597, Nov. 2022.

[59] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical image segmentation review: The success of U-Net," 2022, *arXiv:2211.14830*.

[60] Y. Deng, Y. Hou, J. Yan, and D. Zeng, "ELU-Net: An efficient and lightweight U-Net for medical image segmentation," *IEEE Access*, vol. 10, pp. 35932–35941, 2022.

[61] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.

[62] J. Xu, Z. Li, B. Du, M. Zhang, and J. Liu, "Reluplex made more practical: Leaky ReLU," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2020, pp. 1–7.

[63] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," 2015, *arXiv:1511.07289*.

[64] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017, *arXiv:1710.05941*.

[65] R. Kumar, "APTx: Better activation function than MISH, SWISH, and ReLU's variants used in deep learning," 2022, *arXiv:2209.06119*.

[66] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[67] O. Sharma, "A new activation function for deep neural network," in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, Feb. 2019, pp. 84–86.

[68] R. Raza, U. I. Bajwa, Y. Mehmood, M. W. Anwar, and M. H. Jamal, "DResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 103861.

[69] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.

[70] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, *arXiv:1508.04025*.

[71] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6450–6458.

[72] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 683–687.

[73] S. Wu, S. Zhong, and Y. Liu, "Deep residual learning for image steganalysis," *Multimedia Tools Appl.*, vol. 77, no. 9, pp. 10437–10453, May 2018.

[74] L. Bi, J. Kim, A. Kumar, M. Fulham, and D. Feng, "Stacked fully convolutional networks with multi-channel learning: Application to medical image segmentation," *Vis. Comput.*, vol. 33, nos. 6–8, pp. 1061–1071, Jun. 2017.

[75] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 16, Jun. 2016, pp. 770–778.

[76] J. Cheng, "Brain magnetic resonance imaging tumor dataset," *Figshare MRI Dataset Version*, vol. 5, 2017.

[77] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, pp. 1–9, Aug. 2018.

[78] H. Huang, P. Liu, and J. Liu, "TAGU-Net: Transformer convolution hybrid-based U-Net with attention gate for atypical meningioma segmentation," *IEEE Access*, vol. 11, pp. 53207–53223, 2023.

[79] S. A. Zargari, Z. S. Kia, A. M. Nickfarjam, D. Hieber, and F. Holl, "Brain tumor classification and segmentation using dual-outputs for U-Net architecture: O2U-Net," *Stud. Health Technol. Inform.*, vol. 305, pp. 93–96, Jul. 2023.

[80] A. M. Mostafa, M. Zakariah, and E. A. Aldakheel, "Brain tumor segmentation using deep learning on MRI images," *Diagnostics*, vol. 13, no. 9, p. 1562, Apr. 2023.

[81] M. Masood, T. Nazir, M. Nawaz, A. Javed, M. Iqbal, and A. Mehmood, "Brain tumor localization and segmentation using mask RCNN," *Frontiers Comput. Sci.*, vol. 15, no. 6, Dec. 2021, Art. no. 156338.

[82] P. R. Kumar, K. Bonthu, B. Meghana, K. S. Vani, and P. Chakrabarti, "Multi-class brain tumor classification and segmentation using hybrid deep learning network model," *Scalable Comput., Pract. Exper.*, vol. 24, no. 1, pp. 69–80, Apr. 2023.

[83] R. Kadirappa and S.-B. Ko, "An automated multi-class skin lesion diagnosis by embedding local and global features of dermoscopy images," *Multimedia Tools Appl.*, vol. 82, no. 22, pp. 34885–34912, Sep. 2023.

[84] H. Basak, R. Kundu, and R. Sarkar, "MFSNet: A multi focus segmentation network for skin lesion segmentation," *Pattern Recognit.*, vol. 128, Aug. 2022, Art. no. 108673.

[85] P. T. Le, B.-T. Pham, C.-C. Chang, Y.-C. Hsu, T.-C. Tai, Y.-H. Li, and J.-C. Wang, "Anti-aliasing attention U-Net model for skin lesion segmentation," *Diagnostics*, vol. 13, no. 8, p. 1460, Apr. 2023.

[86] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.

TAHIR HUSSAIN received the master's degree in computer and communication engineering from National Cheng Kung University (NCKU), Tainan, Taiwan. He is currently pursuing the Ph.D. degree with the Shouno Laboratory, Department of Informatics, The University of Electro-Communications, Tokyo, Japan. He is a MEXT Scholar. His research interest includes utilizing deep learning models for the detection, classification, and segmentation of medical images.
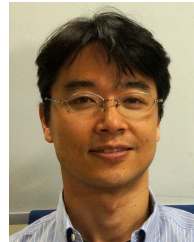
HAYARU SHOUNO (Member, IEEE) received the Ph.D. degree in engineering from Osaka University, Osaka, in 1999. He is currently a Professor with the Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan. His research interests include computer vision, machine learning, and neural networks. He is an Acting Editor of *Neural Networks* and an Elected Governor of Asia Pacific Neural Network Society (APNNS).

● ● ●