

Toward more accurate diagnosis of multiple sclerosis: Automated lesion segmentation in brain magnetic resonance image using modified U-Net model

Bakhtiar Amaludin¹  | Seifedine Kadry²  | Fung Fung Ting³  | David Taniar¹ 

¹Faculty of Information Technology, Monash University, Clayton, Australia

²Department of Applied Data Science, Noroff University College, Kristiansand, Norway

³School of Information Technology, Monash University, Kuala Lumpur, Malaysia

Correspondence

David Taniar, Faculty of Information Technology, Monash University, Clayton, VIC 3800, Australia.
Email: david.taniar@monash.edu

Abstract

Early diagnosis of multiple sclerosis (MS) through the delineation of lesions in the brain magnetic resonance imaging is important in preventing the deteriorating condition of MS. This study aims to develop a modified U-Net model for automating lesions segmentation in MS more accurately. The proposed modified U-Net uses residual dense blocks to replace the standard convolutional stacks and incorporates three axes (axial, sagittal, and coronal) of 2D slice images as input. Furthermore, a custom fusion method is also introduced for merging the predicted lesions from different axes. The model was implemented on ISBI2015 and OpenMS data sets. On ISBI2015, the proposed model achieves the best overall score of 93.090% and DSC of 0.857 on the OpenMS data set.

KEYWORDS

brain, computer vision, deep learning, medical image, MRI, MS lesion, segmentation, U-net

1 | INTRODUCTION

Multiple sclerosis (MS) is a chronic autoimmune disorder that causes significant damage to the central nervous system, particularly the brain, spinal cord, and optic nerves. MS patients often experience long-term issues such as impaired balance, vision, and muscle control. MS affects 2.5 million people worldwide.¹

Early detection of MS is crucial to delay the more severe progression of the disease. Some studies confirm that late treatment for MS results in greater deterioration compared to those who begin treatment earlier.^{2,3} For early detection, a doctor or radiologist will perform a

physical and neurological examination as an initial diagnosis, followed by several tests to further confirm MS. One of the critical stages of the MS examination is brain investigation using magnetic resonance imaging (MRI) images.⁴ Through MRI analysis, experts will examine and delineate abnormal tissues found in brain MRI images considered MS lesions. However, accurate detection of MS lesions is a challenging task.⁵ This is because delineating lesions manually within brain MRI images is labor-intensive⁶ and is influenced by the skill and experience of the experts.⁷ Consequently, researchers have proposed various automated lesion segmentation schemes.

However, unlike other applications such as brain tumor detection, MS lesion segmentation remains a challenging task. The difficulty arises from the unique characteristics of brain lesions, which are often hard to detect

Abbreviations: MRI, magnetic resonance imaging; MS, multiple sclerosis.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *International Journal of Imaging Systems and Technology* published by Wiley Periodicals LLC.

and may intersect with the brain's white matter.⁸ Lesions also vary in size (typically quite small) and are distributed across various areas of the brain.^{5,9} It is not surprising if two or more experts occasionally disagree when determining lesion shape, size, and location.

Given its significance and urgency, numerous researchers persistently explore and propose deep learning (DL) models based on convolutional neural networks (CNNs) to enhance the accuracy of MS lesion prediction. However, this remains a formidable challenge due to the unsatisfactory performance of models developed to date.¹⁰ This study aims to develop a more accurate approach by optimizing the U-Net model, which is highlighted as follows:

1. Replace the standard convolutional stack in U-Net model using residual dense block in the encoder and decoder layers. This block is the combination of residual block and dense block as originally introduced as part of ResNet¹¹ and DenseNet.¹² This architecture gives an advantage by achieving better performance and eliminating unimportant features.

2. The model is designed to make a segmentation based on different input axes. Rather than using the simple technique (such as union, voting, and intersection), a custom fusion (CF) will be used for merging the predictions from these axes.

2 | RELATED WORKS

In recent years, numerous DL methods related to computer vision for disease detection tasks have been extensively developed. The popular DL model used for these tasks is based on convolutional neural networks (CNNs), as they can effectively extract complex patterns from an image.^{10,13–19} In terms of automating MS lesion segmentation, plenty of automatic segmentation models also have been proposed, as summarized in Table 1.

In the early studies, the traditional CNN scheme was commonly used in MS lesion automatic segmentation. This scheme is generally constituted by some convolution operations, pooling, and a fully-connected (dense) layer.

TABLE 1 State-of-art deep learning for MS lesion segmentation.

Base	Authors	Input	Main contribution	Data set used
CNN	Ansari et al. ⁵	2D	Employ Inception Modules	ISBI2015
	Aslani et al. ⁷	2D	Employ ResNet Block	ISBI2015
	Valverde et al. ²⁰	3D	Introduce 11 CNN + 3 dense layers	ISBI2015 + Private
	Birenbaum and Greenspan ²¹	3D	Process MRI in Longitudinal way	ISBI2015
	Afzal et al. ²²	2D	Using parallel path to reduce FP	MICCAI2008
	Yildirim and Dandil ²³	3D	Introduce Mask R-CNN	ISBI2015
	Zhang et al. ²⁴	2D	Employ Anatomical Convolutional	ISBI2015
FCNN	Roy et al. ⁹	2D	Employ Cascaded CNN	ISBI2015
	Gessert et al. ²⁵	3D	Employ Attention-guided module	Private Data Set
	Kamnitsas et al. ²⁶	3D	Parallel pathways CNN	ISBI2015
	Essa et al. ²⁷	2D	Introduce Region-based CNN	MICCAI2008
	Chen et al. ²⁸	3D	Present Modified Attention Gate	ISBI2015
	Zhang et al. ²⁹	2.5D	Use 100 layers CNN	ISBI2015
U-Net	Kumar et al. ³⁰	2D	Add dense skip connection	MICCAI2016
	Salehi et al. ³¹	3D	Employ Tversky Loss	Private
	Hashemi et al. ³²	3D	Employ Asymmetric Loss	MICCAI2016
	Hu et al. ³³	3D	Employ Attention Module	ISBI2015
	Hashemi et al. ³⁴	2D	Employ Attention Module	ISBI2015
	Hou et al. ³⁵	3D	Introduce Cross Attention module	ISBI2015
	Alijamaat et al. ³⁶	3D	Using Wavelet Pooling	MICCAI 2016
	Kamraoui et al. ³⁷	3D	Using 2 modalities for robustness	ISBI2015 and MICCAI 2016

Note: The table shows how various research uses different models and input handling. The contribution from each model and different data sets used for evaluating their approach is also highlighted.

Abbreviations: FCNN, fully convolution neural network; MS, multiple sclerosis.

Birenbaum and Greenspan²¹ proposed multiview longitudinal CNN with two main steps: extracting brain white matter as lesions candidate and then using CNN to extract the lesions from the candidate of two points of time images.

Another earliest proposed model is a cascaded CNN scheme using 3D patches proposed by Valverde et al.²⁰ They used the CNN scheme to select candidate voxels and another CNN scheme to detect lesions from candidate voxels. Afzal et al.²² also introduced a similar scheme,²⁰ with the difference in the input they used (2D) instead of 3D patches to alleviate the computational complexity. While Ulloa et al.³⁷ proposed a CNN model that contains three convolutional layers and dense layers. Ansari et al.⁵ proposed a CNN module with inception blocks adapted from the GoogLeNet model and optimized using Binary Cross Entropy (BCE) Loss and Structural Similarity Index Measure (SSIM) Loss. Yıldırım et al.²³ proposed using the Mask R-CNN architecture introduced by He et al.³⁸ for specific MS lesion segmentation tasks. They used ResNet50 and ResNet101 as the backbone of their Mask R-CNN architecture. Similarly, Aslani et al.⁷ also proposed to use ResNet-based architecture with multi-branch down-sampling to enable information from different image modalities to be encoded separately. Finally, the multiscaled feature maps from different modalities are fused and up-sampled to obtain the final prediction.

Fully convolution neural network (FCNN) is then implemented to improve traditional CNN. Compared to CNN, FCNN comprises only convolutional layers without any fully-connected layers. This architecture makes the network more efficient and allows it to obtain more spatial context rather than predict pixel-level lesion like traditional CNN-based models do.³⁹ Brosch et al.⁴⁰ proposed the encoder–decoder model scheme by employing residual blocks for the encoder part with the whole brain MRI images as input. Roy et al.⁹ introduced an FCNN scheme to annotate white matter lesions from MRI. The proposed model contains a parallel convolution for each modality of 2D slices. Then, the learned features are concatenated and followed by another convolutions layer. Kamnitsas et al.²⁶ proposed an FCNN scheme with a pair of parallel paths, the first for processing high-resolution modalities and the other for lower-resolution images. Such an arrangement is designed to minimize contextual information loss. These paths are fused through concatenation and used as input of another fully connected layer.

FCNN scheme is also used by Gessert et al.²⁵ for a temporal segmentation model for new and emerging lesions that combines the lesions' information from two points of time. They used a two-time processing path connected by attention-guided interactions for exchanging information. While Essa et al.²⁷ proposed a region-based

CNN (R-CNN) using 3D patches as input. They use two different R-CNN models with different inputs, and the prediction from those models is then fused using the adaptive neuro-fuzzy inference system. Another form of FCNN, called 1000-layers CNN or Tiramisu scheme, introduced by Jegou et al.⁴¹ was also adapted for MS segmentation by Zhang et al.²⁹

Built on the top of FCNN architecture,³⁹ Ronneberger et al.⁴² proposed an extended model called U-Net by adding a contracting path from high-resolution input to up-sampling outputs in each layer level. Due to its superiority, the U-Net model has been used as a standard building block for many medical image segmentation tasks. However, despite their success, some extensive works have been conducted to explore the possibility of making a better version of U-Net by incorporating other DL techniques.⁴³ Accordingly, for MS lesion segmentation, U-Net has been implemented in many ways. Some interventions to the original model include improving the encoder–decoder part and modifying the skip connection.

Kumar et al.³⁰ modified the original U-Net by replacing the convolution layer on the encoder part with a dense block to alleviate the vanishing gradient. Another model based on the combination of U-Net and Recurrent Neural Network has been introduced by Zhang et al.⁴⁴ They incorporated a Recurrent Slice-wise Attention block to capture long-range voxel dependencies. Also, Hu et al.³³ introduced a 3D attention context module in addition to 3D attention blocks in their U-Net architecture. This module was developed to improve more detailed spatial information on lesion features in the decoding phase. Alijamaat et al.³⁶ proposed a modified U-Net scheme by incorporating wavelet transform-based pooling as a replacement for standard max pooling in their decoder layers. Zhang et al.²⁴ modified their 3D U-Net with two anatomic convolutional modules and optimized using a novel anatomical lesion-wise loss objective function.

Chen et al.²⁸ proposed an encoder–decoder scheme by introducing novel attention and a graph-driven network (DAG-Net) to capture the spatial correlations and the global context for better lesion representations. An attention module to U-Net is also proposed by Hashemi et al.³⁴ and Hou et al.³⁵ to improve segmentation. Feng et al.⁴⁵ used an optimized U-Net by proposing a dropout mechanism for their input modalities to make a more robust model. Hashemi et al.³² proposed the FC-Dense-Net model and Asymmetric Loss function in their scheme. While Salehi et al.³¹ and Hashemi et al.³² proposed a U-Net scheme optimized using Tversky Loss as the objective function.

To conclude, all of the proposed models attempt to improve the segmentation performance and yields in different performances in some benchmark data sets. As

detecting lesions in MS is challenging,^{5,9} no particular model has perfect performance although some complex network modification was made.

Thus, this study will focus on building an alternative model that can improve lesion segmentation performance. We also will focus on developing the model with reduced model parameters to obtain more efficient architecture.

3 | METHODOLOGY

3.1 | Proposed model

The proposed model is built on top of the U-Net architecture (as the baseline model) introduced by Ronneberger et al.⁴² The model comprises four encoder and decoder layers with a skip connection between the encoder and decoder at the same level. The encoder layer extracts spatial information from the input image, while the decoder generates a prediction based on information from the encoder.

In U-Net, features are extracted during two convolutional operations, followed by Rectified Linear Unit (ReLU) activation function. The two stacked convolution layers extract features with a kernel size of 3×3 with the initial number of filters set to 64. After the last convolution, down-sampling in the encoder part will be performed using max-pooling so that the next deeper layer will receive lower-resolution features but double the number of filters. The down-sampling is accomplished

until the fourth layer before coming into the bottleneck. In contrast, the decoder part will up-sample the features from the previous layer, starting from the bottleneck (the deepest layer) to the last layer. In this phase, the feature map size will be doubled, but the number of filters in each layer will be halved. Finally, In the final output layer, Features extracted from the decoder will be concatenated with the features from the corresponding encoder layers through the skip connection, and the sigmoid activation function is used to classify whether the pixel will be considered a lesion or not.

To improve the U-Net, we make two key modifications to the U-Net architecture which are: making architecture less complex and incorporation of the residual dense block. The complete architecture of the proposed model is shown in Figure 1.

3.1.1 | Reduce model complexity

The proposed model is shallower as it contains three encoders and three decoder layers. Thus the model is more compact and expected to have fewer parameters than the original. The number of features of each layer is expected to be 64, 128, 256, and 512, respectively. Furthermore, in the input layer, the group convolution with $g = 4$ (value of g reflects the number of available modalities) will be performed immediately. The aim is to generate independent features from each modality. It also reduces the computation cost as it has fewer parameters than standard convolution.

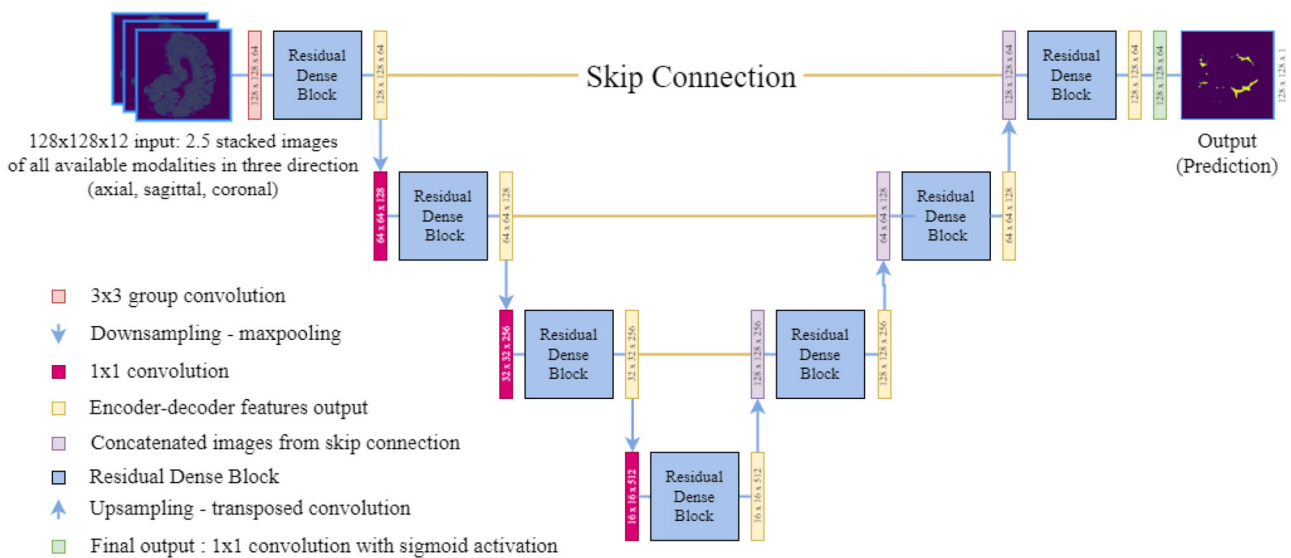


FIGURE 1 The structure of the proposed model with three encoders and three decoders which receives input of 2.5 stacked images of all modalities. The standard convolutions in each layer of the base U-Net are replaced with a residual dense block. In the last layer, the sigmoid is used to obtain final prediction.

3.1.2 | Residual dense block

The proposed model, supported by the incorporation of a residual dense block, is used in each encoder–decoder layer to replace the standard convolutional stack. So, instead of making the model deeper, it uses more convolution stacks through the residual dense block in each layer. The idea of the residual dense block is to combine and take advantage of the residual block and dense block, respectively, as part of ResNet¹¹ and DenseNet.¹² Although in the original work, the dense block is primarily designed and adapted from the residual block to overcome the vanishing gradient problem,¹² some past works have shown that these blocks are not mutually exclusive. Instead, they can combine them to make better image feature extraction.^{46–50}

As shown in Figure 2, a residual block is extended from standard convolutions operations, except it has a skip connection from the first to the last convolution. This skip connection combines information from the original information before the first convolution operation is performed to the result of the last convolution in each block using the addition operator. This connection enables the model to skip information from the convolution layers that do not contribute much to the performance.

In the dense block, the number of feature maps will be increased in every layer. The increase caused by the number of filters in each layer will be the concatenation of previous layers with the number of filter growth. So

the dense block retains the collective information as the filter since the first convolution operation will be carried to the end of the networks. Thus, for the input of the i -th layer is the result dense function H (convolution, batch normalization, and activation operation) of concatenated output from the previous convolution.

Compared to conventional convolutions, the number of weighted parameters in the dense block is lesser as the model learns only the weight of the new growth filters. Accordingly, a dense block with a skip connection is chosen for the proposed model to replace the standard convolutions block in the U-Net architecture. Each layer in the residual dense block comprises a 3×3 convolution with a dilation rate of 1. The result of each convolution is activated using ReLU followed by Batch Normalization.⁵¹ However, as the number of input channels on each block will differ from the number of channels in the last convolution, a 1×1 convolution operation is performed with the number of channels $n \times \text{channel growth } k$. This block is denoted as x_i given H is the dense block function and f is the 1×1 convolution mapping operation. $x_i = f(x_i) + H([x_0, \dots, x_{i-1}])$

Thus, regarding the parameter complexity, the proposed model has 12 685 953 parameters, which is significantly smaller than the original U-Net model (34 507 201 parameters). Some factors that make the model less complex are the layer's depth of only three and the incorporation of a residual dense block that generates the number of filters of each CNN operation linear to the growth rate. Compared to standard convolutional stacked, the number

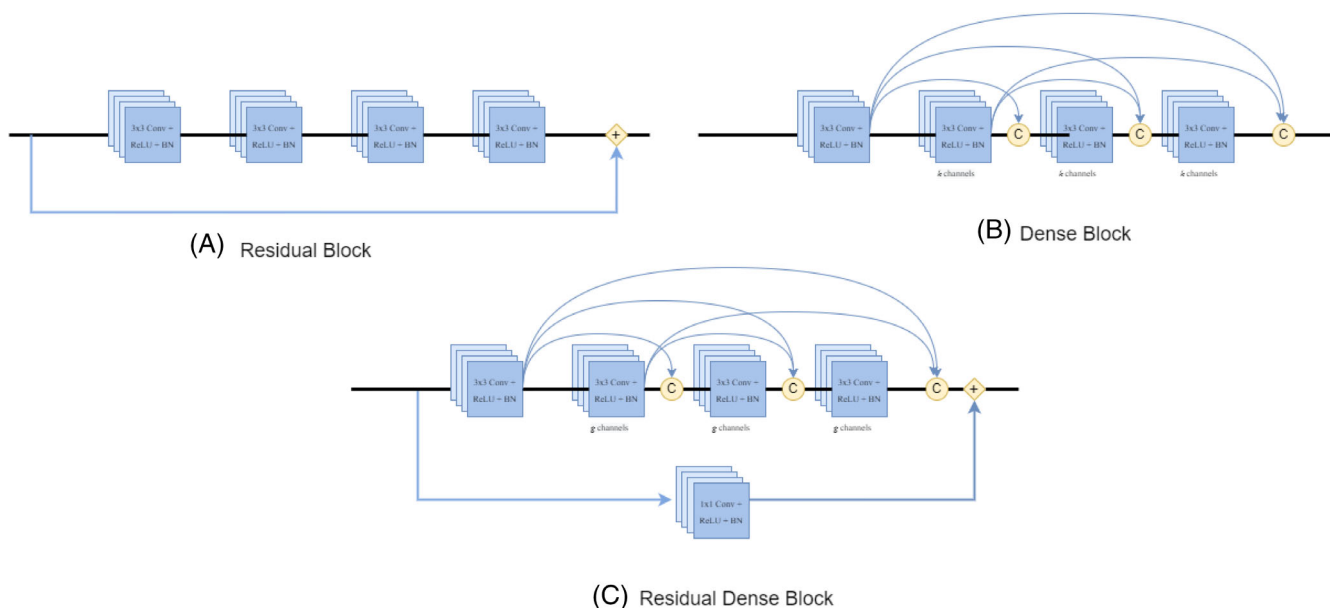


FIGURE 2 In the residual block (A), the features from the last convolution will be combined using the addition operator with the original input information. While in dense block (B), each convolution operation receives input from all previous convolutions operations. The residual dense block (C) is a combination of both blocks.

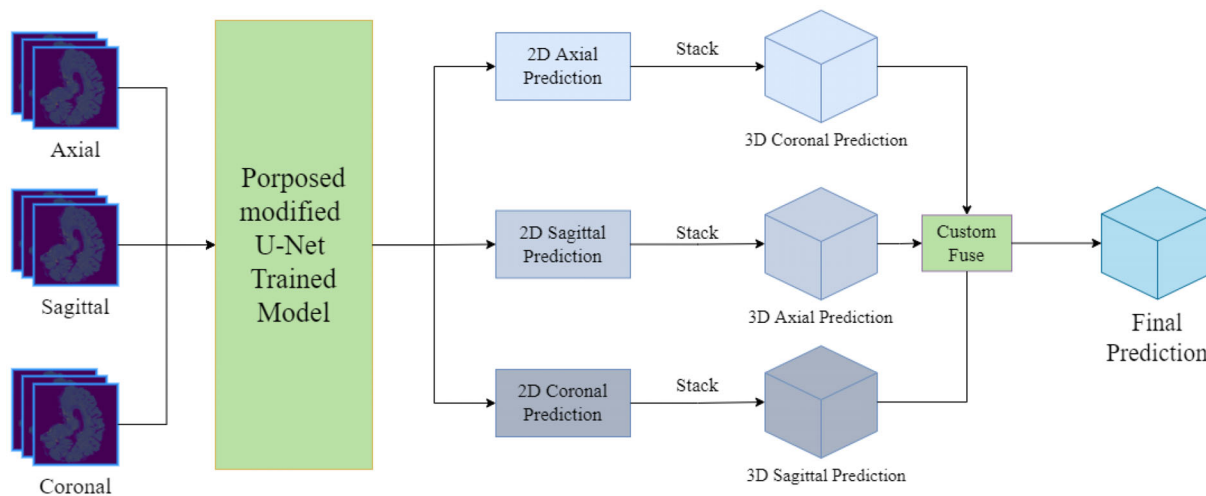


FIGURE 3 Complete scheme for 3D prediction reconstruction with the custom fusion (CF) technique. A single model used for predicting MS lesions with three different image axes. Then 2D slice images from all orientations are stacked and fused using the CF technique to reconstruct 3D prediction. MS, multiple sclerosis.

of filters from each convolutions operation will be multiplied based on the number of filters in the previous convolutions operation.

3.2 | Model final prediction

3.2.1 | 3D prediction reconstruction

Since the original 3D MRI image are sliced into 2D, it is essential to reconstruct them back into 3D. The model is also expected to predict the lesion based on three different image directions (axial, coronal, and sagittal). Prediction using all directions may increase the probability of detecting lesions, as each axis will expose different volumes of lesions.^{52,53} This scheme is inspired by Hitziger et al.⁵⁴ and Aslani et al.⁷ However, it is different in terms of how they merge predictions from all axes. The first used the average of the predicted probabilities and then compared it with three different fusion techniques (union, majority, and intersection), while the latter used the major voting method. Hence, this reconstruction method in this study is presented in Figure 3.

3.2.2 | Custom fusion

The CF strategy comes as an alternative to two simple operations (i.e., union and intersect) as used by past studies.^{7,29,34,54} This method is inspired by the calculation of lesion-wise metrics performance of two raters, as explained by Carass et al.⁵⁵ To better understand this technique, a fusion of two simple 1D binary predictions is presented in Figure 4.

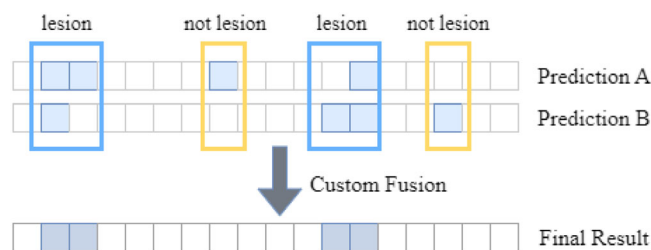


FIGURE 4 In custom lesion fusion with simple 1D array prediction, the lesion will be considered based on the same region.

Thus, instead of finding similarities between two predictions at the pixel level, the technique focuses on the agreement of lesions' region location. If two or more binary predictions agree with a particular lesion region regardless of the shape and size, it will be considered a lesion. In contrast, the nonoverlapping lesion region will not be considered a lesion as it might potentially become a false lesion.

The pseudo-code of the CF method is described in Algorithm 1. First, n different outputs are merged through a simple addition operation to obtain the overlapping pixels from P predictions resulting M . Then, iteratively, another non-overlapping pixel adjacent to each overlapping pixel in M is checked.

4 | EXPERIMENTS

4.1 | Data set

Primarily we evaluate the model on the ISBI2015 data set. However, in addition, for more comprehensive results, we also evaluate it on the OpenMS data set as the

Algorithm 1 Custom fusion

```

Date:  $P$ 
Result:  $O$ 
 $M \leftarrow$  element-wise addition of predictions in  $P$ ;
 $V \leftarrow$  min overlap pixels considered as a lesion;
 $X \leftarrow$  index of all elements  $M$  with value  $\geq 1$ ;
while  $True$  do
   $A \leftarrow$  elements of  $X$  with respective value in  $M < V$ ;
  if  $length\ A == 0$  then
    break;
  end
  for  $i$  in  $A$  do
     $p \leftarrow$  pixel  $i$  in  $M$ ;
     $d \leftarrow$  the distance  $p$  with any member of  $X$ ;
    if  $d < 2$  then
      set pixel  $p$  value with  $V$ ;
      append  $i$  to  $O$ 
    end
  end for
  if  $O < is\ not\ changed$  then
    break;
  end
end

```

secondary evaluation. Both contain 3D MRI images of MS lesion patients.

a. ISBI2015 data set

Four modalities are available in the data set: FLAIR, T2, PD, and MPRAGE (T1). The MRI images with dimensions of $181 \times 217 \times 181$. The characteristic of this data set is that the training size is significantly smaller than the testing. Given by the challenge, this data set is divided into training and testing containing MRI from 5 and 14 patients, respectively. Each patient has four to five observations. In total, the training set contains 84 3D MRI images with corresponding ground-truth masks from two experts. Figure 5 illustrated some sample data from the data set with the ground truths. While the testing data contains 265 3D MRI images without the ground truths (Figure 6).

b. OpenMS data set

The data set contains 3D MRI images from 30 patients with four available modalities: FLAIR, T2W, T1W, and T1WKS and each patient's data are supplemented with a single consensus of ground truth images. We use the pre-processed version of MRI which has dimensions of $154 \times 240 \times 240$.

4.2 | Data preprocessing

Due to the insufficiency of the training data, 3D MRI images will not be used as direct input to the model. Thus, image slicing is executed over all directions (axial, sagittal, and coronal) on all modalities and ground truth masks (Figure 6). Due to its different from all four available modalities. Some preprocessing tasks have already been performed in both data sets, such as skull stripping, N4 bias correction, and co-registration. Additional preprocessing follows what Zhang et al.²⁹ did, kernel density estimation is applied to normalize the MRI image. The intensity normalization of MRI images is the approach to improve the performance of the DL model.⁵⁶

As the lesions are primarily located in the brain area in the center of images, the surrounding black area in each slice is removed as it does not contribute much to the final segmentation. Thus, the 2D slices are cropped to 128×128 pixels. Also, since lesions do not always present in 2D slices, non lesion slices are discarded to reduce the imbalance class problem. The number of 2D slices from the data sets is shown on Table 2.

A set of 2D image slices is then prepared as input. The process follows Zhang et al., Roth et al.^{29,57} that use

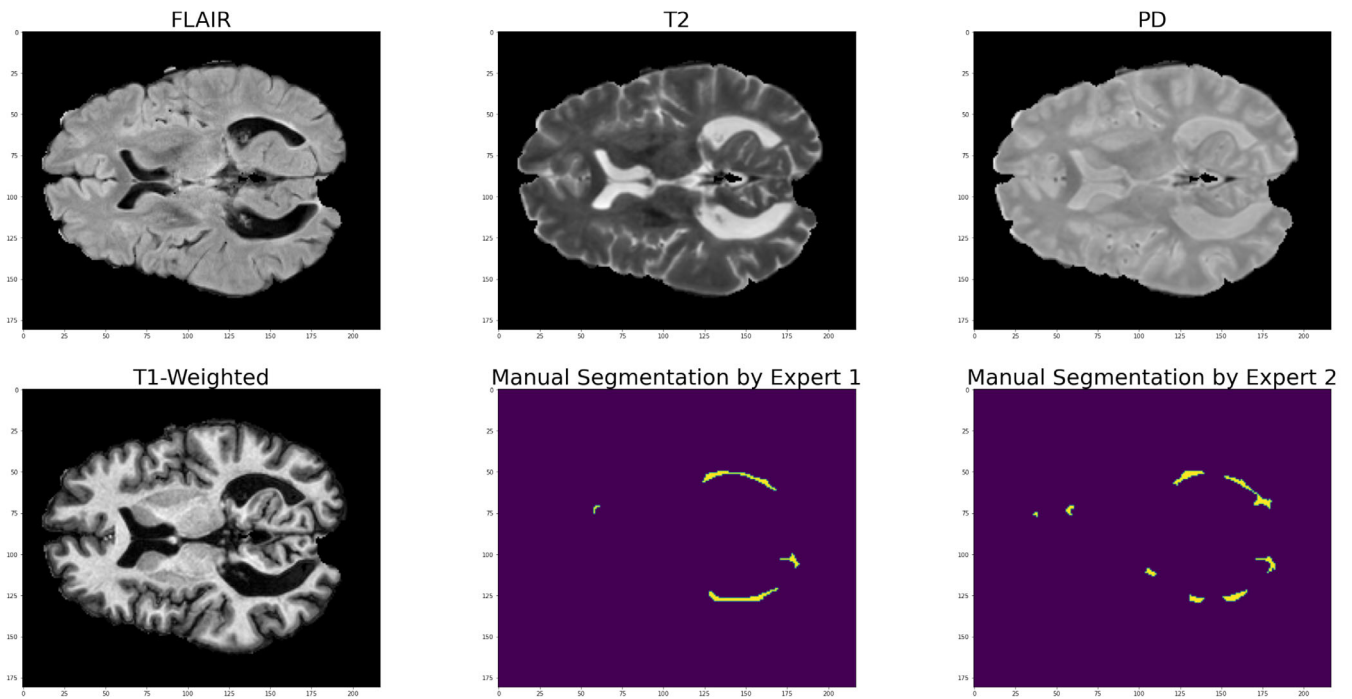


FIGURE 5 Sample MRI slice from the ISBI2015 data set displayed in 2D. Each subject has four image modalities (FLAIR, T2, PD, and MPRAGE). The ground truths from two experts are also provided in the training set. MRI, magnetic resonance imaging.

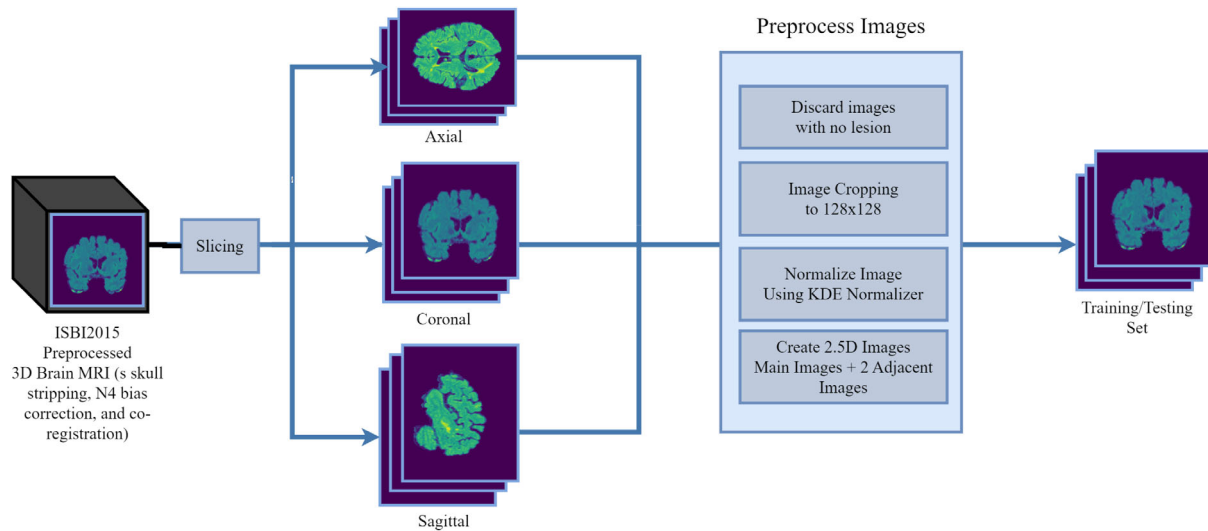


FIGURE 6 Data preprocessing steps from original 3D MRI images into 2D slices followed by image manipulation (eliminating image with no lesion, cropping, normalization, and 2.5D stacking). MRI, magnetic resonance imaging.

TABLE 2 Number of MRI 2D slices from dataset.

Data set	Train	Validation	Test
ISBI2015	4387	1096	143 895
OpenMS	5219	2610	2610

Abbreviation: MRI, magnetic resonance imaging.

2.5D stacked slices from all modalities of all directions as the input. A combination of different image modalities may increase the model's performance.⁷ A 2.5D stacked slice is defined as the 2D image slice plus two adjacent slices. It will allow the model to learn global context but with less computational cost.⁵⁷ Thus, for each input data,

TABLE 3 Several metrics used for evaluating model performance used in this MS lesion segmentation study as used in ISBI2015 evaluation platform.

Metrics	Formula
Dice Score (DSC)	$2TP/(2TP + FP + FN)$
Jaccard	$TP/(TP + FP + FN)$
PPV (Precision)	$TP/(TP + FP)$
TPR (Sensitivity)	$TP/(TP + FN)$
LTP (Lesion Wise TP)	Total overlapped lesions
LFP (Lesion Wise FP)	Total misidentified lesions
LTPR	$LTP/(LTP + LFN)$
LFPR	$LFP/(LFP + LTN)$
Vol. Difference (VD)	$(V_y - V_{gt})/V_{gt}$
Vol. Correlation (VC)	Average Pearson correlation of volume changes overtimes

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive.

there are 12 images (4 modalities \times 3 stacked slices). Hence the dimension of the input is $128 \times 128 \times 12$.

4.3 | Performance evaluation

To assess the model performance, the model implementation on the first data set was evaluated based on metrics defined in the ISBI2015 challenge⁵⁵; thus, it can be compared to another model that participated in the challenge. These metrics (Table 3) are constituted based on these metrics:

- True positive (TP): actual lesion (1) predicted as a lesion (1).
- False positive (FP): not a lesion (0) predicted as a lesion (1).
- False negative (FN): actual lesion (1), but not predicted as a lesion (0).
- True negative (TN): not a lesion (0) and not predicted as a lesion (0).
- V: total segmentation voxels.

For fair and objective evaluation, the performance result will be accomplished by submitting predictions to the ISBI2015 evaluation platform.¹ The platform also

calculates the overall score performance (Equation 1) based on the weighted sum of metrics in Table 3.

$$\text{Overall} = \frac{1}{2N} \sum^{2n} \left(\frac{\text{DSC}}{8} + \frac{\text{PPV}}{8} + \frac{1 - \text{LFPR}}{4} + \frac{\text{LTPR}}{4} + \frac{1 - \text{VC}}{4} \right) \quad (1)$$

Otherwise, for assessing model performance on the second data set, we perform an internal evaluation on the test data set based on more general metrics such as Dice Score, Jaccard, Precision, and Sensitivity.

4.4 | Model training

A total of 80% of the data in the training data set are used for fitting the proposed model and 20% for validation. This study does not consider using K-fold validation as it will prolong the training process. However, before dividing the data set into train and validation sets, stratified shuffling is performed based on the patient or observations to randomly distribute the data. The manually delineated lesion from the second rater is used as ground truth. The reason for this is that the second has more years of experience (10 years) than the first (4 years).⁵⁵ Also, since the sample images from the training set are limited, data augmentation techniques are performed in the training process to improve the model's robustness. The augmentation is executed by performing a random shift between 5 and 15 pixels vertically/horizontally, random rotation within a range of 15° , and a random horizontal flip.

$$\text{Dice Loss} = 1 - \text{DSC} = 1 - \frac{2 \times \text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (2)$$

To achieve model convergence, the Dice Loss (Equation 2), which is calculated from DSC,⁵⁸ is used as the objective function for the model training and evaluated during the training along with Adam to optimize gradient descent calculation as it is computationally efficient.⁵⁹ The training process will be executed for 200 epochs with an initial learning rate of 0.001 and batch size 8. The learning rate then will exponentially decay after reaching the 50th epoch. The validation loss is monitored during the training. The early learning stop strategy is also used to save training time if validation loss is not improving during 20 epochs in a row. The training is executed on NVidia A40 GPU using *The Multi-modal Australian ScienceS Imaging and Visualization Environment (MASSIVE)*.

4.5 | Ablation study

The proposed model is compared against the original U-Net model, U-Net model with residual block, and U-Net with dense block. All models use the same data processing and training configuration except for the model architecture and the number of filters, as summarized in Table 4. To see the effectiveness of the CF algorithm, the experiment will also compare this technique with other techniques (i.e., union, intersection, and voting as implemented in the past works^{29,34}). Given individual segmentation results Y_1, Y_2, Y_3 , the simple fusion function denoted as \mathcal{M} below:

1. Union: Consider all positively predicted pixels as lesions. $\mathcal{M}(Y_1, Y_2, Y_3) = Y_1 \cup Y_2 \cup Y_3$

2. Intersection: Consider a lesion if two or more predictions give positive predictions. $\mathcal{M}(Y_1, Y_2, Y_3) = Y_1 \cap Y_2 \cap Y_3$

3. Voting: Consider a lesion based on the majority prediction. $\mathcal{M}(Y_1, Y_2, Y_3) = Y_1 \cap Y_2 + Y_1 \cap Y_3 + Y_2 \cap Y_3 + Y_1 \cap Y_2 \cap Y_3$

TABLE 4 Parameter settings for ablation study.

Model	Filters	Convolutions
U-Net	[64, 128, 256, 512, 1024]	2 in encoder, 1 in decoder
U-Net with Residual Block	[64, 128, 256, 512, 1024]	2 in encoder, 1 in decoder
U-Net with Dense Block	[64, 128, 256, 512, 1024]	6 in encoder, 6 in decoder with Growth Rate 32
Proposed Model	[64, 128, 256, 512]	6 in encoder, 6 in decoder with Growth Rate 32

Note: Each model has different number of filters and stacked convolutions operations. The length of filters indicates the number of encoder-decoder in the architecture.

TABLE 5 Performance comparison of the proposed model and baseline models evaluated on ISBI2015.

Metrics	Basic U-Net	U-Net with Residual Block	U-Net with Dense Block	Proposed
Overall	92.777	92.792	92.861	93.034
DSC	0.663 ± 0.11	0.659 ± 0.12	0.655 ± 0.12	0.661 ± 0.12
Jaccard	0.506 ± 0.12	0.503 ± 0.13	0.498 ± 0.13	0.506 ± 0.13
PPV	0.806 ± 0.13	0.828 ± 0.12	0.832 ± 0.12	0.833 ± 0.12
TPR	0.602 ± 0.18	0.588 ± 0.19	0.579 ± 0.18	0.589 ± 0.19
LTPR	0.510 ± 0.21	0.477 ± 0.21	0.470 ± 0.21	0.496 ± 0.21
1-LFPR	0.880 ± 0.11	0.893 ± 0.11	0.898 ± 0.10	0.903 ± 0.10
1-VD	0.640 ± 0.18	0.625 ± 0.18	0.623 ± 0.18	0.624 ± 0.19
Params	34 507 201	35 911 361	45 004 065	12 685 953

Note: Presented in the table are the average score and standard deviation of all subjects in the testing data set.

5 | RESULT AND DISCUSSION

5.1 | Proposed model evaluation

Table 5 shows the performance and model parameters of the proposed and baseline models. The proposed model successfully got an overall score of 93.034% and surpassed the original U-Net model (92.777%). It is also better than when using U-Net with the residual block (92.792%) and U-Net with the dense block (92.861%).

The superiority of U-Net is confirmed in this experiment. Although it has the lowest overall score, it is still performing well when looking at some individual metrics. The main drawback of basic U-Net is the number of parameters is relatively high as it performs convolution operations with a large number of filters which is potentially redundant.

When incorporating residual blocks, the segmentation performance is then slightly increased. The skip connection in the residual has successfully eliminated the vanishing gradients problems in the network and increased individual metrics like PPV and LFPR. On the other hand, the employment of dense blocks in the U-Net also increases these individual metrics but with fewer parameters in the model. Finally, the incorporation of dense blocks and residual blocks together achieves the best result. Hence, there are two important things about the proposed model that can be highlighted:

a. Performance improvement

The proposed model successfully increases the model's overall score by significantly improving PPV and 1-LFPR without big reduction to other metrics. The reason is the incorporation of residual blocks and dense blocks at the same time. The residual blocks play important role to eliminate useless features in the network, while the dense blocks retain the collective information

during the convolution operations from the first layer to the last layer.

b. Reduce model complexity

The proposed model is more efficient as it has less number of parameters. The reason is that incorporating dense blocks in U-Net is better than the residual block, although the number of parameters is not expected to be higher and contradicts the original idea of DenseNet.³⁸ The layer depth triggers this excessive number of parameters as the original U-Net has and the large channel growing rate.

Visually, the sample of segmentations presented in Figure 7 shows the input, predictions, ground truth, GradCAM,⁶⁰ and saliency map.⁶¹ GradCAM and saliency map validate the correctness of the lesion location in association with the input modalities and can clearly emphasize the lesion (indicated as red in GradCAM).

5.2 | CF evaluation

The comparison of CF performance with the individual axis is presented in Table 6. Consistent with past

works,^{52,62} using individual axis, prediction based on axial input is better than sagittal and coronal. This study also confirms that the CF method leads to better performance than using only individual axis for prediction. As shown in Table 6, no individual axis prediction gets overall score above 92%.

Otherwise, against other fusion techniques, as presented in Table 7, the CF also yields better results by achieving an overall score of 93.034%. It can increase DSC (0.661), Jaccard (0.506), VC (0.867), and retain the 1-LFPR (0.903). Generally, predictions fusion using common techniques is somewhat better than individual predictions, as presented in Table 5. For example, although the overall score is slightly lower, the union successfully increases the LTPR and can increase DSC (0.605). It confirms past work where the union method can increase DSC when combining two predictions.³⁴ However, as it tries to detect as many lesions as possible, the possibility of getting FP increases (reflected in lowest PPV and LFPR) (Table 6).

In contrast, fusion using intersection successfully increases PPV and 1-LFPR significantly (0.944 and 0.607), but it also decreases the TP as it will eliminate

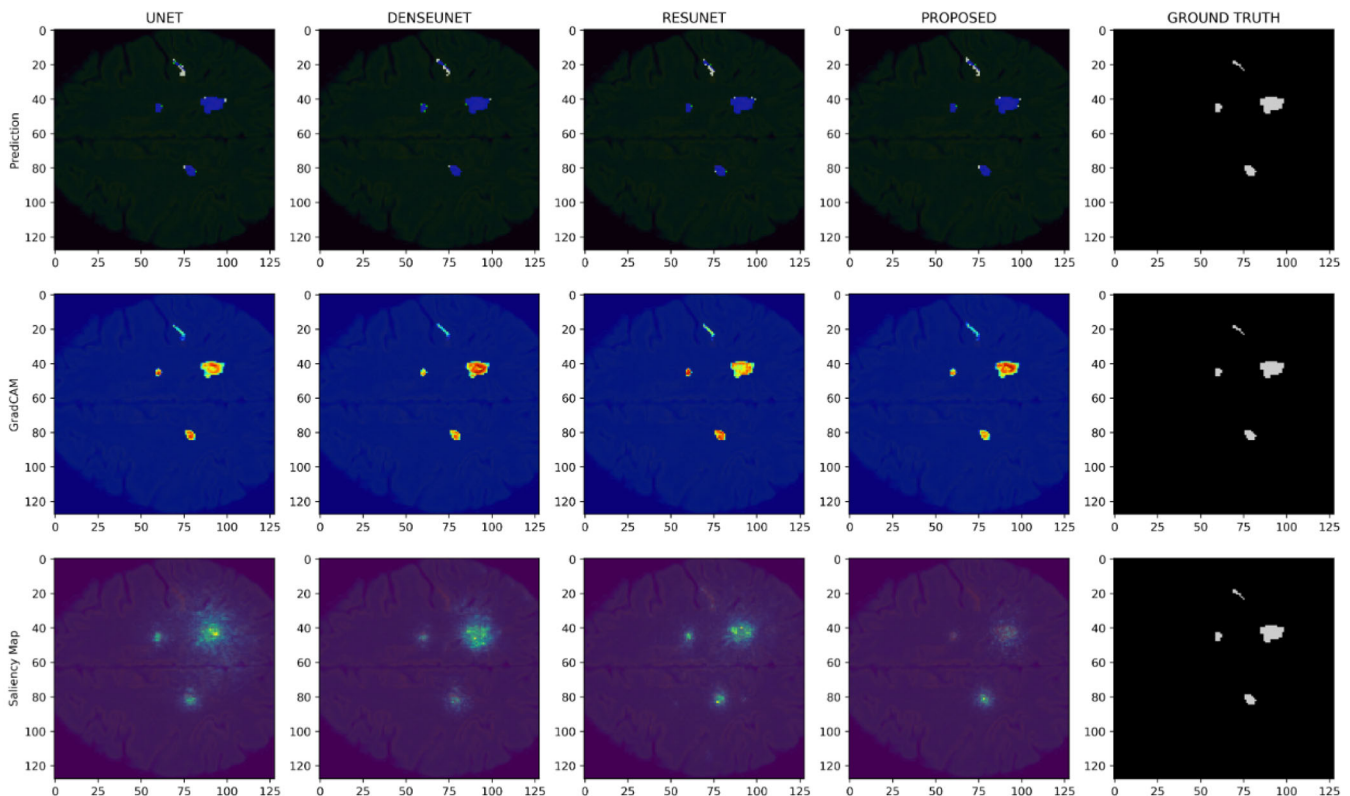


FIGURE 7 Sample lesion prediction, GradCAM, and Saliency Map result from baseline models and the proposed model. From left to right are predictions from each model U-Net, Dense U-Net, Residual-U-Net, and the proposed model. In the images, the blue pixels indicate TP, the green pixels indicate FN, and the white pixels indicate FP. FN, false negative; FP, false positive; TP, true positive.

Metrics	Axial	Sagittal	Coronal	CF
Overall	91.985	91.316	91.883	93.034
DSC	0.623 ± 0.13	0.598 ± 0.14	0.604 ± 0.14	0.661 ± 0.12
Jaccard	0.466 ± 0.13	0.441 ± 0.14	0.447 ± 0.14	0.506 ± 0.13
PPV	0.854 ± 0.12	0.815 ± 0.14	0.849 ± 0.12	0.833 ± 0.12
TPR	0.523 ± 0.18	0.503 ± 0.17	0.500 ± 0.18	0.589 ± 0.19
LTPR	0.581 ± 0.17	0.550 ± 0.20	0.550 ± 0.21	0.496 ± 0.21
1-LFPR	0.690 ± 0.18	0.666 ± 0.17	0.719 ± 0.16	0.903 ± 0.10
1-VD	0.588 ± 0.20	0.582 ± 0.19	0.573 ± 0.20	0.422 ± 0.22

Note: The table presents the average score and standard deviation of all subjects in the testing data set.

Metrics	Union	Intersection	Voting	CF
Overall	90.105	92.785	92.816	93.034
DSC	0.650 ± 0.12	0.542 ± 0.16	0.621 ± 0.14	0.661 ± 0.12
Jaccard	0.494 ± 0.13	0.389 ± 0.15	0.465 ± 0.14	0.506 ± 0.13
PPV	0.759 ± 0.15	0.944 ± 0.63	0.894 ± 0.96	0.833 ± 0.12
TPR	0.408 ± 0.18	0.607 ± 0.17	0.508 ± 0.18	0.589 ± 0.19
LTPR	0.638 ± 0.21	0.471 ± 0.21	0.556 ± 0.22	0.496 ± 0.21
1-LFPR	0.535 ± 0.18	0.903 ± 0.85	0.805 ± 0.14	0.903 ± 0.10
1-VD	0.663 ± 0.21	0.625 ± 0.18	0.554 ± 0.20	0.422 ± 0.22
Runtime	20.3 ms	19.4 ms	19.6 ms	861 ms

Note: The table presents the average score and standard deviation of all subjects in the testing data set.

non-overlapping pixels of lesions. In comparison, the voting technique balances these methods by increasing these performance metrics (i.e., PPV and LFPR) without getting significant losses in individual aspects by obtaining a score of 92.815%. The CF technique comes as the most conservative one. It successfully increases lesion detection by carefully selecting the predictions from all input images with different orientations. The qualitative evaluation for these techniques is presented in Figure 8. The figure confirms that the CF technique can significantly eliminate the chance of getting FP and FN.

Instead of fusing all individual axis predictions, this technique also evaluated against the combinations of two axes predictions (i.e., axial + sagittal, sagittal + coronal, and axial + coronal). As can be seen in Table 8, the combination of axial + sagittal predictions yields the best results and slightly higher than when using all axes. However, to get a fused prediction, both using two axes or all axes, it is worth to be addressed that the prediction will unquestionably take longer than a single prediction. This is because it needs three single predictions to be inferred first and then spends more runtime to execute the algorithm. Furthermore, since the CF performs iterative checking to find the lesion regions, thus the execution time is slower (see Table 7).

TABLE 6 Performance comparison of the custom fusion (CF) and different individual axis predictions.

TABLE 7 Performance comparison of custom fusion (CF) technique to others techniques using all input axes.

5.3 | Comparison to the state-of-art

For consistency and fairness, the proposed model is also compared with other state-of-art models based on the result reported on the ISBI2015 website (Table 9). Although the proposed model's overall score is not the best; however, it significantly outperformed other models in the aspect of LFPR, which contributes much to the overall score.

The proposed models are also comparable to other models as it is not significantly different in terms of DSC. The DSC of the proposed model is closer to the performance of the Tiramisu model²⁹ and nnUnet⁶³ but with a lower standard deviation score (0.12). Moreover, for TPR, the proposed model also shows a promising result: it is the second best with a score of 0.589. Unfortunately, since there is no information about the number of parameters used in those models, their complexity cannot be compared.

5.4 | Evaluation on second data set

Table 10 presents the performance of our model on the second data set. When trained on this data set, the model achieves a DSC of 0.857 for axial slices, along with strong

FIGURE 8 Example analysis on various fusions using prediction from the proposed model. The blue, red, and yellow pixels are TP, FN, and FP, respectively.

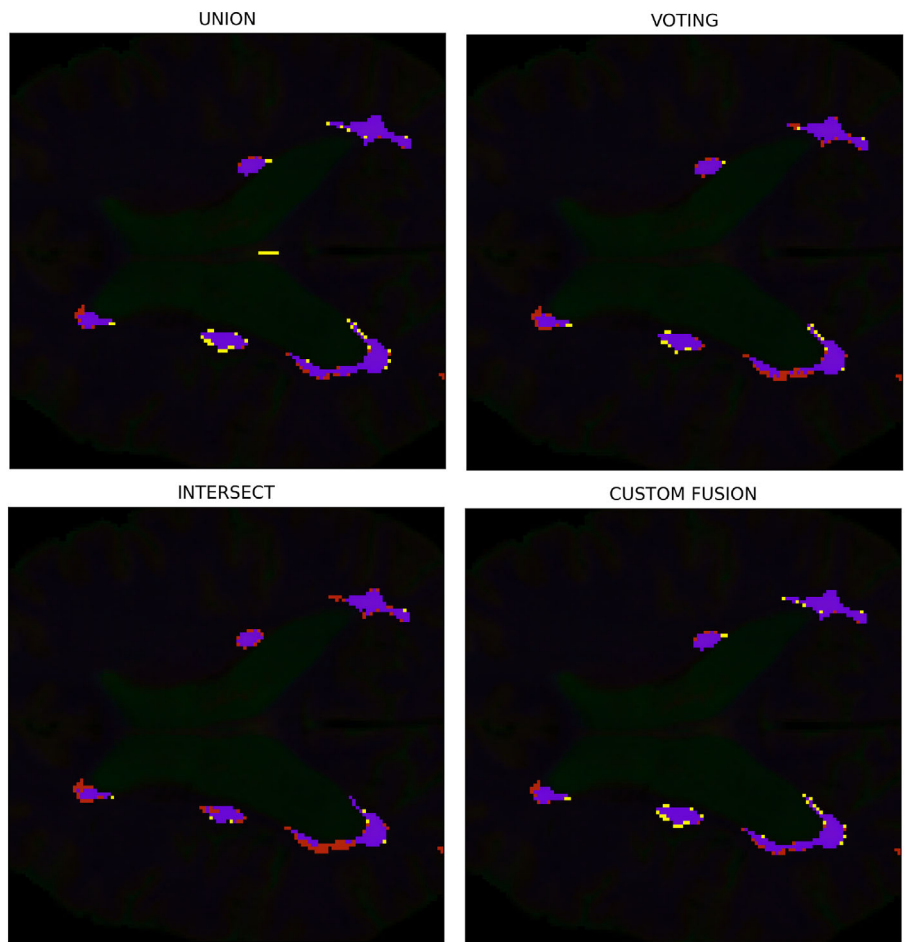


TABLE 8 Custom fusion performance comparison using all input axes and only two axes.

Metrics	Axial + saggital	Saggital + coronal	Axial + coronal	All axes
Overall	93.090	92.841	92.903	93.034
DSC	0.642 ± 0.13	0.649 ± 0.13	0.65 ± 0.13	0.661 ± 0.12
Jaccard	0.500 ± 0.13	0.485 ± 0.13	0.500 ± 0.13	0.506 ± 0.13
PPV	0.848 ± 0.12	0.843 ± 0.12	0.844 ± 0.12	0.833 ± 0.12
TPR	0.574 ± 0.19	0.556 ± 0.19	0.564 ± 0.19	0.589 ± 0.19
LTPR	0.521 ± 0.22	0.507 ± 0.22	0.526 ± 0.22	0.496 ± 0.21
1-LFPR	0.882 ± 0.11	0.873 ± 0.11	0.857 ± 0.12	0.903 ± 0.10
1-VD	0.385 ± 0.20	0.607 ± 0.20	0.615 ± 0.20	0.422 ± 0.22

Note: The table presents the average score and standard deviation of all subjects in the testing data set.

scores in other key metrics, such as precision and sensitivity. These encouraging results not only validate the model's robustness but also suggest its potential for broader implementation across different data sets and disease contexts.

A notable limitation of our study is that it does not employ a transfer learning-based approach. Instead, we preprocess and retrain the model on the second data set. In future research, a more optimal strategy for achieving model generalization should be pursued.

6 | CONCLUSION

In this study, we present a modified 2D U-Net model for MS lesion segmentation, incorporating residual dense blocks and a CF technique to merge diverse predictions. The model was evaluated on both the ISBI2015 and OpenMS data sets, achieving a top overall score of 93.09% on ISBI2015 and DSC of 0.857 on the OpenMS data set. Future research will explore the application of our model for segmenting other diseases, such as lung cancer and

TABLE 9 The proposed model performance compared to other state-of-art models reported on ISBI2015 evaluation platform.

Model	Score	DSC	Jaccard	PPV	TPR	1-LFPR	LTPR
Tiramisu ²⁹	93.358	0.668 ± 0.14	0.519 ± 0.15	0.886 ± 0.10	0.575 ± 2.01	0.875 ± 0.12	0.537 ± 0.23
All-NET ²⁴	93.317	0.636 ± 0.15	0.485 ± 0.16	0.916 ± 0.87	0.524 ± 2.03	0.883 ± 0.12	0.527 ± 0.23
nnUNet ⁶³	93.087	0.686 ± 0.13	0.537 ± 0.14	0.849 ± 0.11	0.614 ± 0.19	0.832 ± 0.14	0.553 ± 0.23
Ours-all axes	93.034	0.661 ± 0.12	0.506 ± 0.13	0.833 ± 0.12	0.589 ± 0.19	0.903 ± 0.10	0.496 ± 0.21
Ours-Ax. + Sag.	93.090	0.642 ± 0.12	0.500 ± 0.13	0.848 ± 0.12	0.574 ± 0.19	0.882 ± 0.11	0.512 ± 0.22
DLB ³⁷	92.849	0.646 ± 0.14	0.494 ± 0.15	0.888 ± 0.10	0.486 ± 2.02	0.869 ± 0.11	0.486 ± 0.22
MMU ⁵⁴	92.661	0.641 ± 0.14	0.487 ± 0.14	0.852 ± 0.12	0.559 ± 2.07	0.803 ± 0.14	0.550 ± 0.27
Longitudinal ²¹	90.070	0.627 ± 0.14	0.471 ± 0.141	0.789 ± 0.154	0.555 ± 0.181	0.502 ± 0.177	0.568 ± 0.23

Note: The table presents the average score and standard deviation of all subjects in the testing data set.

TABLE 10 Model performance on the second data set measured in terms of DSC, precision, and sensitivity.

Metrics	Axial	Sagittal	Coronal
TP	71 453	67 295	64 107
FP	9414	11 247	9527
FN	13 209	14 671	13 207
DSC	0.857	0.838	0.847
Jaccard	0.878	0.859	0.867
Precision	0.883	0.856	0.870
Sensitivity	0.843	0.821	0.829

brain tumors, while employing transfer learning methods to enhance the model's generalizability.

Our CF approach for combining prediction masks demonstrated superior performance compared to other methods, increasing the number of TPs without raising the number of FPs. Nevertheless, two limitations were identified: the study did not compare the CF technique with other advanced fusion methods, and the execution time was longer than that of simpler techniques.

ACKNOWLEDGMENTS

This work was supported by the MASSIVE HPC facilities (www.massive.org.au). We also would like thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions. Open access publishing facilitated by Monash University, as part of the Wiley - Monash University agreement via the Council of Australian University Librarians.

DATA AVAILABILITY STATEMENT

MS Lesion data used in this study are available at <https://smart-stats-tools.org/lesion-challenge-2015> and https://github.com/muschellij2/open_ms_data.

ORCID

Bakhtiar Amaludin  <https://orcid.org/0000-0001-7607-0193>

Seifedine Kadry  <https://orcid.org/0000-0002-1939-4842>

Fung Fung Ting  <https://orcid.org/0000-0001-9855-6020>

David Taniar  <https://orcid.org/0000-0002-8862-3960>

ENDNOTE

¹ <https://smart-stats-tools.org/>

REFERENCES

- Walton C, King R, Rechtman L, et al. Rising prevalence of multiple sclerosis worldwide: insights from the Atlas of MS, third edition. *Mult Scler J*. 2020;26(14):1816-1821.
- Miller JR. The importance of early diagnosis of multiple sclerosis. *J Manag Care Pharm*. 2004;10(3 Suppl B):S4-S11.
- Waubant E. Improving outcomes in multiple sclerosis through early diagnosis and effective management. *Prim Care Companion CNS Disord*. 2012;14(5):27363.
- Compston A, Coles A. Multiple sclerosis. *Lancet*. 2008;372:1502-1517.
- Ansari SU, Javed K, Qaisar SM, Jillani R, Haider U. Multiple sclerosis lesion segmentation in brain MRI using inception modules embedded in a convolutional neural network. *J Healthc Eng*. 2021;2021:4138137.
- Danelakis A, Theoharis T, Verganelakis DA. Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. *Comput Med Imaging Graph*. 2018;70:83-100. doi:10.1016/j.compmedimag.2018.10.002
- Aslani S, Dayan M, Storelli L, et al. Multi-branch convolutional neural network for multiple sclerosis lesion segmentation. *Neuroimage*. 2019;196:1-15.
- Sahraian MA, Radue EW. *MRI Atlas of MS Lesions*. Springer Science & Business Media; 2007.
- Roy S, Butman JA, Reich DS, Calabresi PA, Pham DL. Multiple sclerosis lesion segmentation from brain MRI via fully convolutional neural networks. ArXiv. 2018:abs/1803.09172.
- Zeng C, Gu L, Liu Z, Zhao S. Review of deep learning approaches for the segmentation of multiple sclerosis lesions on brain MRI. *Front Neuroinform*. 2020;14:610967.

11. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016:770-778.
12. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017: 2261-2269.
13. Kaur A, Kaur L, Singh A. State-of-the-art segmentation techniques and future directions for multiple sclerosis brain lesions. *Arch Comput Methods Eng*. 2021;28:951-977.
14. Sah M, Direkoglu C. A survey of deep learning methods for multiple sclerosis identification using brain MRI images. *Neural Comput Appl*. 2022;34:7349-7373.
15. Khan M, Khan A, Alhaisoni M, et al. Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm. *Int J Imaging Syst Technol*. 2023;33:572-587.
16. Behera TK, Khan MA, Bakshi S. Brain MR image classification using superpixel-based deep transfer learning. *IEEE Journal of Biomedical and Health Informatics*. 2022;1-11. <https://ieeexplore.ieee.org/document/9926058/>
17. Masood M, Maham R, Javed A, Tariq U, Khan MA, Kadry S. Brain MRI analysis using deep neural network for medical of internet things applications. *Comput Electr Eng*. 2022;103: 108386. <https://www.sciencedirect.com/science/article/pii/S0045790622006036>
18. Zhu Z, Khan MA, Wang S-H, Zhang Y-D. RBEBT: a ResNet-based BA-ELM for brain tumor classification. *Comput Mater Contin*. 2023;74(1):101-111. <http://www.techscience.com/cm/v74n1/49800>
19. Zahid U, Ashraf I, Khan MA, et al. BrainNet: optimal deep learning feature fusion for brain tumor classification. *Comput Intell Neurosci*. 2022;2022:1-13.
20. Valverde S, Cabezas M, Roura E, et al. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *Neuroimage*. 2017;155: 159-168.
21. Birenbaum A, Greenspan H. Multi-view longitudinal CNN for multiple sclerosis lesion segmentation. *Eng Appl Artif Intel*. 2017;65:111-118.
22. Afzal HMR, Luo S, Ramadan S, et al. Automatic and robust segmentation of multiple sclerosis lesions with convolutional neural networks. *Comput Mater Contin*. 2020;66:977-991.
23. Yildirim M, Dandil E. Automated Multiple Sclerosis Lesion Segmentation on MR Images Via Mask R-CNN; 2021:570-577.
24. Zhang H, Zhang J, Li C, et al. ALL-Net: anatomical information lesion-wise loss function integrated into neural network for multiple sclerosis lesion segmentation. *NeuroImage Clin*. 2021;32:102854. <https://www.sciencedirect.com/science/article/pii/S2213158221002989>
25. Gessert N, Krüger J, Opfer R, et al. Multiple sclerosis lesion activity segmentation with attention-guided two-path CNNs. *Comput Med Imaging Graph*. 2020;84:101772.
26. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. 2017;36:61-78. <https://www.sciencedirect.com/science/article/pii/S1361841516301839>
27. Essa E, Aldesouky D, Hussein S, Rashad M. Neuro-fuzzy patch-wise R-CNN for multiple sclerosis segmentation. *Med Biol Eng Comput*. 2020;58:2161-2175.
28. Chen Z, Wang X, Huang J, Lu J, Zheng J. Deep attention and graphical neural network for multiple sclerosis lesion segmentation from MR imaging sequences. *IEEE J Biomed Health Inform*. 2022;26(3):1196-1207.
29. Zhang H, Valcarcel A, Bakshi R, et al. Multiple sclerosis lesion segmentation with tiramisu and 2.5D stacked slices. *Med Image Comput Comput Assist Interv*. 2019;11766:338-346.
30. Kumar A, Murthy OS, Ghosal P, Mukherjee A, Nandi D. A Dense U-Net Architecture for Multiple Sclerosis Lesion Segmentation; 2019.
31. Salehi SS, Erdogmus D, Gholipour A. Tversky Loss Function For Image Segmentation Using 3D Fully Convolutional Deep Networks; 2017.
32. Hashemi SR, Salehi SSM, Erdoğan D, Prabhu SP, Warfield S, Gholipour A. Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: application to multiple sclerosis lesion detection. *IEEE Access*. 2019;7:1721-1735.
33. Hu C, Kang G, Hou B, Ma Y, Labeau F, Su Z. Acu-Net: a 3D attention context U-net for multiple sclerosis lesion segmentation. In: ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2020: 1384-1388.
34. Hashemi M, Akhbari M, Jutten C. Delve into multiple sclerosis (MS) lesion exploration: a modified attention U-Net for MS lesion segmentation in brain MRI. *Comput Biol Med*. 2022;145: 105402. doi:10.1016/j.combiomed.2022.105402
35. Hou B, Kang G, Xu X, Hu C. Cross attention densely connected networks for multiple sclerosis lesion segmentation. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2019:2356-2361.
36. Aljamaat A, NikravanShalmani A, Bayat P. Multiple sclerosis identification in brain MRI images using wavelet convolutional neural networks. *Int J Imaging Syst Technol*. 2021;31(2):778-785. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ima.22492>
37. Kamraoui R, Ta VT, Tourdias T, Mansencal B, Phd J, Coupé P. DeepLesionBrain: towards a broader deep-learning generalization for multiple sclerosis lesion segmentation. *Med Image Anal*. 2021;76:102312.
38. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN; 2017.
39. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation; 2015:3431-3440.
40. Brosch T, Tang L, Yoo Y, Li D, Traboulsee A, Tam R. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Trans Med Imaging*. 2016;35:1229-1239.
41. Jegou S, Drozdal M, Vázquez D, Romero A, Bengio Y. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation; 2017:1175-1183.
42. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation; vol. 9351; 2015: 234-241.
43. Siddique N, Sidike P, Elkin C, Devabhaktuni V. U-Net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access*. 2021;9:82031-82057.
44. Zhang H, Zhang J, Zhang Q, et al. RSANet: Recurrent Slice-wise Attention Network for Multiple Sclerosis Lesion Segmentation; 2020.
45. Feng Y, Pan H, Meyer C, Feng X. A Self-Adaptive Network For Multiple Sclerosis Lesion Segmentation From Multi-Contrast MRI With Various Imaging Protocols; 2018.
46. Rabbi J, Ray N, Schubert M, Chowdhury S, Chao D. Small-object detection in remote sensing images with end-to-end

- edge-enhanced GAN and object detector network. *Remote Sens.* 2020;12:1432.
47. Wang H, Su D, Liu C, Jin L, Sun X, Peng X. Deformable non-local network for video super-resolution. *IEEE Access.* 2019;7:177734-177744.
 48. Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image restoration. *IEEE Trans Pattern Anal Mach Intell.* 2021;43(7):2480-2495.
 49. Zhang Z, Tang Z, Wang Y, et al. Dense residual network: enhancing global dense feature flow for character recognition. *Neural Netw.* 2021;139:77-85. <https://www.sciencedirect.com/science/article/pii/S0893608021000472>
 50. Zhao X, Zhang J, Tian J, Zhuo L, Zhang J. Residual dense network based on channel-spatial attention for the scene classification of a high-resolution remote sensing image. *Remote Sens.* 2020;12:1887.
 51. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (PMLR); 2015:448-456.
 52. Galler S, Stellmann JP, Young K, et al. Improved lesion detection by using axial T2-weighted MRI with full spinal cord coverage in multiple sclerosis. *Am J Neuroradiol.* 2016;37(5):963-969.
 53. Wilms G, Marchal G, Kersschot E, et al. Axial vs sagittal T2-weighted brain MR images in the evaluation of multiple sclerosis. *J Comput Assist Tomogr.* 1991;15(3):359-364.
 54. Hitziger S, Ling WX, Fritz T, D'Albis T, Lemke A, Grilo J. Triplanar U-Net with lesion-wise voting for the segmentation of new lesions on longitudinal MRI studies. *Front Neurosci.* 2022;16:964250. <https://www.frontiersin.org/articles/10.3389/fnins.2022.964250>
 55. Carass A, Roy S, Jog A, et al. Longitudinal multiple sclerosis lesion segmentation: resource and challenge. *Neuroimage.* 2017;148:77-102. <https://www.sciencedirect.com/science/article/pii/S1053811916307819>
 56. Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the impact of intensity normalization on MR image synthesis. *Medical Imaging 2019: Image Processing.* Vol 10949. International Society for Optics and Photonics; 2019:109493H.
 57. Roth H, Lu L, Seff A, et al. A New 25D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations, vol. 17; 2014.
 58. Dice LR. 8. Ecology; 1945;26:297-302.
 59. Kingma D, Ba J. Adam: a method for stochastic optimization. International Conference on Learning Representations; 2014.
 60. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV); 2017:618-626.
 61. Simonyan K, Vedaldi A, Zisserman A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps; preprint 2013 12.
 62. Weier K, Mazraeh J, Naegelin Y, et al. Biplanar MRI for the assessment of the spinal cord in multiple sclerosis. *Mult Scler J.* 2012;18(11):1560-1569.
 63. Isensee F, Jaeger P, Kohl S, Petersen J, Maier-Hein K. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2021;18:1-9.

How to cite this article: Amaludin B, Kadry S, Ting FF, Taniar D. Toward more accurate diagnosis of multiple sclerosis: Automated lesion segmentation in brain magnetic resonance image using modified U-Net model. *Int J Imaging Syst Technol.* 2024;34(1):e22941. doi:10.1002/ima.22941