Received: 28 May 2024 | Revised: 30 July 2024 | Accepted: 17 October 2024 | Published online: 24 October 2024

RESEARCH ARTICLE

Medinformatics
2025, Vol. 2(1) 36–48
DOI: 10.47852/bonviewMEDIN42023518

BON VIEW PUBLISHING

VSegNet – A Variant SegNet for Improving Segmentation Accuracy in Medical Images with Class Imbalance and Limited Data

Iyyakutty Dheivya[1] and Gurunathan Saravana Kumar[1,*]

[1]*Department of Engineering Design, Indian Institute of Technology Madras, India*

Abstract: Deep learning methods for many medical image segmentation task encounter challenges like smaller datasets and class imbalance. This study proposes a variant SegNet (vSegNet) designed to deliver significantly accurate and reliable segmentation results on such datasets. The novelty lies in designing encoder and decoder blocks with an appropriate number of convolution layers and using the Dice score and Hausdorff distance (HD) as compound loss function in learning. This study used public datasets consisting of chest X-rays, axial CT slices, foot ulcer images, and subset of SPIDER dataset to benchmark the segmentation task of the proposed neural network model with other popular networks like U-Net, SegNet, DeepLabv3+, VGG16, MobileNetV2, and fully convolutional network (FCN). For the segmentation of lungs in chest X-rays, vertebral body in CT, augmented data for the previous case, foot ulcer dataset, and segmentation of vertebrae, intervertebral disks, and spinal canal in SPIDER dataset (MRI dataset) respectively, the proposed vSegNet performed with a Dice score of 0.96 ± 0.01, 0.90 ± 0.20, 0.95 ± 0.02, 0.86 ± 0.07, and 0.95 ± 0.01 and the HD of 14.33 ± 7.74, 8.45 ± 7.08, 7.99 ± 6.05, 29.32 ± 25.64, and 8.45 ± 2.81 with respect to the ground truth on the test dataset. These results highlight the effectiveness of the proposed model in delivering both higher segmentation accuracy and improved boundary delineation. The proposed network, vSegNet, has been demonstrated as an effective model for semantic segmentation on class-imbalanced smaller datasets, surpassing all other networks considered in this study in terms of mIoU, BF score, Dice score, HD, accuracy, precision, recall, and $F1$ score on a variety of anatomical regions and medical imaging modalities.

Keywords: deep neural network, semantic segmentation, Dice score, Hausdorff distance, compound loss function

1. Introduction

Medical image segmentation is one of the essential steps for accurate diagnosis and treatment planning, yet it faces several challenges like class imbalance and limited data to name a few. For instance, segmenting the foreground of different organs from medical images often involves identifying smaller or less common structures compared to larger, more prevalent ones. This imbalance leads to models that may underperform in detecting these less common regions. Additionally, creating large annotated datasets is both challenging and time-consuming. Addressing class imbalance and data scarcity is crucial for developing robust segmentation models and improving diagnostic precision. However, traditional handcrafted methods face challenges when applied on complex segmentation tasks on images with varied illumination, making them less adaptable to different applications [1, 2]. Neural networks for image segmentation evolved from U-Net to DeepLab and Mask R-CNN [3–5]. However, most architectures are developed and evaluated on large datasets, overlooking the challenges of smaller databases that commonly occur in many medical image segmentation task. To address this, approaches like transfer learning and domain adaptation are crucial [6–8].

Decoder structures have improved deep neural network performance on new datasets with fewer training samples [9–11]. Preprocessing techniques can enhance image analysis, but segmentation of smaller and infrequent objects remains challenging. Atrous convolution and spatial pyramid pooling can extract features for segmentation from image classification models [12, 13]. Current medical image segmentation methods using deep learning face challenges with organ deformations, weak edges, and limited adaptability to diverse-scale regions [14–16]. Self-supervised models encounter difficulties with inhomogeneous backgrounds and distinguishing distinct regions within medical images [17]. Approaches like VLUU and DeepLABNet address partial labels and stability but can cause over-segmentation [18, 19]. These methods may also suffer from limitations such as blurry boundaries, over-segmentation, prolonged training time, and reliance on prior knowledge [20–23]. Due to class imbalance, traditional loss functions like cross-entropy are not enough for medical image segmentation [24–26]. Compound loss functions (CL) are necessary to optimize the model's performance across various objectives [25, 27, 28]. Focal loss [29] can mitigate the impact of dominant classes, but small databases remain a challenge.

Despite advancements in image segmentation, existing methods struggle with complexities such as small datasets, class imbalance, and issues like blurry boundaries and over-segmentation as revealed by the comprehensive literature review. Many methods often rely on large datasets and may not

*Corresponding author: Gurunathan Saravana Kumar, Department of Engineering Design, Indian Institute of Technology Madras, India. Email: gsaravana@iitm.ac.in

effectively handle these complexities. Thus, there is a need for new architectures and loss functions that can address these limitations and advance the effectiveness of segmentation methods for complex medical images. Our study proposes a neural network and a CL function to address class imbalance and limited data availability, especially in the medical image segmentation. We demonstrate the effectiveness of the proposed variant SegNet (vSegNet) for segmentation tasks in multi-modal images consisting of both grayscale and RGB images. We also present a comprehensive comparative analysis with many state-of-the-art neural models and discuss the prospects of the method.

2. Materials and Methods

This section describes the proposed deep fully convolutional neural network architecture with CL function, details of dataset used, and the state-of-the-art networks chosen for performance comparison.

2.1. vSegNet model

This study proposes a novel deep fully convolutional neural network architecture – vSegNet, shown in Figure 1, for semantic segmentation and tested the same on four public image datasets. It differs from regular SegNet in two ways.

1) vSegNet neural network model uses a varied number of convolution layers, which distinguishes it from the SegNet architecture [9] and Xavier initialization. Each encoder network convolution layer extracts features using a 3-by-3 filter bank and is followed by batch normalization and ReLU activation. Encoder extracts hierarchical features from input images, which are decoded to reconstruct segmented outputs with spatial precision using indices from Maxpooling layers. Maxpooling layers with stride 2 and kernel of 2 are used for all datasets.
2) To improve boundary precision and object localization, Dice loss (DL) and two-sided Hausdorff distance loss function (HDL) as given in the Equations (1) and (2) are combined to form a CL function as in Equation (5). DL helps to delineate overlapping structures, while HDL is useful when the region of interest (ROI) has shape dissimilarity. The non-weighted loss function combination helps to improve the efficiency in segmentation tasks involving non-overlapping regions and boundaries. The CL with the DL and the HDL addresses the challenges of class imbalance and accurate boundary delineation. The DL, which measures overlap, ensures robust segmentation of small or irregularly shaped regions. The HDL, focusing on the maximum boundary error, enhances the precision of the segmented contours. This combination is especially required for medical image segmentation, where accurate identification of boundaries and balancing the classes are crucial for clinical decision-making. By integrating these metrics, our approach significantly improves segmentation performance, especially in datasets with limited samples and pronounced class imbalance.

$$DL = 1 - \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{1}$$

where p_i is the values of predicted posterior probability and g_i is that of the ground truth of corresponding pixel.

$$HDL = H(A,B) = max\{h(A,B), h(B,A)\} \tag{2}$$

where,

$$h(A,B) = \max_{x \in X} \min_{y \in Y} ||x - y||_2 \tag{3}$$

$$h(B,A) = \max_{y \in Y} \min_{x \in X} ||x - y||_2 \tag{4}$$

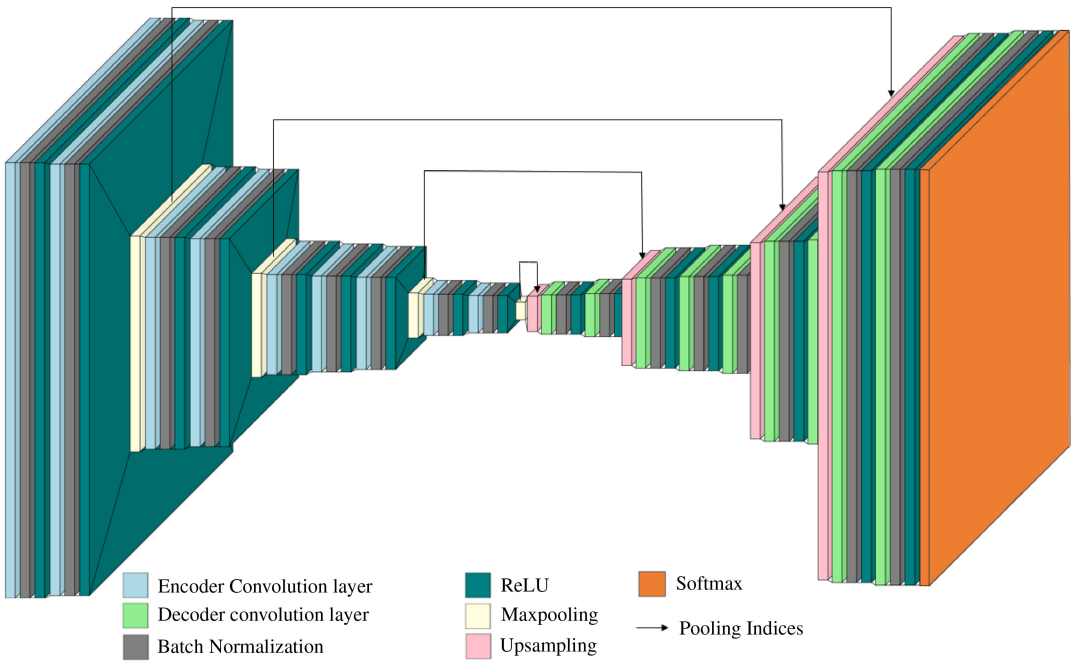

Figure 1. Proposed vSegNet model. Input layer is placed before the first encoder convolution layer, and the pixel classification layer is included after softmax layer.

$$CL = DL + HDL \tag{5}$$

A deeper encoder in vSegNet enables the extraction of more complex and detailed features from input images, which is particularly beneficial for medical image segmentation tasks. The decoder network in vSegNet mirrors the encoder's structure but focuses on up-sampling the feature maps to match the original image resolution. The skip connections between the corresponding encoder and decoder layers preserve fine-grained details that might be lost during pooling operations. Improved encoder-decoder connections and the proposed CL function ensure that the feature maps are reconstructed with better resolution. This enhances the model's ability to accurately segment small or intricate structures in medical images, which could address the class imbalance challenge and provide improved segmentation in the smaller dataset. Alongside the proposed loss function, we also use the vSegNet with cross-entropy loss function to do a comparison between the performance of the vSegNet with the proposed loss function and the standard cross-entropy loss.

2.2. Dataset

vSegNet's segmentation performance was benchmarked using four open-source datasets. Sample images and their ground truth label are shown in Figure 2. The datasets are enumerated and described below.

1) Segmentation of lungs from the chest X-rays of Montgomery County (MC) and Shenzhen X-ray sets (SC): MC dataset is a deidentified chest X-rays repository and manual annotation for lungs from the Department of Health and Human Services of Maryland, USA, accessible for reference studies by the National Library of Medicine (NLM). SC is a collection of labeled chest X-rays for normal lungs and pulmonary TB manifestations from Shenzhen No.3 hospital in Shenzhen, China, and made publicly available by NLM [30–34]. Each of the grayscale images is 256*256 pixel in.png format.
2) Vertebral body (VB) segmentation from computed tomography (CT) images: CT axial slices of 5 patients with the annotation for the VB from Computer vision and image processing lab, university of Louisville (CVIP) [35, 36]. This dataset did not stratify the samples based on patients. The dimension of the 2D axial slice is 512*512 in.bmp format. The ROI in the image contains a significantly lesser number of pixels compared to the background. Thus, we assessed the performance of the networks on the class-imbalanced images and augmented dataset by increasing the number of samples by augmenting with rotation in the range of $\pm 20^{\circ}$.
3) Foot ulcer (FU) dataset from Advancing the Zenith of Healthcare Wound and Vascular Center [37]. Ground truth for the wound area is marked and available along with the dataset. All the samples are 512*512*3 pixels in dimension in.png format. The ulcer is not always a single connected region, as can be seen in the image Figure 2(C).
4) T2-weighted sagittal slices of lumbar spine magnetic resonance imaging scans (MRI) images from SPIDER dataset (MRI dataset) [38]: MRI images are available with annotation for vertebrae, intervertebral disks, and spinal canal in each series. The annotated regions are together combined as foreground. A total of 257 patients' MRI series are used in the study, comprising 3,535 sagittal slices. Samples are resized to 256*256 in.png format. In our study, the dataset includes three anatomical structures: vertebrae, spinal canal, and intervertebral discs, each marked with different labels. For our binary segmentation tasks, we unified these labels by treating all three structures as a single foreground class, while assigning the rest of the image a background. This simplification was adopted to streamline the segmentation process and concentrate on distinguishing the spinal structures from the background. Combining these structures into a single class allows us to enhance the model's ability to detect and segment relevant spinal anatomy without the added complexity of differentiating between individual spinal components. This approach is particularly advantageous when the primary objective is to isolate

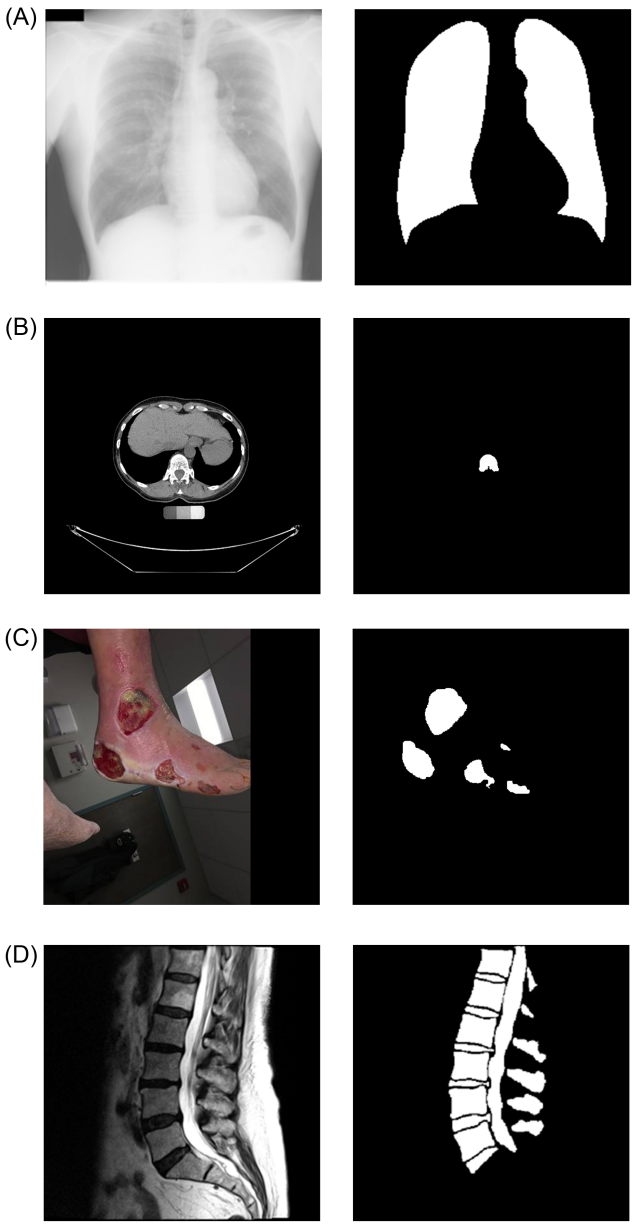

Figure 2. Samples from datasets and the corresponding label (segmentation ground truth) used in this study to benchmark the proposed neural network model vSegNet for segmentation. (A) X-ray image of chest: Sample image from MC SC dataset. (B) CT image of vertebrae: Sample image from CVIP dataset. (C) RGB image of foot ulcer: Sample image from FU dataset. (D) MRI image of lumbar spine: Sample image from SPIDER (MRI dataset).

Table 1. Overview of training. The proposed model achieves good training accuracy, although with a longer training duration allowing for a more comprehensive understanding of the dataset.

Dataset	Characteristics	vSegNet (CL)	vSegNet (Cross-entropy)	Pretrained FCN	Pretrained DeepLabv3+	MobileNetV2	VGG16	SegNet	Pretrained SegNet	U-Net
MC SC	Training Time (HH:MM:SS)	0:37:48	0:38:41	0:41:30	0:39:07	1:03:25	0:17:52	0:33:24	0:51:00	0:24:42
	Forward pass time (in ms)	115.9	119.4	149.1	236.1	346.6	128.54	126.94	203.4	101.4
	Backward pass time (in ms)	121.2	123	162.8	277.2	412.8	155.3	137.2	245.8	102.6
	Training Accuracy (in %)	98.85	97.48	97.73	99.69	92.88	98.5	98.6	99.38	98.61
	Number of Learnable	**594822**	594822	134285124	20607636	2268068	14719812	520710	29444166	7696258
	Model size (in MB)	**4.35**	4.35	1390	156	16.4	103	5.67	341	54.7
CVIP	Training Time (HH:MM:SS)	2:00:11	2:03:51	3:20:53	1:18:40	1:59:30	1:19:15	1:56:31	3:18:58	2:12:41
	Forward pass time (in ms)	127.4	128.2	501.7	159.9	287.1	160.5	295.3	522.1	103.7
	Backward pass time (in ms)	157.7	159.5	540.1	197.1	334.8	211	331.6	556.7	126.4
	Training Accuracy (in %)	99.91	99.88	99.88	99.99	99.08	99.86	99.74	99.79	99.88
	Number of Learnable	**594822**	594822	134285124	20607636	2268068	14719812	520710	29444166	7696258
	Model size (in MB)	**4.32**	4.32	1390	156	16.3	103	5.65	313	13.3
Augmented CVIP	Training Time (HH:MM:SS)	2:08:21	2:12:02	3:33:26	1:20:13	1:56:30	1:21:42	2:00:55	3:30:27	2:20:07
	Forward pass time (in ms)	139.1	141.6	504.8	160.5	301.1	161.8	314.3	530.2	129.5
	Backward pass time (in ms)	161.6	165.4	556.5	201.7	346.7	244.4	354.7	597.9	138.1
	Training Accuracy (in %)	99.93	99.83	99.83	99.88	99.41	99.63	99.82	99.68	99.9
	Number of Learnable	**594822**	594822	134285124	20607636	2268068	14719812	520710	29444166	7696258
	Model size (in MB)	**4.44**	4.44	1390	156	16.5	103	5.76	313	54.8
FU	Training Time (HH:MM:SS)	5:47:44	6:12:54	8:03:47	3:48:45	4:47:24	3:32:21	5:34:55	9:00:35	5:56:25
	Forward pass time (in ms)	116.5	118.2	186.6	140.6	245.8	211.8	216.1	362.8	101.1
	Backward pass time (in ms)	165.1	169.4	238.2	210.1	311	275.7	308.6	418.1	151.4
	Training Accuracy (in %)	99.37	99.25	99.83	99.94	99.53	99.72	99.34	99.86	86
	Number of Learnable	**595974**	595974	134285124	20607636	2268068	14719812	521862	29444166	7697410
	Model size (in MB)	**4.82**	4.82	1390	156	16.8	103	6.35	314	35
MRI	Training Time (HH:MM:SS)	10:08:30	10:53:04	11:46:17	5:45:22	6:03:58	5:11:07	9:56:39	11:24:15	7:42:27
	Forward pass time (in ms)	119.6	124.8	153.4	245.2	387.3	142.3	127.9	208.1	102.8
	Backward pass time (in ms)	124.3	128.9	180.3	281.4	419.5	159.2	144.3	250.6	106.1
	Training Accuracy (in %)	99.67	99.41	97.42	99.71	91.46	96.88	97.14	98.24	97.68
	Number of Learnable	**595974**	595974	134285124	20607636	2268068	14719812	521862	29444166	7697410
	Model size (in MB)	**4.82**	4.82	1390	156	16.8	103	6.35	314	35

any spinal structure from the surrounding tissues, thus improving segmentation accuracy and reducing computational complexity. This dataset is also large and thus helped in evaluating the proposed networks generality in segmentation in larger datasets as well.

For training, validating, and testing the performance, data in each dataset are split as follows:

1) MC SC dataset: 197 images for training, 25 for validation, and 25 for testing
2) CVIP dataset: 279 samples for training, 35 for validation, and 35 for performance evaluation
3) FU dataset: 648 images for training, 81 samples for validation, and 81 for testing
4) MRI dataset: 2829 images for training, 353 images for validation, and 353 samples for testing

In this study, we did not apply any preprocessing techniques to the images. This approach was chosen to evaluate the segmentation performance of the proposed model without any potential bias due to preprocessing steps.

2.3. Benchmarking networks

We compared vSegNet model with various state-of-the-art neural networks on four public datasets as described in Section 2.2 and evaluated the performance for two different loss functions. We utilized U-Net [3], SegNet (without weights) [9], SegNet (with VGG16 pretrained weights from the ImageNet dataset) [9], VGG16 [6], MobileNetV2 [39], Pretrained DeepLabv3+ with Resnet18 weights [13], and Pretrained fully convolutional network (FCN) with VGG16 weights from ImageNet dataset [40] to benchmark and compare the performance of vSegNet. In our study, we employed original weight initialization methods for the networks, U-Net, SegNet, Decoder layers of Pretrained SegNet used the He method, and models VGG16 and MobileNetV2 used the Gloret method. As the original architectures of U-Net, SegNet, VGG16, MobileNetV2, DeepLabv3+, and FCN models use cross-entropy as their loss function, we have maintained the same in this study. All models were trained for 100 epochs for all datasets except the MRI dataset. Due to the larger training set in MRI dataset, the number of epochs is reduced to 20 during training to prevent overfitting and improve computational efficiency. The training was end-to-end with a constant learning rate of 0.001 with mini-batch size of 2. All computations were performed using MATLAB® R2021a on NVIDIA® Titan Xp. Titan Xp GPU used had Pascal architecture with memory speed 11.4 Gbps, boost clock 1582 MHz and 12 GB standard memory configuration.

3. Results

For faster convergence and better generalization, mini-batch size of 2 is used to train models considering image dimension and GPU memory. Validation dataset is used for unbiased evaluation and hyperparameter tuning. Stochastic Gradient Descent with Momentum is used to optimize training process.

3.1. Overview of training and testing

Models are trained without dropout and early stop and evaluated using testing, and validation data. Convolution layers use 3*3 kernels. Training accuracy, number of weights learned, and size of the models in MegaBytes are presented in Table 1 for comparison. Forward and backward pass times are calculated as the average of 10 iterations with a Minibatch size of 1. vSegNet model has 18 convolution layers together in its encoder and decoder, making it deeper than SegNet. However, it has fewer learnable parameters resulting in a shorter training time and better training accuracy. vSegNet shows training efficacy with minimal loss of accuracy (less than 1% compared to the best training efficiency), as presented in Table 1. Figure 3 shows the results of computational efficacy of segmentation using 10 samples from the test set. vSegNet model performed efficiently, while most time expensive was for the pretrained SegNet. This is due to number of learnable parameters and model size (see Table 1), which are the lowest for vSegNet. FU dataset takes the longest test time for all the models, due to the higher resolution images.

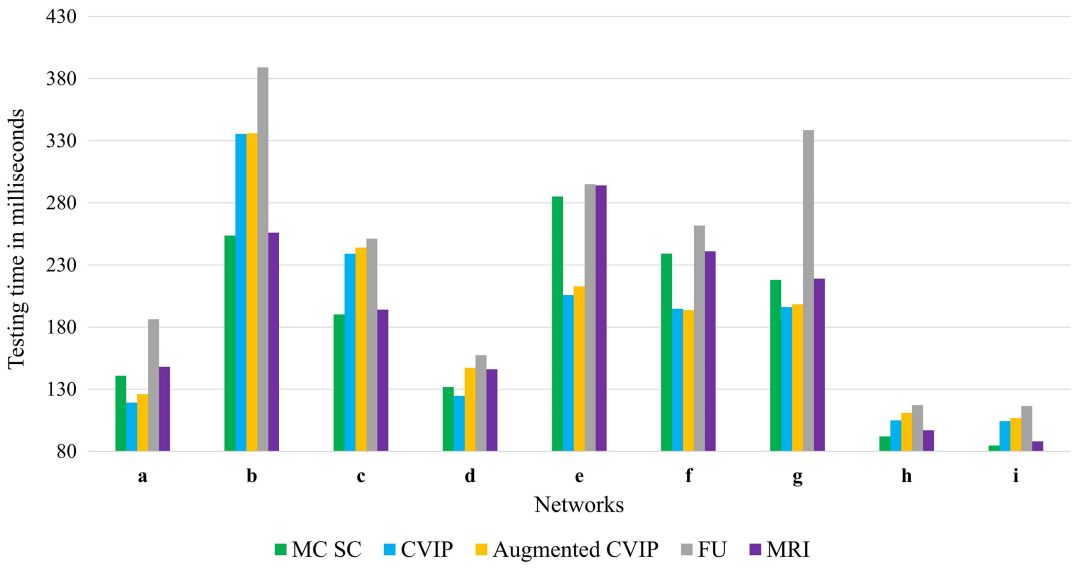

Figure 3. Comparison of testing time. vSegNet (both h and i) has notably lesser testing time. Model labels, a: U-Net, b: Pretrained SegNet, c: SegNet, d: VGG16, e: MobileNetV2, f: Pretrained DeepLabv3+, g: Pretrained FCN, h: vSegNet (cross-entropy), i: vSegNet (CL).

3.2. Qualitative comparison of segmentation results

Figure 4 shows the segmentation performance of different networks on one example from each dataset's test group, with ground truth and the segmentation done by the different models. In each of the images, correctly segmented regions are shown in white and the incorrect regions are shown in red (background misclassified) and blue (foreground misclassified). vSegNet with a combined loss function demonstrated good performance in MC SC dataset, while the boundary errors are particularly significant in segmentation done by other networks. MobileNetV2 failed to segment vertebral bodies in the CVIP dataset and U-Net in the FU dataset (refer Figure 4(E) and 4(I)). U-Net segmented many unwanted features along with the label in the CVIP datasets (refer Figure 4(E)). In the FU dataset, vSegNet demonstrated superior segmentation quality compared to U-Net, especially using the proposed combined loss function compared to cross-entropy. The results clearly highlight vSegNet's ability to accurately delineate regions of interest, showcasing superior performance in comparison to other models and affirming its robustness in handling complex medical image segmentation tasks.

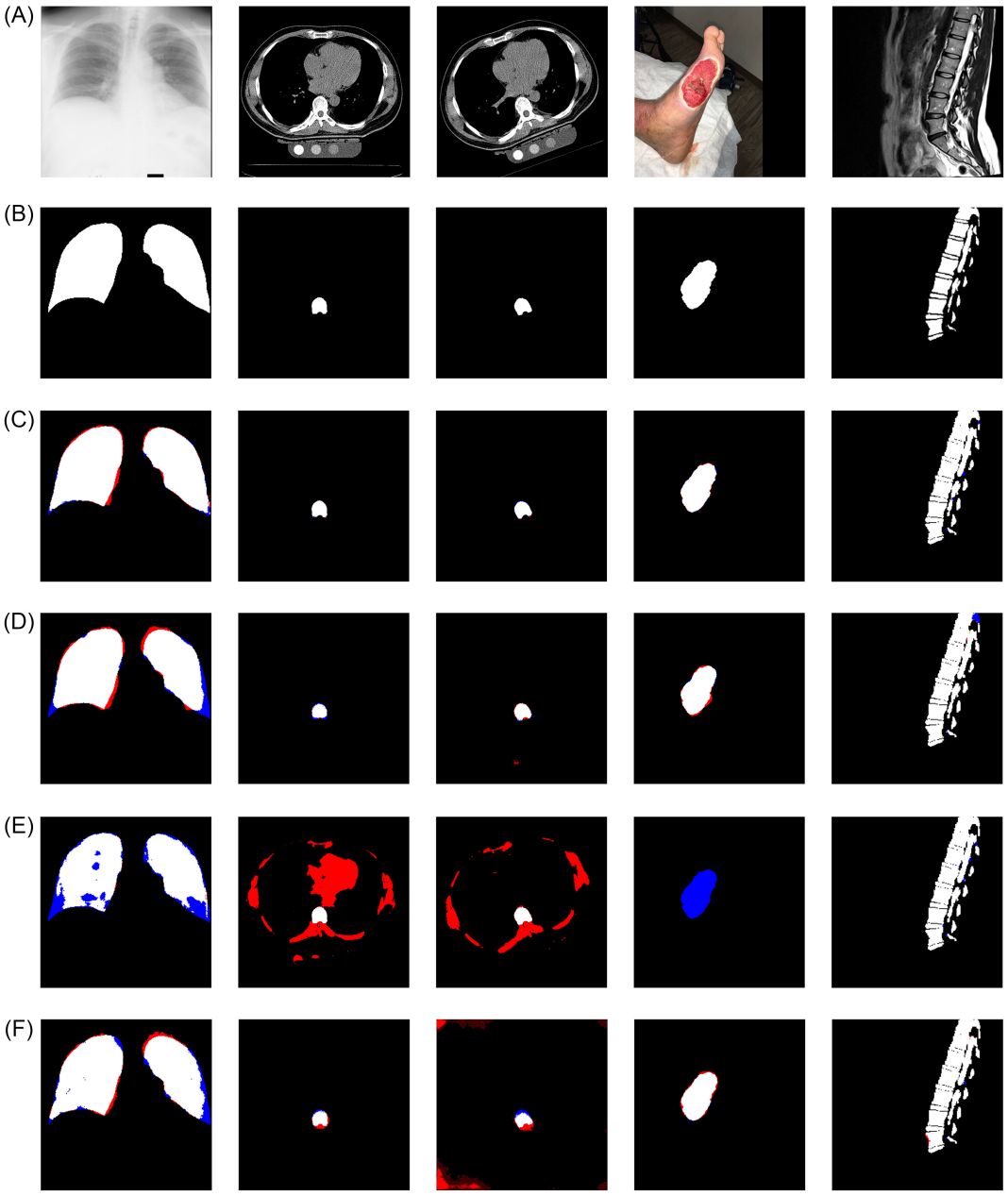

Figure 4. Sample illustration of the performance of segmentation by different models considered in this study from each dataset, showing true positives (white), true negatives (black), false positives (red), and false negatives (blue). Column 1 is a sample from MC SC dataset. Columns 2 & 3 show a sample from the CVIP dataset without and with augmentation, respectively. Column 4 is a sample from FU dataset. Column 5 is a sample from MRI dataset. (A) Input image. (B) Ground truth. (C) Segmentation by vSegNet with CL. (D) Segmentation by vSegNet with cross-entropy as loss function. (E) Segmentation by U-Net. (F) Segmentation by Pretrained SegNet. (G) Segmentation by SegNet. (H) Segmentation by VGG16. (I) Segmentation by MobileNetV2. (J) Segmentation by Pretrained DeepLabv3+. (K) Segmentation by Pretrained FCN.

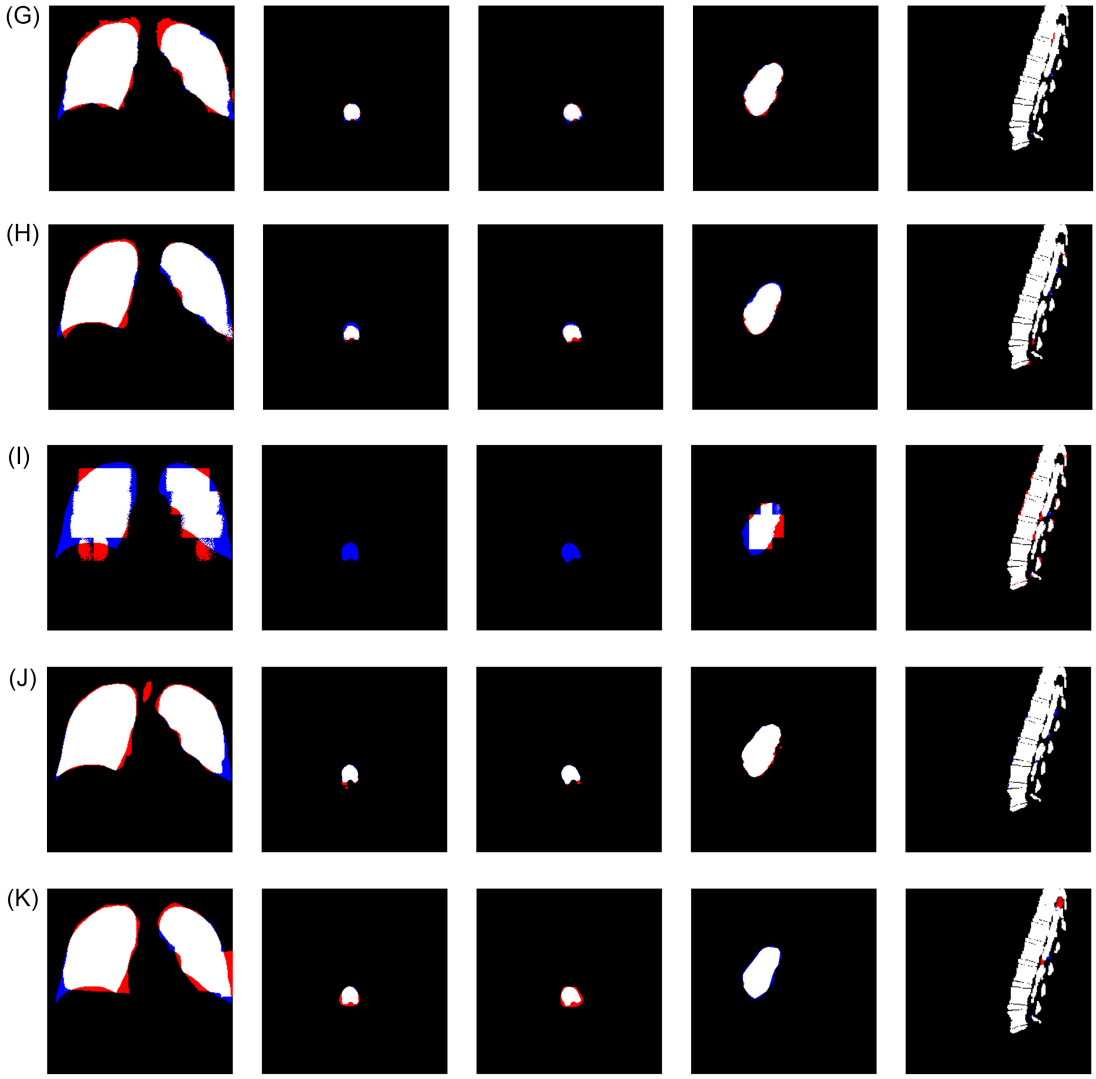

Figure 4. (Continued)

Table 2. Performance metrics for different models on MC SC dataset and MRI dataset. Model labels, a: U-Net, b: Pretrained SegNet, c: SegNet, d: VGG16, e: MobileNetV2, f: Pretrained DeepLabv3+, g: Pretrained FCN, h: vSegNet (cross-entropy), i: vSegNet (CL).

	MC SC dataset			MRI dataset		
Models	Mean accuracy	Mean IoU	Mean BF score	Mean accuracy	Mean IoU	Mean BF score
a	0.500	0.143	0.003	0.939	0.904	0.92
b	0.957	0.932	0.803	0.828	0.73	0.793
c	0.964	0.926	0.780	0.977	0.961	0.924
d	0.974	0.938	0.814	0.889	0.824	0.836
e	0.900	0.832	0.824	0.81	0.719	0.801
f	0.977	0.947	0.878	0.983	0.92	0.951
g	0.963	0.908	0.905	0.946	0.914	0.929
h	0.969	0.936	0.842	0.925	0.892	0.946
i	**0.980**	**0.958**	**0.917**	**0.987**	**0.97**	**0.982**

3.3. Comparison based on evaluation metrics

We evaluated models for accurately segmenting foregrounds using mean accuracy, mean Intersection of Union (IoU), and mean BF score. The higher mean IoU values show better segmentation accuracy, indicating regions of interest are segmented more effectively. Similarly, the higher mean BF scores ensure precise segmentation. The results are presented in Tables 2, 3, and 4. They show that vSegNet model performed better than other models consistently, with a higher mean IoU and mean BF score. However, the performance of all networks is generally poor in FU dataset as compared to other datasets, and the variability in performance is also more for all models for this dataset. Considering both Dice score (as shown in Figure 5) and HD evaluation (given in Figure 6), vSegNet exhibits high Dice scores and low HD, indicating excellent overlap with ground truth

Table 3. Performance metrics for different models on CVIP and augmented CVIP dataset. Model labels, a: U-Net, b: Pretrained SegNet, c: SegNet, d: VGG16, e: MobileNetV2, f: Pretrained DeepLabv3+, g: Pretrained FCN, h: vSegNet (cross-entropy), i: vSegNet (CL).

	CVIP dataset			Augmented CVIP dataset		
Models	Mean accuracy	Mean IoU	Mean BF score	Mean accuracy	Mean IoU	Mean BF score
a	0.642	0.615	0.411	0.797	0.764	0.861
b	0.894	0.642	0.811	0.924	0.659	0.857
c	0.913	0.796	0.920	0.948	0.791	0.926
d	0.787	0.753	0.856	0.942	0.824	0.927
e	0.500	0.500	0.000	0.500	0.497	0.467
f	0.897	0.802	0.937	0.950	0.829	0.947
g	0.939	0.815	0.890	0.946	0.838	0.927
h	0.931	0.495	0.481	0.933	0.757	0.694
i	**0.963**	**0.914**	**0.961**	**0.976**	**0.945**	**0.974**

Table 4. Performance metrics for different models on FU dataset. Model labels, a: U-Net, b: Pretrained SegNet, c: SegNet, d: VGG16, e: MobileNetV2, f: Pretrained DeepLabv3+, g: Pretrained FCN, h: vSegNet (cross-entropy), i: vSegNet (CL).

Models	Mean accuracy	Mean IoU	Mean BF score
a	0.500	0.007	0.000
b	0.875	0.804	0.791
c	0.780	0.724	0.726
d	0.813	0.749	0.800
e	0.685	0.655	0.802
f	0.868	0.845	0.849
g	0.887	0.841	0.850
h	0.799	0.725	0.792
i	**0.889**	**0.851**	**0.852**

and minimal boundary errors. This demonstrates the model's precision and reliability in segmentation tasks, providing accurate and consistent results, for all the datasets considered in this study. For each dataset, the quantitative metrics accuracy, precision, recall, and $F1$ score are presented in Figure 7. The widespread interquartile range (IQR) of Pretrained SegNet, VGG16 for CVIP dataset in Figure 7(B) and wider IQR for models in Figure 7(D) suggest higher variability in the prediction performance of these models. MobileNetV2's performance on the FU dataset exhibits more outliers, indicating a tendency to produce unreliable segmentation results. The compact IQR in these plots corresponding to vSegNet suggests its consistent performance across various datasets, with fewer outliers compared to other models. This consistency highlights the robustness and reliability of the proposed model and loss function. The whiskers of the boxplot, representing the range of scores, are shorter and thus further demonstrate vSegNet's enhanced efficacy in delineation.

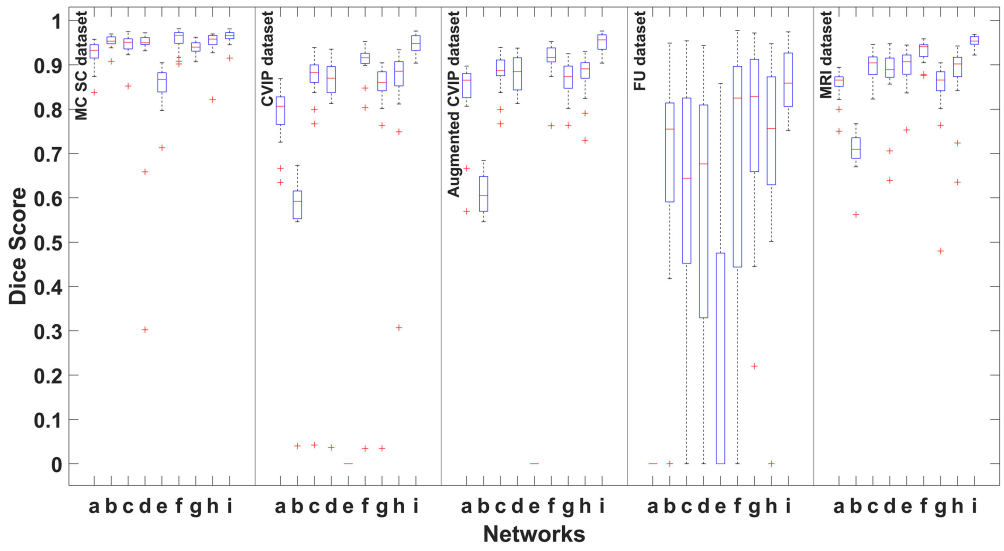

Figure 5. Comparison of Dice score for various models for different datasets given as box plot. The red line indicates the median and the blue box represents the 25% and the 75% in the Dice score. Model labels, a: U-Net, b: Pretrained SegNet, c: SegNet, d: VGG16, e: MobileNetV2, f: Pretrained DeepLabv3+, g: Pretrained FCN, h: vSegNet (cross-entropy), i: vSegNet (CL).

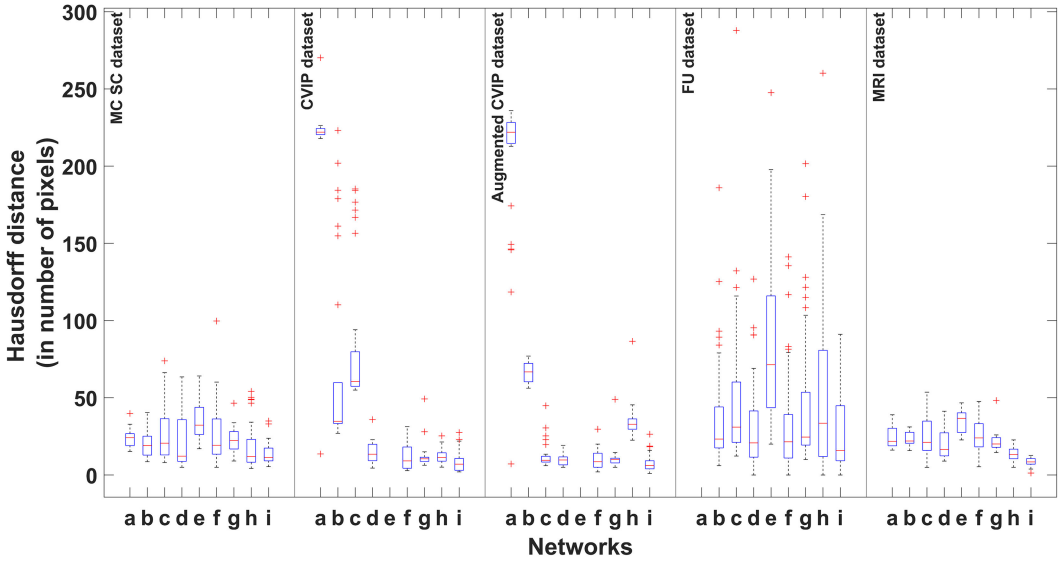

Figure 6. Comparison using Hausdorff distance. The red line indicates the median and the blue box represents the 25% and the 75% in the Hausdorff distance. Model labels, a: U-Net, b: Pretrained SegNet, c: SegNet, d: VGG16, e: MobileNetV2, f: Pretrained DeepLabv3+, g: Pretrained FCN, h: vSegNet (cross-entropy), i: vSegNet (CL).

4. Discussion

This study presents a variant deep fully convolutional neural network architecture – vSegNet aimed at enhancing medical image segmentation performance in the presence of class-imbalanced limited datasets. This architecture is coupled with a combined loss function to significantly improve the model's ability to accurately segment minority classes. Additionally, in this work, we present segmentation on X-rays, CT scans, MR images, and RGB images, demonstrating the model's applicability across various medical imaging modalities. The analysis of segmentation accuracy with various qualitative and quantitative metrics and performance comparisons with many state-of-art networks collectively demonstrate a substantial improvement in segmentation accuracy and robustness of the proposed model vSegNet, providing a promising solution for medical practitioners dealing with limited and imbalanced data.

Chest X-ray segmentation for normal lungs and pulmonary TB manifestations have been demonstrated in some contemporary works. In a work [41] that uses nn-UNET, segmentation accuracy with Dice score and HD of 0.955 ± 0.068 and 44.79 ± 60.31 for MC dataset, and 0.949 ± 0.055 and 61.31 ± 67.97 for SC dataset is reported. In another work [42], region-based Fuzzy C-Means, and Mean Shift models (Dice score 0.68) struggle with intensity similarity, causing mixing of background and lung regions, making them perform very poorly for limited datasets with class imbalance. A saliency model, with a higher Dice score of 0.87, shows better performance but not suitable for medical image segmentation tasks where clinical decision is crucial [42]. Similarly, FractalCovNet, U-Net, and 3D-CNN models show poor performance (accuracy: 65%, 76%, 85%; recall: 14%, 28%, 38%, respectively) [43]. Their low precision (85%, 65%, 76%) and $F1$-scores (30%, 42%, 53%, respectively) highlight their ineffectiveness in handling small, imbalanced datasets due to poor positive sample identification and low overlap between predicted and actual labels. In comparison, the proposed model vSegNet achieved a Dice score of 0.96 ± 0.01 and a HD of 14.33 ± 7.74 in this dataset showing considerable improvement in the segmentation accuracy in this application.

Various level set methods have been used for vertebrae segmentation from CT and MRI datasets and reported Dice scores ranging from 68.53 ± 3.06 by the Chan-Veses model to 92.08 ± 2.37 by the automatic global level set approach method presented in Li et al. [44]. The residual U-Net ensemble model achieves Dice scores of 0.88 [45] and HDs of 11.7 [46] on this class of segmentation task. The main challenges in this application include the variability in spinal shapes, poor signal-to-noise ratios in MRI, and difficulties in segmenting severely stenosed regions [45, 46]. These factors contribute to the lower scores and reduced reliability for datasets with limited data and class imbalance. When using vSegNet to delineate vertebrae, intervertebral disks, and spinal canal in the MR images, we obtained Dice scores and HD of 0.95 ± 0.01 and 8.45 ± 2.81, respectively. Our results highlight the effectiveness of using CL function and data augmentation in vertebrae segmentation in CT images, achieving a Dice score of 0.95 ± 0.02 and HD of 7.99 ± 6.05.

Wound and ulcer segmentation in RGB images is another significant image processing task finding application in clinical diagnosis. A study on wound segmentation using a deep learning approach observed a mean IoU of 0.78 ± 0.02 [47]. A Detect-and-Segment (DS) approach [47] improved segmentation metrics like Matthews' correlation coefficient from 0.77 to 0.85 and IoU from 0.63 to 0.75. But this approach could improve the IOU from 0.53 to 0.60 only on another dataset consisting of systemic sclerosis digital ulcers. Thus, the DS method is poor in generalizing on limited datasets with class imbalance as it cannot effectively handle the heterogeneity in wound types and imaging conditions. In a study on Diabetic FU Segmentation Challenge [48], various models achieved a Dice score of 0.69, but this varied with dataset size and quality. Higher HD was observed for complex wounds. Precision ranged from 0.72 to 0.82. In comparison, when using the proposed vSegNet on the FU delineation, we obtained precision, Dice score, and HD of 0.89 ± 0.54, 0.86 ± 0.07, and 29.32 ± 25.64, respectively.

From the various extensive set of training and testing of different model considered in this study, it is observed that pretrained models like SegNet, DeepLabv3+, and FCN with more learnable parameters converge faster but underperform on test data showing high variability when compared to vSegNet due to limited training samples. DeepLabv3+ excelled in capturing fine-

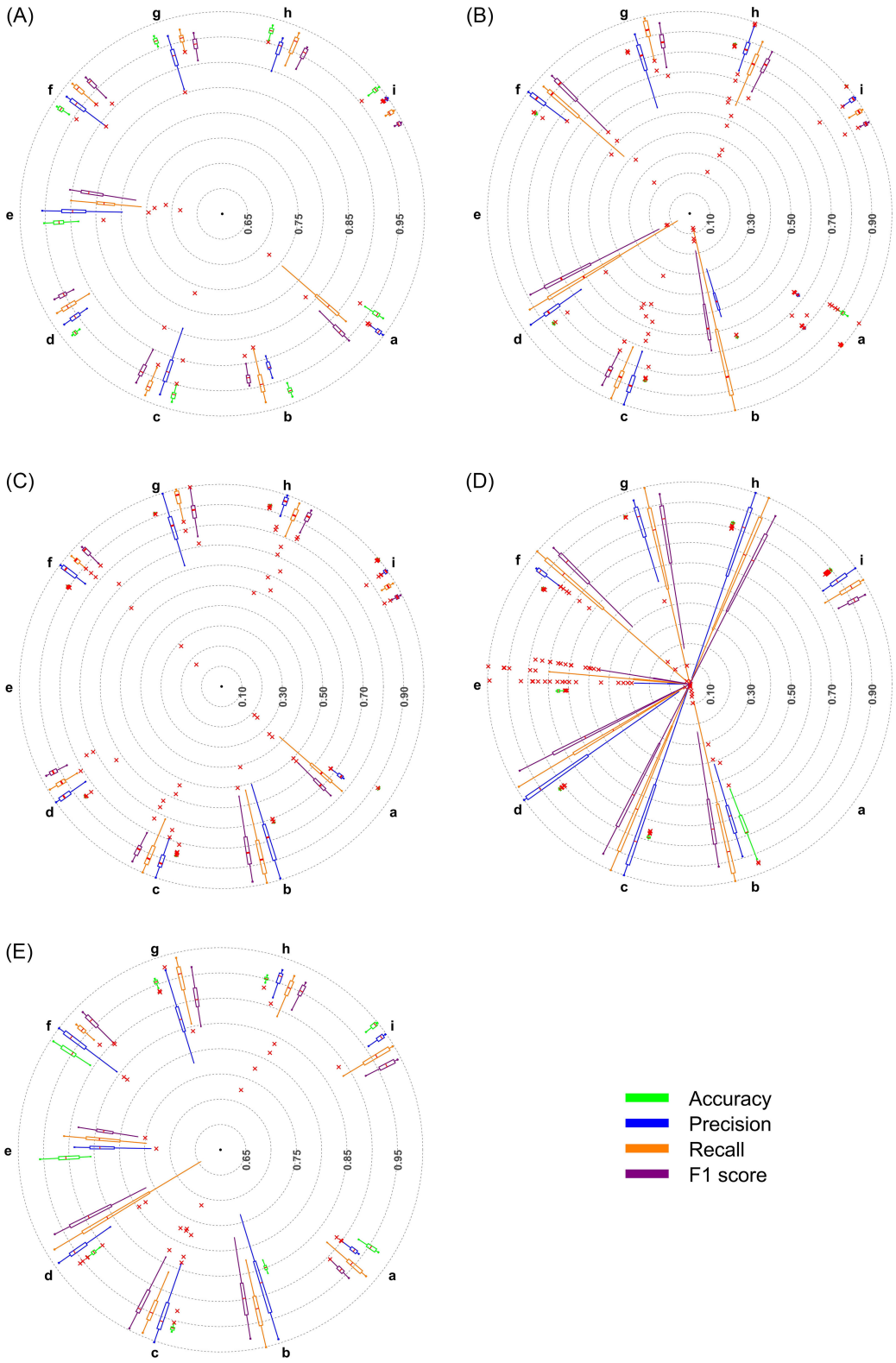

Figure 7. Comparison using quantitative metrics. Model labels, a: U-Net, b: Pretrained SegNet, c: SegNet, d: VGG16, e: MobileNetV2, f: Pretrained DeepLabv3+, g: Pretrained FCN, h: vSegNet (cross-entropy), i: vSegNet (CL). (A) MC SC dataset. (B) CVIP dataset. (C) Augmented CVIP dataset. (D) FU dataset. (E) MRI dataset.

grained details and object boundaries in X-ray images. SegNet performed well on MC SC and CVIP datasets, but its performance was poor in FU dataset. U-Net performs generally well on grayscale images but not on RGB images. MobileNetV2's performance is significantly poor in FU dataset. Unlike the MC SC dataset, ROI in CVIP dataset has fewer foreground pixels than the background. Augmentation did not improve the performance of most models except U-Net, whose performance improved

significantly (refer Figure 4 and metrics from Table 3). Thus, we see generalization remains a challenge for medical image segmentation models when exposed to new and diverse imaging contexts.

Deeper neural networks often perform poorly on smaller datasets with a class imbalance in image segmentation tasks due to their high complexity and large number of parameters. These networks require extensive data to learn effective feature representations. When data are limited, they tend to overfit, capturing noise as features instead of intricate features. Class imbalance further exacerbates this challenge, as the networks may become biased towards the majority class or background, leading to poor segmentation performance for minority classes or ROI. Consequently, deeper networks have poor generalization and cannot provide accurate and robust segmentation results under constraints like class imbalance and limited datasets. To summarize, selecting a neural network model for medical image segmentation is not one-size-fits-all but requires careful consideration of the task requirements and characteristics of dataset. The findings of this study offer valuable insights into various models' performance for different medical imaging datasets, aiding in decision-making for real-world medical applications. The proposed vSegNet with CL function excels in intricate segmentation tasks trained on a limited number of samples and/or with smaller ROI.

This study reported the performance of the neural network models on segmenting ROI in different imaging modalities, by training them individually on different modalities. One can extend the method to train and evaluate the proposed model to handle multi-modal inputs simultaneously. This shall enhance the model's versatility and applicability in different clinical scenarios. Another improvement that can be implemented is adaptive learning. This shall allow the model to continuously learn from new data and adapt to changed scenarios in the clinical application, thereby maintaining high performance over time. Increasing the dataset diversity by including more images from various demographics and clinical settings will help improve the robustness and generalizability of the proposed model.

5. Conclusion

This work proposed a novel neural network architecture, vSegNet, designed specifically for medical image segmentation in the datasets that have challenges of limited samples and class imbalance. This architecture features convolutional layers and a CL function aimed at enhancing feature extraction and segmentation performance in such datasets. The CL function involving Dice score and HD balances precision and recall and provides a more comprehensive evaluation of segmentation quality, which helps in effectively segmenting smaller regions of interest with good boundary delineation. The article also reports a comprehensive comparison of the performance of the proposed model against several state-of-the-art neural network architectures, namely U-Net, Pretrained SegNet, SegNet, VGG16, MobileNetV2, Pretrained DeepLabv3+, and Pretrained FCN, on four different datasets namely, chest X-ray images, vertebral CT scans, RGB FU images, and spine MRI scans. The choice of diverse imaging modalities and anatomical features in the study enabled a comprehensive performance analysis. The results reveal that the proposed model vSegNet significantly performs better than the other models considered in this study in terms of segmentation accuracy and handling class imbalance, as demonstrated by improvements in the mean IOU, mean accuracy, mean BF score, accuracy, precision, recall, *F*1 score, Dice score, and HD. The proposed vSegNet architecture and loss function provide effective solutions for achieving high-quality segmentation in class-imbalanced medical images. This improvement is crucial for accurate diagnosis and treatment planning in many medical imaging applications.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Openi at https://openi.nlm.nih.gov/imgs/collections/NLM-MontgomeryCXRSet.zip, https://openi.nlm.nih.gov/imgs/collections/ChinaSet_AllFiles.zip, and https://openi.nlm.nih.gov/faq#faq-tb-coll; in SpineWeb at http://spineweb.digitalimaginggroup.ca/Index.php?%20n=Main.Datasets#Dataset_4.3A_CVIP_Spinal_CT_Database; in Github at https://github.com/uwm-bigdata/wound-segmentation, and in Zenodo at https://zenodo.org/records/10159290.

Author Contribution Statement

Iyyakutty Dheivya: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Gurunathan Saravana Kumar:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing – review & editing, Supervision.

References

[1] Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314

[2] Gill, G., Toews, M., & Beichel, R. R. (2014). Robust initialization of active shape models for lung segmentation in CT scans: A feature-based atlas approach. *International Journal of Biomedical Imaging*, *2014*(1), 479154. https://doi.org/10.1155/2014/479154

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems*, 1–9.

[5] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . . , & Li, F. F. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*, 211–252. https://doi.org/10.1007/s11263-015-0816-y

[6] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv Preprint:1409.1556*.

[7] Dubey, A. K., & Jain, V. (2020). Automatic facial recognition using VGG16 based transfer learning model. *Journal of Information and Optimization Sciences*, *41*(7), 1589–1596. https://doi.org/10.1080/02522667.2020.1809126

[8] Howard, A. G. (2013). Some improvements on deep convolutional neural network based image classification. *arXiv Preprint:1312.5402*.

[9] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

[10] Theckedath, D., & Sedamkar, R. R. (2020). Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Computer Science*, *1*(2), 79. https://doi.org/10.1007/s42979-020-0114-9

[11] Qassim, H., Verma, A., & Feinzimer, D. (2018). Compressed residual-VGG16 CNN model for big data places image recognition. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference*, 169–175. https://doi.org/10.1109/CCWC.2018.8301729

[12] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

[13] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision*, 801–818.

[14] Kim, M., & Lee, B. D. (2021). Automatic lung segmentation on chest X-rays using self-attention deep neural network. *Sensors*, *21*(2), 369. https://doi.org/10.3390/s21020369

[15] Lu, H., Tian, S., Yu, L., Liu, L., Cheng, J., Wu, W., . . . , & Zhang, D. (2022). DCACNet: Dual context aggregation and attention-guided cross deconvolution network for medical image segmentation. *Computer Methods and Programs in Biomedicine*, *214*, 106566. https://doi.org/10.1016/j.cmpb.2021.106566

[16] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., . . . , & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv Preprint:1804.03999*.

[17] Hansen, S., Gautam, S., Jenssen, R., & Kampffmeyer, M. (2022). Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. *Medical Image Analysis*, *78*, 102385. https://doi.org/10.1016/j.media.2022.102385

[18] Dong, N., Kampffmeyer, M., Liang, X., Xu, M., Voiculescu, I., & Xing, E. (2022). Towards robust partially supervised multi-structure medical image segmentation on small-scale data. *Applied Soft Computing*, *114*, 108074. https://doi.org/10.1016/j.asoc.2021.108074

[19] Hryniowski, A., & Wong, A. (2019). DeepLABNet: End-to-end learning of deep radial basis networks. *Journal of Computational Vision and Imaging Systems*, *5*(1), 1–1.

[20] Cui, X., Chang, S., Li, C., Kong, B., Tian, L., Wang, H., . . . , & Li, Z. (2021). DEAttack: A differential evolution based attack method for the robustness evaluation of medical image segmentation. *Neurocomputing*, *465*, 38–52. https://doi.org/10.1016/j.neucom.2021.08.118

[21] Zhou, Q., Wang, Q., Bao, Y., Kong, L., Jin, X., & Ou, W. (2022). LAEDNet: A lightweight attention encoder–decoder network for ultrasound medical image segmentation. *Computers and Electrical Engineering*, *99*, 107777. https://doi.org/10.1016/j.compeleceng.2022.107777

[22] Tao, G., Li, H., Huang, J., Han, C., Chen, J., Ruan, G., . . . , & Cai, H. (2022). SeqSeg: A sequential method to achieve nasopharyngeal carcinoma segmentation free from background dominance. *Medical Image Analysis*, *78*, 102381. https://doi.org/10.1016/j.media.2022.102381

[23] Tang, P., Yang, P., Nie, D., Wu, X., Zhou, J., & Wang, Y. (2022). Unified medical image segmentation by learning from uncertainty in an end-to-end manner. *Knowledge-Based Systems*, *241*, 108215. https://doi.org/10.1016/j.knosys.2022.108215

[24] Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision*, 565–571. https://doi.org/10.1109/3DV.2016.79

[25] Karimi, D., & Salcudean, S. E. (2020). Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on Medical Imaging*, *39*(2), 499–513. https://doi.org/10.1109/TMI.2019.2930068

[26] Baumgartner, C. F., Koch, L. M., Pollefeys, M., & Konukoglu, E. (2018). An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*, 111–119. https://doi.org/10.1007/978-3-319-75541-0_12

[27] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., . . . , & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

[28] Yeung, M., Sala, E., Schönlieb, C. B., & Rundo, L. (2022). Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, *95*, 102026. https://doi.org/10.1016/j.compmedimag.2021.102026

[29] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *arXiv Preprint: 1708.02002*.

[30] Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., . . . , & McDonald, C. J. (2014). Automatic tuberculosis screening using chest radiographs. *IEEE Transactions on Medical Imaging*, *33*(2), 233–245. https://doi.org/10.1109/TMI.2013.2284099

[31] Jaeger, S., Karargyris, A., Candemir, S., Siegelman, J., Folio, L., Antani, S., . . . , & McDonald, C. J. (2013). Automatic screening for tuberculosis in chest radiographs: A survey. *Quantitative Imaging in Medicine and Surgery*, *3*(2), 89–99. https://doi.org/10.3978%2Fj.issn.2223-4292.2013.04.03

[32] Candemir, S., Jaeger, S., Palaniappan, K., Musco, J. P., Singh, R. K., Xue, Z., . . . , & McDonald, C. J. (2014). Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Transactions on Medical Imaging*, *33*(2), 577–590. https://doi.org/10.1109/TMI.2013.2290491

[33] Rajaraman, S., Folio, L. R., Dimperio, J., Alderson, P. O., & Antani, S. K. (2021). Improved semantic segmentation of tuberculosis—Consistent findings in chest X-rays using augmented training of modality-specific U-Net models with weak localizations. *Diagnostics*, *11*(4), 616. https://doi.org/10.3390/diagnostics11040616

[34] Jaeger, S., Candemir, S., Antani, S., Wáng, Y. X. J., Lu, P. X., & Thoma, G. (2014). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative Imaging in Medicine and Surgery*, *4*(6), 475–477. https://doi.org/10.3978%2Fj.issn.2223-4292.2014.11.20

[35] Aslan, M. S., Ali, A., Rara, H., Arnold, B., Farag, A. A., Fahmi, R., & Xiang, P. (2009). A novel 3D segmentation of vertebral bones from volumetric CT images using graph cuts. In *Advances in Visual Computing: 5th International Symposium*, 519–528. https://doi.org/10.1007/978-3-642-10520-3_49

[36] Aslan, M. S., Shalaby, A., & Farag, A. A. (2013). Clinically desired segmentation method for vertebral bodies. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, 840–843. https://doi.org/10.1109/ISBI.2013.6556606

[37] Wang, C., Anisuzzaman, D. M., Williamson, V., Dhar, M. K., Rostami, B., Niezgoda, J., ..., & Yu, Z. (2020). Fully automatic wound segmentation with deep convolutional neural networks. *Scientific Reports*, *10*(1), 21897. https://doi.org/10.1038/s41598-020-78799-w

[38] van der Graaf, J. W., van Hooff, M. L., Buckens, C. F., Rutten, M., van Susante, J. L., Kroeze, R. J., ..., & Lessmann, N. (2024). Lumbar spine segmentation in MR images: A dataset and a public benchmark. *Scientific Data*, *11*(1), 264. https://doi.org/10.1038/s41597-024-03090-w

[39] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

[40] Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683

[41] Gaggion, N., Mansilla, L., Mosquera, C., Milone, D. H., & Ferrante, E. (2023). Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: Applications to chest X-ray analysis. *IEEE Transactions on Medical Imaging*, *42*(2), 546–556. https://doi.org/10.1109/TMI.2022.3224660

[42] de Almeida, P. A. C., & Borges, D. L. (2023). A deep unsupervised saliency model for lung segmentation in chest X-ray images. *Biomedical Signal Processing and Control*, *86*, 105334. https://doi.org/10.1016/j.bspc.2023.105334

[43] Junia, R. C., & Selvan, K. (2024). Deep learning-based automatic segmentation of COVID-19 in chest X-ray images using ensemble neural net sentinel algorithm. *Measurement: Sensors*, *33*, 101117. https://doi.org/10.1016/j.measen.2024.101117

[44] Li, Y., Liang, W., Zhang, Y., & Tan, J. (2018). Automatic global level set approach for lumbar vertebrae CT image segmentation. *BioMed Research International*, *2018*(1), 6319879. https://doi.org/10.1155/2018/6319879

[45] Wang, Z., Xiao, P., & Tan, H. (2023). Spinal magnetic resonance image segmentation based on U-Net. *Journal of Radiation Research and Applied Sciences*, *16*(3), 100627. https://doi.org/10.1016/j.jrras.2023.100627

[46] Laiwalla, A. N., Ratnaparkhi, A., Zarrin, D., Cook, K., Li, I., Wilson, B., ..., & Macyszyn, L. (2023). Lumbar spinal canal segmentation in cases with lumbar stenosis using deep-U-Net ensembles. *World Neurosurgery*, *178*, e135–e140. https://doi.org/10.1016/j.wneu.2023.07.009

[47] Scebba, G., Zhang, J., Catanzaro, S., Mihai, C., Distler, O., Berli, M., & Karlen, W. (2022). Detect-and-segment: A deep learning approach to automate wound image segmentation. *Informatics in Medicine Unlocked*, *29*, 100884. https://doi.org/10.1016/j.imu.2022.100884

[48] Yap, M. H., Cassidy, B., Byra, M., Liao, T. Y., Yi, H., Galdran, A., ..., & Kendrick, C. (2024). Diabetic foot ulcers segmentation challenge report: Benchmark and analysis. *Medical Image Analysis*, *94*, 103153. https://doi.org/10.1016/j.media.2024.103153

How to Cite: Dheivya, I., & Kumar, G. S. (2025). VSegNet – A Variant SegNet for Improving Segmentation Accuracy in Medical Images with Class Imbalance and Limited Data. *Medinformatics*, *2*(1), 36–48. https://doi.org/10.47852/bonviewMEDIN42023518