Old Dominion University

ODU Digital Commons

2021

Simultaneous Wound Border Segmentation and Tissue Classification Using a Conditional Generative Adversarial Network

Salih Sarp
Old Dominion University, ssarp001@odu.edu

Murat Kuzlu
Old Dominion University, mkuzlu@odu.edu

Manisa Pipattanasomporn

Ozgur Guler

IET The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Simultaneous wound border segmentation and tissue classification using a conditional generative adversarial network

Salih Sarp[1] | Murat Kuzlu[1] | Manisa Pipattanasomporn[2] | Ozgur Guler[3]

[1] Batten College of Engineering & Technology, Old Dominion University Norfolk, Norfolk, Virginia, USA

[2] Smart Grid Research Unit, Chulalongkorn University, Bangkok, Thailand

[3] eKare, Inc., Fairfax, Virginia, USA

Correspondence
Salih Sarp, 211B Kaufman Hall Norfolk, VA 23529, USA.
Email: ssarp001@odu.edu

Abstract

Generative adversarial network (GAN) applications on medical image synthesis have the potential to assist caregivers in deciding a proper chronic wound treatment plan by understanding the border segmentation and the wound tissue classification visually. This study proposes a hybrid wound border segmentation and tissue classification method utilising conditional GAN, which can mimic real data without expert knowledge. We trained the network on chronic wound datasets with different sizes. The performance of the GAN algorithm is evaluated through the mean squared error, Dice coefficient metrics and visual inspection of generated images. This study also analyses the optimum number of training images as well as the number of epochs using GAN for wound border segmentation and tissue classification. The results show that the proposed GAN model performs efficiently for wound border segmentation and tissue classification tasks with a set of 2000 images at 200 epochs.

1 | INTRODUCTION

Wound management technologies are an essential part of the treatment of chronic wounds, which affect around 6.5 million patients at the cost of $25 billion yearly in the United States. [1]. However, they are lagging technologically, and most caregivers only depend on imprecise optical assessment [2], which brings some complications, such as infection risks, inaccurate measurements, and discomfort to patients [3]. Advanced computer vision methods assist the accurate monitoring of wound healing [4]. Image processing and machine learning automate the evaluation of medical images [5]. The computer vision paired with artificial intelligence (AI) would provide caregivers with continuous and accurate wound healing monitoring at a lower cost. Familiarity with wound tissue types and their sizes play an important role in determining the right chronic wound treatment plan. One of the goals of this study is to contribute to the development of such a system for wound border segmentation and tissue classification utilising the conditional generative adversarial network (GAN) algorithm in a hybrid way.

Yann LeCun, an AI expert in neural networks, called adversarial training 'the most important idea in the last two decades in Machine Learning' [6]. The GAN algorithm of deep learn-

ing (DL) techniques has been used successfully in many applications, such as style transfer, image synthesising, and the famous DeepFake synthetic media creator. The power of the GAN algorithm comes from learning directly from data without human knowledge [7]. That means that GAN does not require a human to select features to predict; it extracts from the data itself. On the other hand, the GAN is challenging to train as the complicated loss functions are hard to interpret [8]. Tweaking hyperparameters such as the number of images and epochs for training in neural networks are still a subject of research and are being done empirically [9]. Finding the right hyperparameters are like a black art where there is no absolute path to follow [10].

The data-driven GAN algorithm provides automatic feature generation, which saves time and labour, but it needs a higher number of images as a trade-off [11]. That is why the number of images is the key parameter to achieve good approximation in GAN-based models. At the same time, data collection and management processes cost tens of millions of dollars in healthcare, such as clinical trials [12]. There are also significant concerns over privacy, confidentiality, and control of the data [13], which makes it difficult to obtain data in healthcare. Collecting data in healthcare is not an easy task, but the GAN algorithm could

www.manaraa.com

generate synthetic images that have no cost and could be used without hesitation.

The number of epochs for training is another critical parameter that requires many trials to find the optimum amount and expertise in the healthcare field [10]. High performance in minimal epoch is needed to achieve significant time and labour savings [14]. The question of how many images and epochs are needed to train a GAN has not been answered. This study also provides a rule of thumb to choose the right number of training images and epochs for GAN algorithms in healthcare applications.

Prior efforts for wound tissue classification and segmentation include the development of an image analysis algorithm that is capable of wound area assessment, segmentation, and extraction of wound colour without correlating to wound tissue from wound images using smartphone cameras [15]. The study in [16] proposes the use of the K-means clustering algorithm, which requires feature engineering, for the wound border segmentation and tissue classification using 113 images. Multispectral imaging is utilised by Thatcher et al. [17]. This study examines the tissue characteristics of burn wounds in the light of medical imaging without segmentation of the wound area and tissue. The authors in [18] explored the feasibility of RGB-D cameras in wound detection, segmentation, and chronic wound area measurement in 3D. However, the use of special RGB-D cameras increases the cost and model complexity of wound management systems. Also, tissue segmentation is not included in this study [18]. K-nearest neighbours, decision tree (DT), and linear discriminant analysis (LDA) are used for the burn tissue classification. This study is limited to burn wounds, whereas the proposed method in this study covers a variety of wound types. The study in [19] proposes a DL and data augmentation model for wound-region segmentation. The model used in the study [19] segments each wound tissue separately. Also, detection and segmentation tasks are done using different machine learning models. The authors in [20] propose an automatic skin ulcer region assessment framework using a convolutional neural network (CNN) and encoder/decoder deep neural network. The study in [20] achieves overall wound segmentation, but the segmentation of different wound tissues is not studied. The study in [21] proposes a CNN-based model for the segmentation of wound tissue types. The authors in [21] provide tissue segmentation of pressure injury wounds with the help of manual pre-processing steps, including external mask application and flashlight removal. The study in [22] describes a chronic wound status monitoring with wound tissue segmentation using LDA, DT, random forests, and naïve Bayesian. This study only segments the wound into two tissue types, whereas our proposed method gives more details. The authors in [23] proposed a model that utilises colour correction and a CNN for wound region segmentation. A two-step pre-processing pipeline is discussed in [23] to segment the overall wound without tissue segmentation. The authors in [24] propose a model that segments solely diabetic wounds using CNN and the removal of artefacts with probability maps after a pre-processing step. The study in [25] investigated CNN with different architectures, that is, U-Net, Segnet, FCN8, and FSN32 for the wound tissue seg-

mentation. The study in [26] proposes a wound segmentation model using both traditional and deep learning methods. In [25], the authors added a pre-processing step that includes detection of the wound and a post-processing step that segments solely the overall wound area. The model also could not be trained end-to-end because of the model complexity. The authors in [27] propose a model for automatic wound region segmentation and wound condition analysis with infection detection and healing progress prediction. This study [27] utilises traditional pre- and post-processing steps to improve segmentation performance and does not have tissue classification. The authors in [28] provide a tool for segmenting and locating chronic wounds to facilitate bioprinting treatment using edge detection and segmentation algorithms. In [28], the authors utilise semi-automatic overall wound segmentation on a limited number of wound images. Pre-processing and feature extraction steps are used to improve the performance of the segmentation task.

The study in [29] proposed a framework for tissue classification based on the appearance and texture of the current and prior visual appearance of the chronic wounds. Pre-processing and feature extraction steps are used for the segmentation task.

In this study, the state-of-the-art GAN algorithm [30, 31] is utilised to develop a model that can classify and segment different wound tissue types simultaneously. Unlike previous studies that require pre- or post-processing steps that increases the model complexity, the proposed method provides wound detection and segmentation without implementing such additional steps. Hence, end-to-end training is possible. Furthermore, while many of the previous studies lack the segmentation of different wound tissues, this study provides segmentation of wound tissues; hence, important information related to wound healing status can be recognised. Additionally, the proposed novel approach could be applied to various wound types such as diabetic, pressure injury, and burn despite the prior studies.

The medical image synthesis using GAN for hybrid wound border segmentation and tissue classification has not been done previously. These two tasks are realised individually by the previous studies with a focus on one type of wound. The main contributions of this study include: (i) The development of a hybrid GAN algorithm to perform wound border segmentation and wound tissue classification in one step on different wound types, and (ii) provide guidance to healthcare researchers with respect to the number of images and epochs needed to perform successful medical image synthesis with GANs for various applications.

2 | METHODOLOGY

A GAN model comprises two neural networks, which are the generator (G), and the discriminator (D). Both generator and discriminator are concurrently trained with real data to capture the data distribution. A random uniform or a Gaussian noise (z) is fed to the generator network to produce fake images (y), G: $z \rightarrow y$ [32]. This makes the output of the generator unique. This newly created fake image is then fed to the discriminator network [32]. The discriminator network aims to determine if
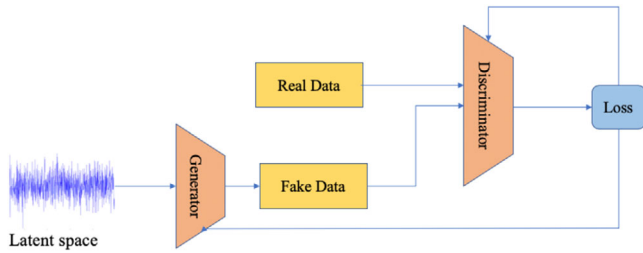
FIGURE 1 The basic structure of the generative adversarial network (GAN) model

FIGURE 2 The conditional GAN (cGAN) model overview

the generated image is from the training set or not, D: $y \rightarrow [0,1]$. The generated images are labelled as fake or real, depending on the training data distribution. The generator is to deceive the discriminator network that generated images are from the training set [33]. Figure 1 shows the basic structure of the GAN model.

Different versions of the GAN model are developed for different applications, such as conditional GAN (cGAN), cycle-consistent GAN (CycleGAN), Gaussian–Poisson GAN, and super-resolution GAN. Results of a cGAN-based model are evaluated in the scope of this study. The CycleGAN-based model was also examined under this study, but after initial trials, this model did not yield good results and suffered from the mode collapse issue, a well-known problem in the GAN field, which causes the generation of a particular output image regardless of different inputs [34], for border segmentation and tissue classification tasks. Since this approach failed to produce output results, it was discarded from this study.

Other deep learning-based segmentation methods [35], to our knowledge, have no evidence of their use to simultaneously perform wound segmentation and classification. Existing research performs, first, the segmentation step and consecutively the classification step. Hence, the proposed algorithm could not be compared to other work as a whole because of its novelty. On the other hand, the proposed novel model in this study accomplishes the border segmentation and tissue classification tasks simultaneously by utilising end-to-end training successfully. In addition to this, the border segmentation task performance is compared with the five different deep learning models, that is, VGG16, Segnet, U-Net, Mask-RCNN, and MobileNetV2, using the Dice coefficient metric.

The cGAN architecture has additional properties over the regular GAN architecture, which is also called vanilla GAN. cGAN gets an image as an input (x) in addition to the random noise (z), and generates an output (y) conditioned on that input image, G: $\{x, z\} \rightarrow y$. The generated image carries similar features with the input image while maintaining the data distribution of the training set, consisting of paired and aligned images. The mapping of input to the output images is learned by the generator network, where the discriminator network learns a loss function to train this mapping, D: $\{x, y\} \rightarrow [0,1]$ [36]. The objective of both generator and discriminator networks is the same as the vanilla GAN algorithm, with the difference that discriminator and generator observe the input image [37]. Figure 2 shows the general architecture of the cGAN model.
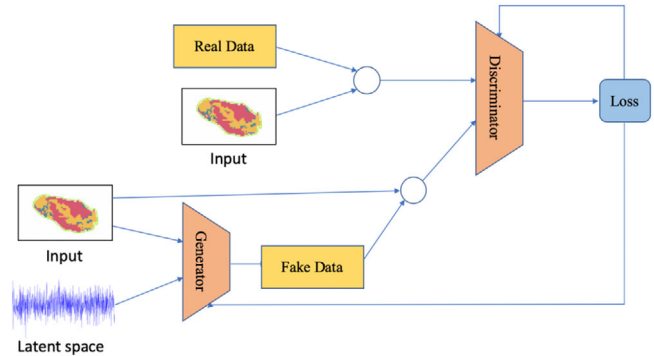
The cGAN model encapsulates two networks and four loss functions to generate plausible fake data. The discriminator networks are updated directly, but generator networks are trained by the feedback coming from the discriminator model while updating the loss function. Trained by the second model, the generator network lacks an objective function, which is the primary reason of GANs' hardship to train [38].

For the generator network, 'U-Net' encoder-decoder with skip connections architecture is utilised to get high resolution. The skip connection is a widely used method to keep the original data between the layers. The input is downsampled and flows through many layers, which concludes the input to a bottleneck. On the other hand, for the image translation, there should be some shared common features. That is why the cGAN is trained over paired and aligned data, which helps to predict the conditioned output. We used a 70 × 70 patch-wise comparison of images by discriminator network to classify the generated image as fake or real.

The discriminator network learns to classify real and fake images with binary cross-entropy loss. There are two loss functions to update the discriminator for real and fake samples, namely, D_real and D_fake. The generator network also has two different losses to provide plausible generated images. The weights of the generator model are then updated with adversarial loss (G_GAN) via the discriminator network and $L1$ loss (G_L1). $L1$ loss is calculated by comparing the generated images with the real image. Adversarial loss and $L1$ loss scores are combined to obtain the loss of the generator network and shown as

$$L_{Gen} = L_{Adv} + (L_{L1}) \times \lambda \tag{1}$$

where L_{Gen} is the generator network loss, L_{Adv} is the adversarial loss from the discriminator network, $L1$ is $L1$ loss, λ is the regularising hyperparameter. $L1$ loss serves as a regularising term in the generator network loss with a hyperparameter lambda, $\lambda = 100$.

The objective of adversarial loss of cGAN architecture can be depicted as:

$$L_{Adv}(G, D) = \mathbb{E}_{x,y}[\log D(x,y)]$$
$$+ \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \tag{2}$$

where the generator (G) competes with the discriminator while G is trying to minimise this objective, and D is trying to maximise it [38]. The final objective function can be expressed as shown in Equation (3).

$$G^* = \text{argmin}_G \text{max}_D L_{Adv} \ (G, D) + \lambda \times L_{L1} \ (G) \qquad (3)$$

3 | DATA COLLECTION, PRE-PROCESSING, ENVIRONMENT, AND VALIDATION

This section discusses data collection, data pre-processing, the simulation environment, and model validation.

3.1 | Data collection and preparation

The chronic wound data repository is provided by eKare Inc., which provides professional wound imaging and analysis services. Images are taken with commercially available cameras by regular users in a natural hospital environment on a normal wound assessment process at the clinic. The chronic wound images, including burn, pressure injury, and diabetic wounds, are semi-automatic segmented for training and testing purposes. The wound tissues are classified as necrotic, sloughy, and granulation, which are represented in blue, yellow, and red colours, respectively, in the segmentation task. The variety of wounds improves the applicability of the algorithm implemented in this study. In this study, anonymised wound images were rescaled to 512×512 pixels. To test the effect of the number of images by the GAN algorithm, we created a set of 100, 500, 1000, 2000, and 4000 images from 13,000 images containing different wound types. The test set was fixed to the same 100 images. Data augmentation, that is, flipping, is used. Some of the images used in this study can be seen below in the result section.

The number of publicly available chronic wound images is very limited and not sufficient for comparison of a training dataset of deep learning-based wound border segmentation and tissue classification tasks. Additionally, it is very challenging or impossible to find chronic wound images with ground truths. Another issue is related to the quality of the images. Medetec wound database [39] is a publicly available dataset that suffers degraded image quality because of the presence of mould growth on the original 35 mm transparencies. This will further decrease the resolution of generated images as well.

In contrast, the unique eKare Inc. chronic wound image repository provides us with a sufficient number of images, higher quality, and above all with ground truth data in order to sustain high-quality training.

3.2 | Environment

We implemented the wound border segmentation and tissue classification model using the PyTorch deep learning framework on the Anaconda platform with Python version 3.6. Our implementations ran on Intel® Core ™ i7 -7800X CPU @3.50 GHz with 16 GB RAM and NVIDIA GeForce GTX 1080 GPU with 8 GB dedicated and 8 GB shared memory. We trained our model 2000 epochs using 100, 500, 1000, 2000, and 4000 images, which took 4, 9, 20, 42, and 76 hours, respectively. The batch size is chosen as 64 to increase the benefit from the GPU. We used a constant learning rate of 0.0002 and 'Adam' optimiser for the first half of the training. The rest of the training was done with a linearly decaying learning rate to zero until convergence.

3.3 | Validation

Validation was done using the mean squared error (MSE) and Dice coefficient metrics for the evaluation of (generated) fake image quality.

MSE, which is a pixel-wise loss function, was used to measure the quality of the generated images in addition to losses of GAN. Minimising the pixel-wise error measurement provides converging results in contrast to GAN loss. Generated segmented images are expected to be very similar to the actual segmented images. In addition, segmented images consist of a combination of three colours, which makes it easy to compare. That is why the MSE metric fits properly for the evaluation of this similarity. MSE score was calculated by comparison of real and fake images on pixel level in three colour channels. MSE metric can be written as

$$\text{MSE} = 1/n \sum_{i=1}^{n} \left[\left(Y_R - Y'_R \right)^2 + \left(Y_G - Y'_G \right)^2 + \left(Y_B - Y'_B \right)^2 \right] \qquad (4)$$

where

n: Number of pixels
Y_R, Y_G, Y_B: RGB values of the real images
Y'_R, Y'_G, Y'_B: RGB values of generated images.

Dice coefficient is used to evaluate the performance of the proposed method in addition to the MSE metric. The harmonic means of recall and precision provides a Dice coefficient, which is also known as the F1-score and is calculated as follows:

$$Dice \ Coefficient = \frac{2 |A \cap B|}{|A| + |B|} \qquad (5)$$

where A and B are the ground truth and model output, respectively. Dice scores range from 0 to 1 where a score of 1 indicates a perfect segmentation.

4 | RESULTS AND DISCUSSION

This section discusses the output of the model, loss graphs, the effect of epoch on border segmentation, tissue classification, and the optimum training conditions of the model.
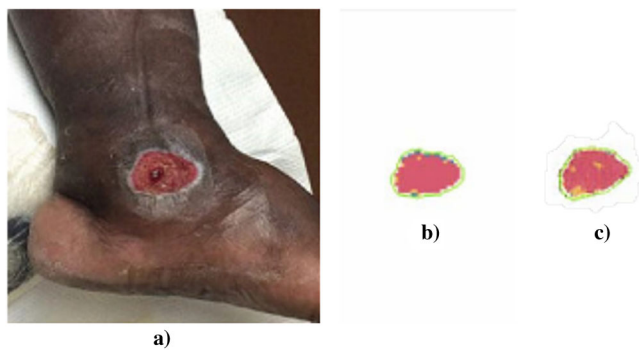
FIGURE 3 cGAN model achieves good results with 2000 images at 200 epochs: (a) Original image, (b) segmentation ground truth, and (c) generated segmented wound

4.1 | Model output

The output of the proposed method was compared with the ground truth. A successful result from the model is given in Figure 3, which indicates a proper border segmentation and tissue classification of the wound by training with 2000 images and 200 epochs.

As shown in Figure 3, the proposed model successfully segments the wound border and classify the wound tissue concurrently. The model learned the wound area in Figure 3, where there are paled areas around the heel and the side of the foot. The model is insensitive to colour changes and could identify the wound in a crowded environment. The background is discarded as well.

4.2 | Effect of number of images on model loss

The loss curves of cGAN are depicted in Figures 4 and 5 when trained with 100, 500, 1000, 2000, and 4000 images, respectively. The G_L1 loss has the most meaningful loss for the generated image quality. G_L1, G_GAN, D_real, and D_fake losses oscillate because the GAN model moves from one type of sample generation to another type of generation before reaching a balance [40]. Training two opposing neural networks concurrently in zero-sum game results in a non-converging problem [40]. G_L1 represents the generator loss only, and it lacks the contribution of the adversarial loss. G_L1 loss could be used only for determining the learning capability of the proposed model with respect to dataset size. However, the training progress is unpredictable from the loss alone. That is why an additional technique is needed to predict the progress of training and the quality of generated images.

A comparison of loss graphs in Figures 4 and 5 reveals the drop rate of the G_L1 loss increases with an increasing number of images for the training. The G_L1 loss drops to 10 at around the 40th epoch and stays stable under five around the 100th epoch with a training set of 100 images (see Figure 4(a)). The drop rate of G_L1 loss increases in Figure 4(b), which is the model loss with a dataset of 500 images. The loss of the
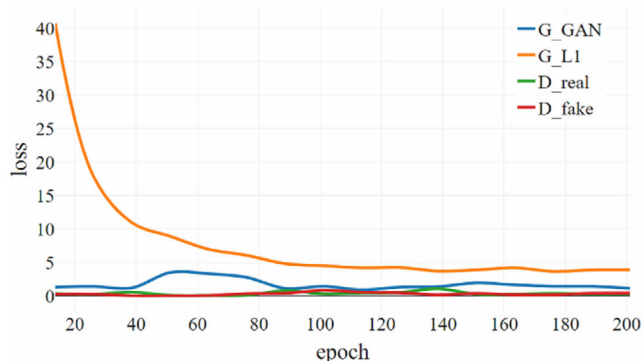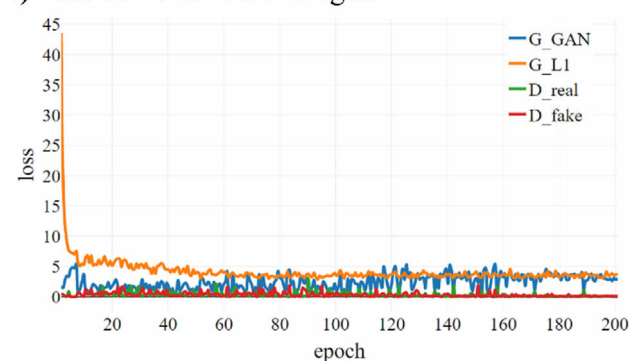

a) Loss curve for 100 images.
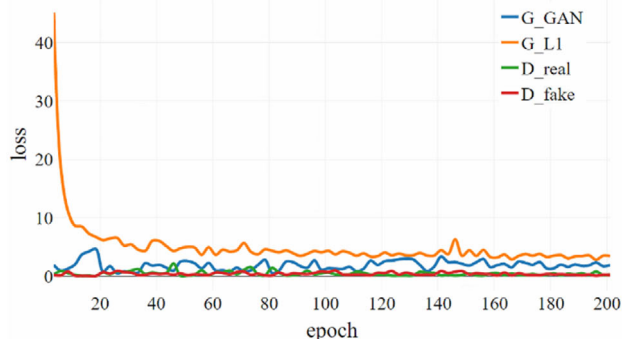
b) Loss curve for 500 images.
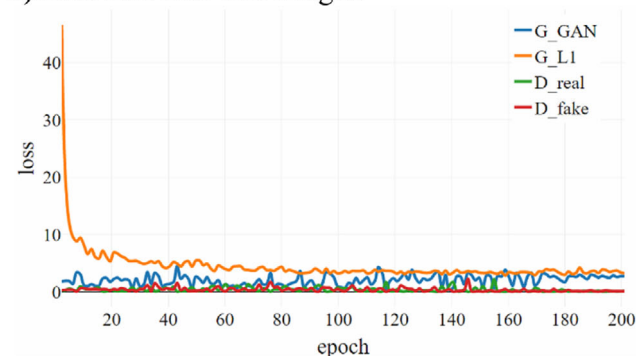
c) Loss curve for 1000 images.

d) Loss curve for 2000 images.

FIGURE 4 Model's loss graphs using a different number of images

proposed model reaches 10 at the 10th epoch and stabilises under five at the 75th epoch. These are four times and one-and-a-half times improvement over the model with 100 images, respectively. A five-fold increase in the training set provides a similar increase in the G_L1 drop rate. The drop rate further
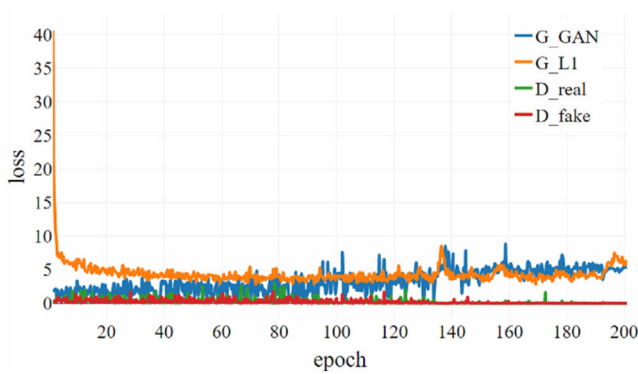
FIGURE 5   Loss curve for the model with 4000 training images

FIGURE 7   Effect of the epoch count on border segmentation and tissue classification tasks

increases two and four times, respectively, in the models trained with 1000 and 2000 images as shown in Figures 4(c) and (d). The results of the model trained with 4000 images are consistent with Figures 4(b)–(d). G_L1 loss converges to 10 in the first couple of epochs and becomes stable under five (5) in 15 epochs. The results indicate that our proposed model learns the data distribution faster with a larger dataset.
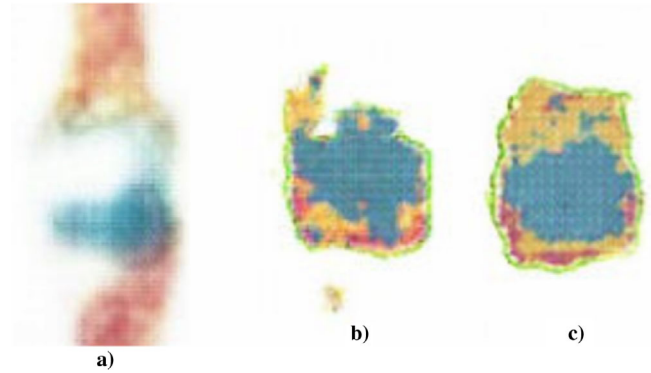
The loss curves for the model with 4000 training images are shown in Figure 5.

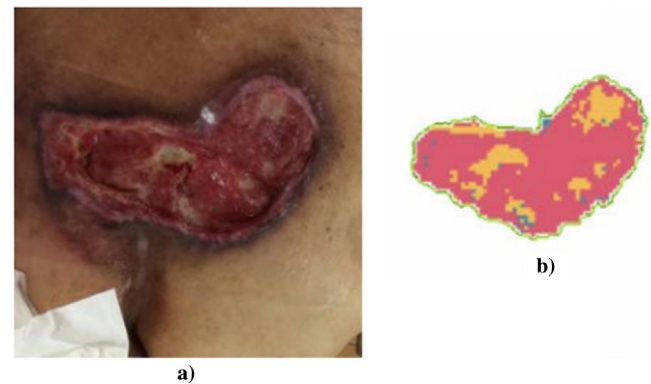## 4.3 | Effect of number of epochs on border segmentation and tissue classification

Figures 6(a) and (b) depict the original wound image and the ground truth wound tissue classification, respectively. By varying the number of epochs, Figures 7(a) to (c) show the results of the model with a dataset of 2000 images, which is trained using 5, 75, and 200 epochs, respectively. With five epochs, as shown in Figure 7(a), the result does not represent the segmented wound, although it carries similar tissue features. Training the model further to 75 epochs in Figure 7(b) provides a better representation. Adequate generation of the wound seg-

FIGURE 8   (a) Original image, (b) ground truth for border segmentation and tissue classification tasks

mentation is achieved after 200 epochs of training as shown in Figure 7(c).

Figure 8 depicts the original wound image and the ground truth for wound tissue classification. Figure 9 illustrates the output of the model with a dataset of 100 images after it was trained for 5, 500, 1000, and 2000 epochs. Training the model for five epochs produces a similar shape but a blurry result in Figure 8(a), which indicated that the model could not get the data distribution yet. An increase in the epoch count generates better-segmented wound images, but these results could not catch the wound shape as a result of inadequate training images.

The results were also analysed using MSE scores as summarised in Table 1, which shows the MSE values for a different number of epochs and training images. MSE score is a good indicator of the model's learning ability to mimic the real image data distribution. The MSE score of the model trained for five epochs with 100 images is the highest and improves with the increase in the number of training images and epochs.

MSE values of the different numbers of images are shown in Figure 10. The model trained with 100 images dataset did not yield efficient results and was omitted for simplicity in Figure 10. The decrease in the MSE score in the first 200 epochs is the highest for all dataset sizes. The dramatic decline in the first 200 epochs indicates that the proposed method successfully learns to segment the wound and classify the tissue type at

FIGURE 6   (a) Original image, and (b) ground truth for border segmentation and wound tissue classification tasks

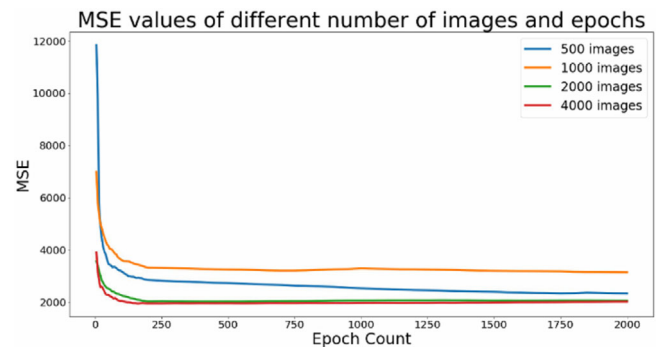**TABLE 1**   Mean squared error (MSE) scores for the conditional generative adversarial network model

Epochs / Images	5	200	500	1000	2000
100	40737	4094	3911	3695	3329
500	11843	2847	2730	2528	2333
1000	6983	3319	3251	3293	3144
2000	3565	2027	2024	2061	2056
4000	3907	1958	1955	1968	2019
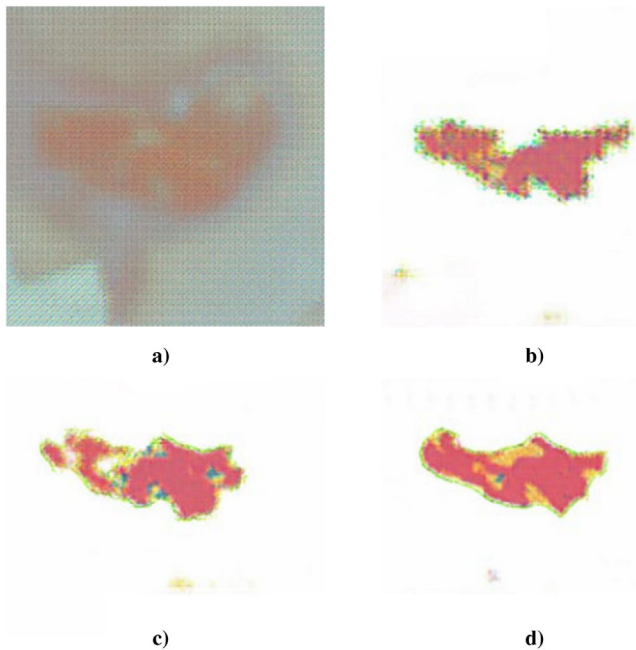
FIGURE 9   The model trained with 100 images is not converging to synthesise border segmentation and tissue classification. (a) Generated image at the 5th epoch, (b) generated image at the 500th epoch, (c) generated image at the 1000th epoch, and (d) generated image at the 2000th epoch

FIGURE 10   Mean squared error (MSE) versus epoch for the different number of images

TABLE 2   The decrease in MSE score between different numbers of epochs

Epochs / Images	5 to 200 epochs	200 to 2000 epochs
100	90.0 %	18.7%
500	76.0%	18.1%
1000	52.5%	5.3%
2000	43.1%	−1.4%
4000	49.9%	-3.1%

200 epochs. Results confirm that increasing epoch count results in a better MSE value for the first 200 epochs, and training for more epochs decrease MSE values slightly. This is a good indicator of optimum training parameter selection that *200 epochs are the optimum epoch number for training the proposed method.*

Note that the model trained with 500 and 1000 images has an equilibrium around 3000 MSE score, and the model trained with 500 images keeps decreasing, which could be a result of a limited number of images representing a few samples and overfitting that data. There may be a potential overfitting problem. The model trained with 2000 and 4000 images share similar MSE values of around 2000. The outcome in Figure 10 indicates that increasing the number of images for training produces lower MSE values, which is a good sign that the proposed model works as expected.

The changes in MSE values with 5–200 and 200–2000 epochs are compared for different datasets, that is, 100–4000 training images as shown in Table 2. It appears that the MSE value

improves significantly during the first 200 epochs of training. Training from 200 to 2000 epochs improves MSE slightly. The changes in MSE score with a smaller number of training images, that is, 100 and 500, are higher than those with a higher number of training images, that is, > 1000. This is because the model with a larger dataset converges faster during the first couple of epochs. Negative values imply the increase in the MSE metric, which results from the deformation of generated images and noises in input images.

The Dice coefficient is also calculated for each epoch and image combination. The Dice scores are shown in Table 3 for further analyses. The correlation between MSE score and Dice coefficient indicates that the model with 2000 images and 200 epochs is the best performing model, requiring a lower number of images and epochs. The differences between MSE scores and the Dice scores are sourced from the calculation of both metrics. The MSE metric considers both overall wound

TABLE 3 Dice coefficients of the proposed model

Images / Epochs	5	200	500	1000	2000
100	0.09	0.77	0.62	0.84	0.73
500	0.09	0.77	0.89	0.88	0.86
1000	0.18	0.74	0.83	0.76	0.82
2000	0.79	0.90	0.88	0.85	0.78
4000	0.42	0.93	0.93	0.91	0.79

TABLE 4 The comparison of Dice coefficients of the proposed model and other models [41]

Model	VGG16	Segnet	U-Net	Mask-RCNN	MobileNetV2	Proposed model
Dice Score	0.81	0.85	0.90	0.90	0.90	0.90

segmentation and the segmentation of the wound tissues. It provides more information about the segmentation performance.

The Dice coefficient metric provides a measurement of wound area segmentation performance regardless of the wound tissue. The models with a lower number of training dataset images, that is, 100 and 500, do not provide higher scores as expected. The Dice coefficient of the model with 1000 training images increases with an increasing number of epochs. The 2000-image model's Dice score is also in line with its MSE score. The 4000-image model has the highest performance metrics, whereas the required number of images doubles compared to the 2000-image model.

The comparison of the proposed model with the previous works is shown in Table 4. Five different previous models are compared with the proposed model.

The comparison indicates that the proposed model has similar performance with other highest performing models. In addition to wound segmentation, the proposed model provides tissue classification and respective segmentation of tissues as well. That is why the proposed model has not only good segmentation performance but also tissue classification capability as well.

4.4 | Effect of number of images on border segmentation and tissue classification

Figure 11 depicts the original wound image and the ground truth for wound tissue classification. The fixed number of epochs at 200 and the results of the models with different training datasets are shown in Figure 12. Input datasets that have fewer than 500 images give poor performance. Therefore, they are excluded from Figure 12. As shown, the proposed method provides efficient segmentation and tissue classification on a dataset consisting of around 2000 images or more. It is a significant conclusion that *having at least 2000 images at hand results in efficient training for GAN to generate qualified images in this study*. Smaller datasets face difficulties in mimicking the data distribution or these models overfit the training images, which is the

case for the model with a training set of 500 images or less, whereas datasets with higher than 2000 images generate plausible images.

4.5 | Discussion

Based on the study results, the following observations regarding the application of the GAN algorithm could be made.

Observation 1: The proposed method can perform both wound border segmentation and tissue type classification in one step.

Observation 2: cGAN has a high potential of producing close to real synthetic images for wound tissue segmentation and classification.

Observation 3: The quality of the generated images are in line with the image count; 2000 image count is the threshold for a valid generated image as the result of our study.

FIGURE 11 (a) Original and (b) ground truth of tissue classification images

a) 500 images   **b) 1000 images**

c) 2000 images   **d) 4000 images**
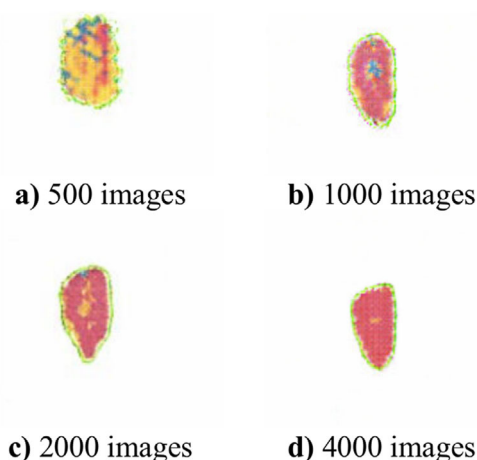
FIGURE 12   Effect of training dataset size at 200 epochs

Observation 4: The epoch count has a significant impact on the generated image quality, but after surpassing the 200-epoch threshold, the model reaches its convergence, and additional training has a marginal effect on the quality of the generated image.

5 | CONCLUSION

This study presents that the cGAN algorithm can achieve chronic wound border segmentation and tissue classification efficiently. The wound border segmentation and the wound tissue type classification using GANs were performed for the first time. Results from different numbers of dataset sizes and epoch counts are evaluated through the MSE metric and visual inspection of generated images. MSE metric provides valuable information in interpreting the quality of the generated segmentation and classification tasks due to the simplicity of the generated images. The optimum training dataset size and epoch count are determined at 2000 images and 200 epochs. This study confirms that the generated image quality increases significantly by increasing the dataset size to 2000 images. After that threshold, the image quality improves marginally. Currently, the data collection in healthcare is an expensive task and process; this study introduces the optimum dataset size for related healthcare applications utilising GAN. The proposed method achieves border segmentation and tissue classification simultaneously without additional processing steps and expertise. The MSE score decreases and the Dice coefficient increases with the increase in generated segmented image quality. The proposed model is in line with these conditions, which are explained in the validation section. The ability to perform end-to-end training and testing ability simplifies the application of the proposed model in healthcare for broader adoption. However, the healthcare industry requires robust and explainable models that will require adopted models to be transparent. The proposed method and deep learning models in general lack transparency and behave as a black box.

The scope of this study includes detection of the various wound types such as burn, lymphovascular, pressure injury, and

classification of three different wound tissues, namely, necrotic, slough, and granulation. Some limitations in this study could be further addressed in the future work. First, the image quality of the overall model could be further improved. The image quality selected for this study is to provide a fast and straightforward implementation, which is the case for many algorithms in the object detection and segmentation field. This is also due to the resolution of available datasets. Since images were collected by various cameras with different settings, it is necessary to format them to a common size for further processing. Second, due to the non-converging nature of the GAN algorithm, the loss curves of our model have also limitations providing the relationship between the training and the generated image quality. That is why the hyperparameter optimisation was performed by observing both the generated images and the loss curves together.

Possible future work may include the modification of the algorithm to generate high-resolution images. The structure of the proposed algorithm resizes the images to 512 × 512 pixels. With model modification, the generated image quality may increase to 2048 × 2048 pixels. Another future research direction can be the consideration of an additional class of tissues, that is, bones or foreign objects such as metal fixations in the wound. This will enhance and increase the use cases of this model. The next iteration of this model may identify wound etiology, such as diabetic, lymphovascular, pressure injury, and surgical. Identification of wound type will enhance wound management further by determining the right wound care plan.

It is expected that this study will help caregivers in deciding the wound treatment plan by understanding the wound tissue classification visually as well as assist researchers in providing an insight into the wound border segmentation and tissue classification through advanced machine learning methods.

ORCID
Salih Sarp 🄳 https://orcid.org/0000-0001-6674-2590

REFERENCES
1. Stern, D., Cui, H.: Crafting polymeric and peptidic hydrogels for improved wound healing. Adv. Healthcare Mater. 8(9), 1900104 (2019)
2. Wu, K.L., et al.: Feasibility Of 3D structure sensing for accurate assessment of chronic wound dimensions. Int. J. Comput. Assisted Radiol. Surg. 10(1), 13–14 (2015)
3. Zahia, S., et al.: Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods. Artif. Intell. Med. 102, 101742 (2020)
4. Mehl, A.A., et al.: Measurement of wound area for early analysis of the scar predictive factor. Revista Latino-Americana de Enfermagem 28, e3299 (2020)
5. Yu, K., et al.: Machine learning applications in the evaluation and management of psoriasis: A systematic review. J. Psoriasis Psoriatic Arthritis 5, 147–149 (2020)

6. Salvaris, M., Dean, D., Tok, W.H.: Generative adversarial networks. In: Deep Learning with Azure, pp. 187–208.Apress, New York (2018)

7. Pinetz, T., Ruisz, J., Soukup, D.: Actual Impact of GAN Augmentation on CNN Classification Performance. In: Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM), Prague, Czech Republic, pp. 15–23 (2019)

8. Mescheder, L., Geiger, A., Nowozin, S.: Which training methods for GANs do actually converge? arXiv:1801.04406 (2018)

9. Young, S.R., et al.: Optimizing deep learning hyper-parameters through an evolutionary algorithm. In: Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments, Austin, Texas, pp. 1–5 (2015)

10. Reimers, N., Gurevych, I.: Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks. arXiv:1707.06799 (2017)

11. Roh, Y., Heo, G., Whang, S.E.: A survey on data collection for machine learning: A big data-ai integration perspective. IEEE Trans. Knowl. Data Eng. 1-1, (2019)

12. Martin, L., et al.: How much do clinical trials cost? Nature Rev. Drug Discovery 16(6), 381–382 (2017)

13. Kostkova, P., et al.: Who owns the data? Open data for healthcare. Front. Public Health 4, 7 (2016)

14. Smith, L.N.: A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. arXiv:1803.09820 (2018)

15. Poon, T.W.K., Friesen, M.R.: Algorithms for size and color detection of smartphone images of chronic wounds for healthcare applications. IEEE Access 3, 1799–1808 (2015)

16. Veredas, F.J., et al.: Wound image evaluation with machine learning. Neurocomputing 164, 112–122 (2015)

17. Thatcher, J.E., et al.: Imaging techniques for clinical burn assessment with a focus on multispectral imaging. Adv. Wound Care 5(8), 360–378 (2016)

18. Filko, D., Cupec, R., Nyarko, E.K.: Wound measurement by RGB-D camera. Mach. Vision Appl. 29(4), 633–654 (2018)

19. Pholberdee, N.: Wound-region segmentation from image by using deep learning and various data augmentation methods. Dissertation, Silpakorn University (2019)

20. Chino, D.Y., et al.: Segmenting skin ulcers and measuring the wound area using deep convolutional networks. Comput. Methods Programs Biomed. 191, 105376 (2020)

21. Zahia, S., et al.: Tissue classification and segmentation of pressure injuries using convolutional neural networks. Comput. Methods Programs Biomed. 159, 51–58 (2018)

22. Chakraborty, C.: Computational approach for chronic wound tissue characterization. Inf. Med. Unlocked 17, 100162 (2019)

23. Lu, H., et al.: Wound intensity correction and segmentation with convolutional neural networks. Concurrency Comput.: Pract. Experience 29(6), e3927 (2017)

24. Cui, C., et al.: Diabetic wound segmentation using convolutional neural networks. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, pp. 1002–1005 (2019)

25. Godeiro, V., et al.: Chronic wound tissue classification using convolutional networks and color space reduction. In: 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), Aalborg, Denmark, pp. 1–6 (2018)

26. Li, F., et al.: A composite model of wound segmentation based on traditional methods and deep neural networks. Comput. Intell. Neurosci. 2018, 1–12 (2018)

27. Li, F., et al.: Wound segmentation network based on location information enhancement. IEEE Access 7, 87223–87232 (2019)

28. Gholami, P., et al.: Segmentation and measurement of chronic wounds for bioprinting. IEEE J. Biomed. Health. Inf. 22(4), 1269–1277 (2017)

29. Khalil, A., et al.: Chronic wound healing assessment system based on different features modalities and non-negative matrix factorization (NMF) feature reduction. IEEE Access 7, 80110–80121 (2019)

30. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM 63(11), 139–144 (2020)

31. Zhao, J., Mathieu, M., LeCun, Y.: Energy-based generative adversarial networks. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France (2019)

32. Huang, H., Yu, P.S., Wang, C.: An introduction to image synthesis with generative adversarial nets. arXiv:1803.04469 (2018)

33. Dosovitskiy, A., Brox, T.: Generating images with perceptual similarity metrics based on deep networks. In: Advances in Neural Information Processing Systems, Barcelona, Spain, pp. 658–666 (2016)

34. Zhu, J.Y., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, pp. 2223–2232 (2017)

35. Minaee, S., et al.: Image segmentation using deep learning: A survey. arXiv:2001.05566 (2020)

36. Bakkay, M.C., et al.: BSCGAN: Deep background subtraction with conditional generative adversarial networks. In: 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, pp. 4018–4022 (2018)

37. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, pp. 1125–1134 (2017)

38. Salimans, T., et al.: Improved techniques for training gans. In: Advances in Neural Information Processing Systems, Barcelona, Spain, pp. 2234–2242 (2016)

39. Thomas, S.: Medetec wound database. http://www.medetec.co.uk/files/medetec-image-databases.html (2019)

40. Goodfellow, I.: NIPS 2016 tutorial: Generative adversarial networks. arXiv:1701.00160 (2016)

41. Wang, C., et al.: Fully automatic wound segmentation with deep convolutional neural networks. Sci. Rep. 10(1), 1–9 (2020)