The Visual Computer (2025) 41:11885–11908
https://doi.org/10.1007/s00371-025-04133-y

REVIEW

Deep learning in chronic wound segmentation: a comprehensive review and meta-analysis

Bill Cassidy[1] · Connah Kendrick[1] · Neil D. Reeves[2] · Joseph M. Pappachan[3] · Moi Hoon Yap[1]

Accepted: 2 August 2025 / Published online: 26 August 2025

Abstract

This work conducts a review of all chronic wound segmentation deep learning studies meeting specific criteria that have been published since research first started in this domain 8 years ago (2015–2023). Management of chronic wounds represents a serious ongoing concern for hospitals and outpatient clinics world-wide. There is a clear need for technological interventions using deep learning approaches that could have a potential significant impact in the automated monitoring of such wounds. We review the existing literature and perform R-squared statistical analysis to form a fresh understanding of the field to gain deeper insights into the issues that are presenting obstacles to research progress. Our findings show a negative correlation between small test set size and test metrics (Dice similarity coefficient and mean intersection over union), indicating smaller test sets are associated with higher test metrics. We also identify other major hurdles in the field, such as a lack of data understanding, a lack of data availability, and a lack of research transparency. The focus of this body of work is to increase understanding of the underlying issues that have pervaded in deep learning chronic wound research. A clear presentation of findings in this work can be used by researchers as a guide to avoiding common pitfalls, and to advance research knowledge.

Keywords Chronic wounds · Segmentation · Deep learning · Pressure ulcers · Venous ulcers · Arterial ulcers · Diabetic foot ulcers

1 Introduction

Chronic wounds are a growing burden on health care systems globally. The incidence of chronic wounds is substantial and is estimated to continue on an upward trend [1]. Diabetic foot ulcers (DFU) and arterial leg ulcers (ALU) are costly and debilitating complications of diabetes [2], with recent research suggesting an association between DFU episodes and all-cause resource utilisation and mortality [3]. Pressure ulcers (PRU) and venous leg ulcers (VLU) are the most common types of complex skin ulcers [4], with ulcer prevalence in the diabetic population estimated to be at least 13% in North America [5].

DFU has a global prevalence of approximately 6.3% among people with diabetes [6], with VLU estimated to have a prevalence of around 1.08% [7]. Global PRU prevalence is estimated to range between 5.2 and 12.3% [8]. However, these figures are likely to be higher as cases are often underreported, especially in lower-income countries where epidemiology data can be scarce and reporting may be inconsistent [9–12].

Occurrence of chronic wounds is often linked with comorbidities including vascular deficits, diabetes, chronic kidney disease, and hypertension [13]. Diabetic peripheral neuropathy is prevalent in the majority of DFU cases and is the main cause of DFU [14]. This condition results in nerve damage

✉ Bill Cassidy
bill.cassidy@stu.mmu.ac.uk

Connah Kendrick
connah.kendrick@mmu.ac.uk

Neil D. Reeves
n.d.reeves1@lancaster.ac.uk

Joseph M. Pappachan
pappachan.joseph@lthtr.nhs.uk

Moi Hoon Yap
m.yap@mmu.ac.uk

[1] Department of Computing and Mathematics, Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK

[2] Faculty of Health and Medicine, Lancaster University Medical School, Health Innovation Campus, Lancaster LA1 4YW, UK

[3] Lancashire Teaching Hospitals NHS Foundation Trust Sharoe Green Lane, Preston PR2 9HT, UK

in the foot leading to a loss of sensation [15]. Patients suffering from this condition may go through prolonged periods of damage to their foot without realising it. Wound condition can worsen, leading to other serious complications. More than 50% of all DFU cases experience infection [16] and is one of the main causes of hospitalisation for diabetic patients [14]. Diabetic leg and foot ulcers are some of the most expensive chronic wound types to treat in the USA [13]. Up to 70% of VLU reoccur within 3 months after wound closure [17].

Patients diagnosed with DFU are up to three times more likely to die compared to patients without the disease, and are at risk to numerous comorbidities, including cardiovascular disease, nephropathy, neuropathy, peripheral arterial disease, and diabetic retinopathy. DFU and VLU often lead to significantly impaired quality of life [17–19]. Occurrence of ulcers is linked to increased risk of both amputation and mortality, particularly when associated with advanced age, anaemia, and peripheral arterial disease [17, 20, 21]. Chronic wounds place a significant emotional and physical burden on patients [22, 23], with depression associated with increased risks at initial and subsequent wound occurrence [24, 25].

Chronic wound management represents a major healthcare system cost and a significant time burden for clinicians and patients. This is particularly true for chronic wounds that are not diagnosed at an early stage and require more intensive treatment methods. Such situations may occur as a result of infection, with worse-case scenarios potentially leading to amputation [20]. These cases can result in frequent clinic and hospital visits for expert assessment [26, 27]. Even after chronic wounds have healed, recurrence rates are high, with minor or major amputation of lower extremities being common [28, 29]. The post-COVID-19 world poses further challenges and risks in the treatment of chronic wounds, especially for diabetic patients who are at higher risk [30–32].

New technologies to address growing clinical needs are becoming ever more prevalent in numerous medical fields [33, 34]. To address the issues associated with chronic wound prevalence, there has been an increase in research interest for fully automated non-contact remote detection and monitoring of chronic wounds [35–37]. Enhancing telemedicine systems to include automated monitoring of wounds can help to reduce risks to vulnerable patients and to ease pressures on overburdened healthcare systems [38]. Furthermore, the growing popularity of low-cost consumer mobile phones allows for these technologies to be distributed in poorer countries and rural areas where patients may have restricted access to healthcare settings.

Non-invasive easy-to-use devices capable of automated detection and monitoring could help to promote patient engagement in monitoring chronic wounds [35] which may help to reduce clinic and hospital visits. Recent scientific evidence shows that convolutional neural networks (CNNs) can be equal to, and in some cases surpass experienced dermatologists for detection and classification in medical domains [39–45]. Wound area changes over time have been shown to provide robust prediction in healing status [46]. Chronic wound segmentation allows for potential assessment of wound development and therefore healing status over time, providing superior accuracy when compared to object localisation techniques, which give a more general indicator of wound development [47].

Subjectivity in medical imaging domains is a challenging aspect of deep learning. Ground truth labelling of wound photographs requires clinical experts to manually delineate the wound regions before segmentation models can be trained, validated, and tested. For such procedures, there are currently no formal standards. Ramachandram et al. [48] found low inter-rater agreement for tissue types in wound images in a recent wound segmentation study. Their results showed Krippendorff alpha values as low as 0.014 for epithelial tissue. This issue is common in deep learning tasks throughout medical imaging domains, including MRI image quality assessment [49], dermoscopic skin lesion evaluation [50–52], and ultrasound diagnosis [53]. Accurate automated delineation of wound regions may also potentially be used as an assisting tool for clinicians to aid in the monitoring of healing progress. By limiting human subjectivity, such advances could help to reduce hospital/clinic burdens when treating patients.

Deep learning research in chronic wound segmentation is a relatively new domain. Early attempts to segment wounds involved the use of traditional computer vision techniques. [54] trained a cascaded two-stage classifier using two state vector machine (SVM) classifiers. Colour and texture descriptors are extracted from superpixels which are used for the classifier training. Colour and wavelet features were used as feature descriptors to distinguish wounds from healthy tissue. The final stage refines the wound boundary using conditional random field statistical modelling. Nonnegative matrix factorisation was utilised by [55]. Their factorisation segmentation approach was used to extract wound bed features from processed wound images. Local spectral histograms were then generated by convolving a filter bank. For each image pixel, local spectral histograms were calculated to construct the feature matrix.

This paper summarises almost all of the deep learning chronic wound segmentation papers published since 2015. The objective of this work is to investigate the relationships between test set sizes used in experiments and corresponding Dice similarity coefficient (DSC) and intersection over union (IoU) test metrics.

2 Methodology

Inclusion criteria for published articles were as follows:

1. Studies involving deep learning architectures used to segment DFU, PRU, or VLU wounds.
2. Written in English.
3. Clearly states the number of images used in train and test sets.
4. Quantified DSC or IoU metrics in test results.
5. If the paper was part of an online challenge event, such as the Diabetic Foot Ulcer Challenge 2022 (DFUC 2022) [56], then only the winning paper was selected for review.

Exclusion criteria for published articles that were not included in our review were as follows:

1. Article was not published in a journal, conference, or workshop.
2. Reported on wound segmentation results only from animals.
3. Focuses on only burn wounds. Due to the paucity of publicly available burn wound datasets, we do not include papers that focus only on those wound types in this review.

The Google, Google Scholar, Research Gate, and PubMed search engines were used to locate relevant publications. Search phrases used were: "wound segmentation", "dfu segmentation", "ulcer segmentation", "pressure segmentation", and "venous segmentation". Additional terms, such as "deep learning", "image segmentation", "computer vision", and "machine learning", in addition to Boolean operators, were also used as search terms. Both paper title and main body text were searched when filtering results. Figure 1 shows the process of identifying relevant studies for inclusion in this review.

3 Semantic segmentation

In this section, we discuss the most prominent chronic wound segmentation papers that focused on semantic segmentation. Semantic segmentation is defined as pixel-based segmentation of wound pixels in an image. This is distinct from instance segmentation, which involves per-wound based detection that identifies individual wound cases in an image. The vast majority of chronic wound segmentation papers use semantic segmentation, a likely consequence of the wider availability of source code and architectures that target this method. Higher computational costs are also associated with instance segmentation which may prove to be a limiting factor in model selection [57–59].

Wang et al. [60] trained a chronic wound segmentation CNN consisting of five encoding layers followed by four decoding layers with rectified linear unit (ReLU) activations, cross-entropy for the loss function, and L2 regularisation with a regularisation coefficient. They used 500 training images and 150 test images sourced from the NYU Wound Database. A modified version of GrabCut [61] was used to crop images to 480 × 640 pixels to reduce non-wound background features. They reported a mean intersection over union (mIoU) of 0.473, with 0.950 pixel accuracy on the test set. This work is notable for being among the first to address chronic wound segmentation using deep learning techniques.

Goyal et al. [62] trained a selection of FCN models to segment DFU wounds and periwounds using a dataset of 600 DFU images with delineated masks provided by clinical experts. Two-tier transfer learning was completed using ImageNet and the Pascal VOC segmentation dataset. The DFU dataset was split into 420 images for training, 60 images used for validation, and a test set comprising 120 DFU images and 105 healthy foot images. For combined segmentation of wound and periwound regions, the best reported model was FCN32-s with a DSC of 0.899. For segmentation of only ulcer regions, the best model was FCN-16 s with a DSC of 0.794. For segmentation of only periwound regions, the best model was FCN-16 s, with a DSC of 0.851. This work observed that FCN-AlexNet and FCN-32 s were not able to accurately segment irregular boundaries. Conversely, they noted that smaller pixel strides used in FCN-16 s and FCN-8 s resulted in improved segmentation of irregular DFU wound contours. They also noted in test image results that wound and periwound features would overlap due to ambiguous feature boundaries. This work is notable for being one of the few deep learning wound segmentation studies to include periwound features (see Fig. 2).

Li et al. [63] proposed a composite model using watershed and thresholding in pre- and post-processing stages to assist in the removal of non-wound features. They trained a segmentation network comprising 13 convolutional layers of MobileNet as the backbone, where a pair of depthwise and pointwise layers is considered as a single layer. The last convolutional layer is upsampled and fused with the previous convolutional layer, followed by a pooling layer to reduce loss of local information. The fusion result is then upsampled 16 times to ensure that the output is of equal resolution to the input. A final post-processing stage is used to perform morphological operations (hole-filling and small region removal) and additional thresholding to remove any persisting background features. For training and testing, a dataset of 950 images was used, of which 389 were collected from patients in hospital settings and 561 images were sourced from the internet. A total of 760 images were used for training and

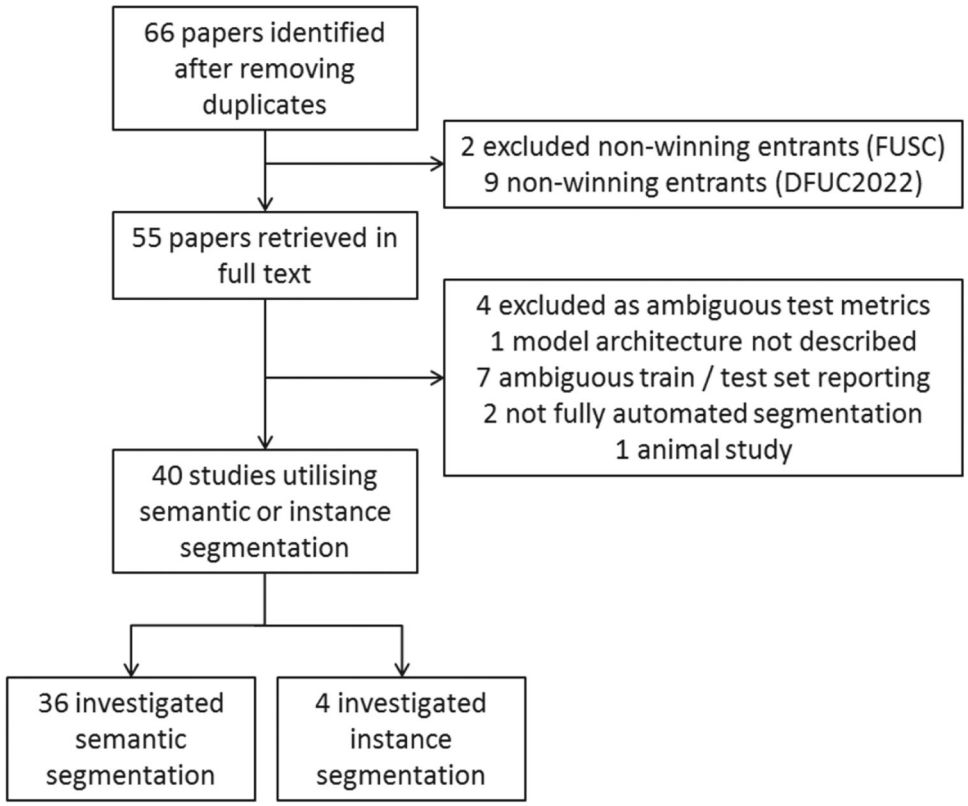


Fig. 1 Study selection flow diagram

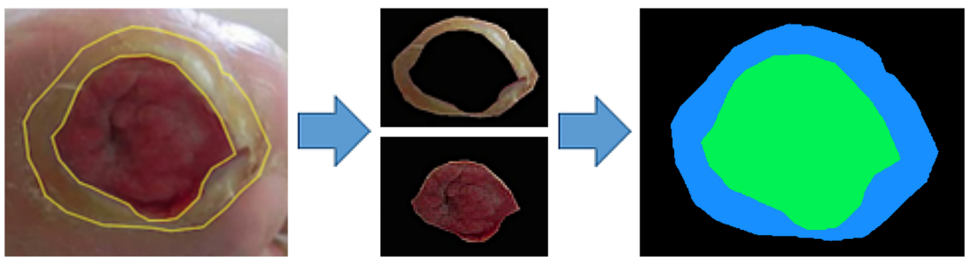


Fig. 2 Illustration of periwound delineation and separation from wound region, originally reported by Goyal et al. [62]

190 used for testing. They reported an mIoU of 0.8589 and precision of 0.9470.

Elmogy et al. [64] proposed a framework to segment and classify tissue types (slough, necrotic eschar, and granulation) in PRU wounds. First, a region of interest extractor (DeepMedic [65], a 3D CNN) uses three different colour spaces (RGB, HSV, and YCbCr) to reduce background details. Next, tissue segmentation is performed using the different spatial outputs from the region of interest models. The framework was trained and tested on 100 PRU images—36 images sourced from a healthcare services company (IGURCO GESTION S. L., Spain), and 64 images from the Medetec [66] dataset. The dataset was split into 60 images for training, 10 images for validation, and 30 images for testing. They reported a DSC of 0.93. They also reported results from a Bland-Altman analysis to assess the degree of agreement between the ground truth delineation of three observers. The results of this analysis showed that most ratings were within the range of $m = \pm 1.96$, with a mean value closer to zero indicating good agreement. The same research group would later conduct similar experiments using a slightly larger dataset of 193 PRU images (test set $\approx$ 58), and reported a reduced DSC of 0.92 [67].

Godeiro et al. [68] tested four segmentation architectures (U-Net, Segnet, FCN8, and FCN32) using chronic wound images and proposed the use of a colour space reduction (CSR) on the CIELab space that increased DSC, accuracy, specificity, and sensitivity for all four networks. They used a dataset of 30 wound images comprising necrosis, granulation and slough tissue types. A total of 10 images were used for training, 5 for validation, and 15 for testing. Due to the small dataset size, pretrained ImageNet VGG16 weights were used to initialise the models. A watershed algorithm was used to assist in dividing images into wound, skin, and other regions. Following CIELab colour space processing, the four segmentation networks were used to obtain the different tissue types

in each wound image. They reported that the U-Net model with CSR provided the highest DSC of 0.9425. This is in comparison with the U-Net without CSR, which resulted in a DSC of 0.9153.

Zahia et al. [69] trained a model capable of performing optimised segmentation of the different tissue types present in pressure injuries (granulation, slough, and necrotic). The model was developed using MATLAB Neural Network Toolbox. They used a preprocessing step to remove flash light artefacts, followed by the creation of a set of 5×5 pixel sub-images used for training. They used a dataset of 22 images comprising stage 3 and 4 pressure injuries. A total of 17 images were used for training and 5 images were used for testing, acquired from the Igurko Hospital, Spain. An additional 4 images were purchased from The National Pressure Ulcer Advisory Panel (NPUAP) online store for validation purposes. All train and test images were 1020×1020 pixels. Examples of infected and necrotic tissues are present, together with healing states evidencing granulation tissue. In an attempt to negate the limited number of train and test images, all images were automatically cropped to $5 \times 5 \times 3$ pixels resulting in 270,762 RGB patches for granulation tissue, 37,146 patches for necrotic tissue, and 80,636 patches for slough. This approach ensured that there was no loss of wound texture features when training the network. They report an average DSC value of 0.9138, an average precision per class of 0.9731 for granulation tissue, 0.9659 for necrotic tissue, and 0.7790 for slough tissue. One notable limitation they observed is that deep areas within certain pressure injuries would appear dark, which the model would confuse with necrotic tissue, which is also generally very dark in colour.

Pathompatai et al. [70] proposed a region-focus training strategy for wound images in the Medetec dataset. They split full-size images into smaller patches for training, and found that network performance could be improved by increasing the number of challenging image patches into the training process. Challenging examples were determined by heatmap analysis. They split full images (typically 600×400 pixels) into 256×256 pixel patches with a 128 pixel stride (50% overlap). For the generation of additional challenging patches, they cropped from different offsets so that the new patches contained different contextual features. To test their method, they trained U-Net using 115 images for training, 29 images for validation, and 36 images for testing—all counts are for full-size images prior to splitting into patches. Expert labelling was not mentioned in the study; therefore, the ground truth masks should be considered to be a weakly supervised component. They compared models with and without region-focus training, and with different numbers of region-focus patches. Their best performing model reported an mIoU of 0.7816. They noted that although the introduction of difficult patch examples would generally improve network performance, the value would need to be tuned for different scenarios. This method could be interpreted as a form of manual attention, prior to the more widespread use of more advanced automated attention mechanisms present in more recent network architectures.

Li et al. [71] proposed a segmentation framework based on human-designed feature maps and artificially assigned convolution kernels using a modified MobileNet. First, a location encoder is used to convert the 2D coordinates of the input image into a location map, which is concatenated with the input image. Next, after downsampling, the location map is fused with the output of the network backbone, which is post-processed using smooth kernels to remove small holes and small non-wound regions in the prediction. Finally, the feature maps are upsampled to the size of the input image, resulting in output maps. This work observed that in order to maintain invariance, CNNs obfuscate the location information of input maps. This is contrary to the usual spatial distribution of wounds and backgrounds in wound images, which tend to be non-uniform in nature. In their experiments, they used a dataset of 950 wound images, as used in their previous work [63]. The training set comprised 760 wound images, and the test set comprised 190 images. They reported an mIoU value of 0.8647, a maxIOU value of 0.8675, and a precision value of 0.9503.

Cui et al. [72] conducted experiments which compared DFU segmentation results from U-Net and a patch-based CNN method, originally proposed by [73]. They used a wound dataset of 445 images acquired from New York University. For training, 392 wound images were used, and 53 images were used for testing. The GrabCut tool was used to remove background features. For the patch-based CNN method, the original images where split into patches, resulting in 4500 pairs of local and global patches (31×31 and 201×201, respectively). An adaptive thresholding method was used as a post-processing step to remove artefacts. They found that U-Net produced sharper wound boundaries when compared to those created by the patch-based CNN method. The reported U-Net results were 0.845 for DSC and 0.761 for mIoU.

Ohura et al. [74] trained four segmentation models (U-Net, U-Net with VGG16 backbone pretrained on ImageNet, SegNet, and LinkNet) using PRU images ($n = 396$) and tested on images of DFU ($n = 20$) and VLU ($n = 20$) which were collected from the Kyorin University Hospital, Japan. The number of training images used was 356, and the number of test images used was 40. The U-Net VGG16 model was found to provide the best results for sensitivity (0.993) and specificity (0.993), while the U-Net with VGG16 model provided the best AUC (0.998), DSC (0.947), and accuracy (0.989). These experiments showed that pretraining on ImageNet improved the standard U-Net with increases in AUC, DSC, and accuracy. However, sensitivity and specificity were

slightly lower for the U-Net VGG16 model when compared to the base U-Net model.

Wagh et al. [75] compared results of various deep learning architectures against traditional associated hierarchical random field (AHRF) segmentation, which reformulates the image segmentation task as a graph optimisation problem. The CNN architectures they experimented with were U-Net, FCN, and DeepLabV3. They devised two experiments, each using a different dataset. Dataset 1 comprised a total of 114 wound images (95 train, 19 val) and was acquired under controlled lighting conditions. Dataset 2 comprised a total of 316 wound images (263 train, 53 val). A total of 202 of these images were acquired via internet scraping, and 114 images were taken from dataset 1. Dataset 3 comprised a total of 1442 wound images (1201 train, 241 val) and were sourced from the University of Massachusetts Medical Center. All datasets contained examples of DFU, arterial, venous, PRU, and surgical wounds. The deep extreme cut algorithm [76] was used to provide ground truth labels; therefore, these experiments should be considered as weakly supervised. For dataset 1, the best performing method was FCN, with a DSC of 0.7822. The best performing method with dataset 2 was also FCN, with a DSC of 0.8418. The best performing method with dataset 3 was DeepLabV3, with a DSC of 0.8554. They also conducted an additional experiment using a common validation set across all datasets. For validation DSC results, the best performing models were FCN on dataset 1 (0.7822), and DeepLabV3 on dataset 2 (0.8537) and dataset 3 (0.8760).

Ong et al. [77] proposed a wound segmentation model with 18.5% fewer parameters than U-Net, with the aim of performing inference on mobile devices. They added an additional layer to both upsampling and downsampling pathways. They also added depthwise separable convolutions (depthwise convolution followed by a 1×1 convolution) to replace the standard convolutions, allowing for independent convolution of each input channel. The depthwise separable convolutions used a stride of 2×2 which effectively downsamples layers by a factor of 2 which is used instead of the max-pooling present in the original U-Net architecture. The upsampling pathway contains transposed convolutions with strides of 2×2. A dropout layer was also added after every depthwise separable convolution with a dropout rate of 0.2. Two parameters ("alpha" and "alpha_up") were added to adjust the number of filters for the upsampling and downsampling pathways, respectively. For training and testing, they used a private dataset sourced from local hospitals comprising 583 wound images. A total of 467 images were used for training, and 116 images were used for testing. A nurse provided the ground truth labels, with a second confirming label quality. After experimenting with different alpha values, they reported an mIoU of 0.869, compared to 0.813 for the unmodified U-Net.

Wang et al. [78] trained a MobileNetV2 wound segmentation model (pretrained on the Pascal VOC segmentation dataset) using a new dataset of 1109 DFU images (831 used for training, 278 used for testing). They used a localisation preprocessing step to remove non-DFU wound features from images prior to segmentation. Following hole-filling and small region removal morphological post-processing algorithms, they reported a mean DSC of 0.9047. However, there are limitations to this work in that the wound images are very small patches that were then heavily padded up to 224×224 pixels. This means that the actual wound pixels constituted very small regions of the total image. Without padding, the average width and height of the training images in this dataset are 71 and 104 pixels, respectively, while the average width and height of the test set images are 70 and 101 pixels, respectively. At such small sizes, as small as 17×18 pixels, many DFU wound features would be lost. Heavily limiting features in such a way, and using only a small number of images for testing ($n = 278$), it would be easier to obtain high test metrics as the network has fewer features to learn.

Wang et al. [79] would later conduct the Foot Ulcer Segmentation Challenge (FUSC) 2021 which introduced a training set of 810 images, a validation set of 200 images, and a test set of 200 images. The newly introduced images contained examples with significantly less padding and with more of the foot and background visible compared to the previous dataset they released. The winner of this challenge, Mahbod et al. [80], achieved an image-based DSC of 0.8880, a reduction of 1.67% from the original score reported by [78]. This may be evidence that the task became more challenging when larger DFU wound images were introduced. This shows that the model would need to learn more complex features that were not present in the original experiments conducted by Wang et al. [79] on the prior smaller dataset which had significantly smaller wound images with significantly fewer features.

Zahia et al. [81] proposed an end-to-end system using 2D wound images and 3D meshes acquired using a structure sensor with the aim of providing automated PRU measurement. For their segmentation experiments, they used a dataset of 175 images for training, and 35 images for testing. They experimented with Mask R-CNN using ResNet-50 and ResNet-101 backbones. They found ResNet-50 (pretrained on MS COCO) to provide the highest performance. For segmentation, they reported a DSC of 0.83, mean sensitivity of 0.85, and a mean precision of 0.87. They attributed the challenging nature of the segmentation task to ambiguous boundaries in some of the PRU images. Such ambiguities may therefore also be reflected in the ground truth labelling, given that such boundaries are likely to be subjective when delineated by expert human annotators.

Niri et al. [82] conducted experiments using superpixels as a preprocessing stage in a wound segmentation work-

flow. They used an FCN32 segmentation model as part of a pipeline to classify wound tissue types. The model was trained using more than 5000 wound superpixel images with no augmentation. They used 5256 images for training, and 1530 images for testing which were divided into granulation, slough, necrosis, and unknown tissue types. The private dataset was sourced from the Hospital Nacional Dos de Mayo (Lima, Peru) and the CHRO Hospital (Orleans, France) and comprises DFU wound images. They also used 219 images from the ESCALE database in their training set, including leg ulcers, diabetic ulcers, and bed sores. The preprocessing stage of using superpixel images means that the model is not exposed to background details. They reported an accuracy of 0.9268, precision of 0.7807, and a DSC of 0.7574.

Chairat et al. [83] trained a U-Net with an EfficientNet-B2 encoder using the WoundsDB dataset of 188 cases acquired from 47 patients. The dataset was split into 132 images for the training set, 28 images for the validation set, and 28 images for the test set. Due to the high resolution of the original dataset images (4896×3264 pixels), random crops were performed, resulting in a training set of 1056 images and a validation set of 224 images. They reported an mIoU of 0.8674, and observed that detection of smaller wounds was generally less accurate when compared to detection of larger wounds. It was also observed that despite the high resolution of the WoundsDB images, most of the images had been acquired at distance, meaning that the images comprised mostly of skin and environment details.

Watanabe et al. [84] proposed RoleNet (role-oriented fully convolutional networks) to segment wound area using sparse estimation and classify the tissue type (granulation, necrotic, and slough). This framework was developed as part of a larger system that could be used for wound area estimation using RGB-D point cloud data taken with an iPhone X mobile phone. They employed depthwise separable and atrous convolutions in their model architecture to reduce the number of model parameters and to increase the receptive field of convolutions. Downsampling with varying strides was also implemented to prevent loss of features. Their training strategy involved training the segmentation and classification networks separately, and then the models were combined and jointly retrained. A dataset of 40 wound images was acquired from a medical book [85], and was split into train ($n = 31$), validation ($n = 3$), and test ($n = 6$) sets. They reported an mIoU of 0.790.

Niri et al. [86] trained a U-Net for segmentation of skin and wounds in 2D images and 3D wound meshes. Following initial segmentation of the 2D images, they devised a strategy to select the best view from the multi-view 3D meshes, which were then segmented and reprojected back to 2D to update the original 2D segmentation result. For their experiments, they created a dataset of 569 chronic wound images which included pathologies such as DFU, burns, and PRU. A total of 426 images were used for training, while 143 images were used for testing. They reported a DSC of 0.9304 and an mIoU of 0.8661. They observed that high angle and distance variations due to camera position and orientation would affect model performance.

Scebba et al. [87] observed the many challenges inherent in wound segmentation, such as the heterogeneity of wound types, tissue colours, shapes, body position, background composition / complexity, image capturing conditions, and variable specifications of capture devices. They noted that initiatives to standardise medical wound photography would likely result in additional workload burdens on clinicians, and that such standards would likely still not guarantee a consistent resolution to generalisation issues experienced in real-world settings. To address these issues, they proposed a wound segmentation method, whereby a MobileNet localisation model would first localise the wound prior to segmentation using U-Net to reduce the number of extraneous features present in the image. The localisation model predictions were automatically adjusted to be increased in size by 50% to allow for surrounding tissue to be present in the cropped inputs required by the segmentation model. They used five wound datasets in their experiments—SwissWOU—a private dataset of DFU ($n = 1096$) and systemic sclerosis digital ulcers ($n = 63$), DFUC 2020 [88], Medetec [66], SIH (second healing intention dataset), and FUSC (Foot Ulcer Segmentation Challenge) [79]. In some cases, without further justification, full datasets were not used in their experiments, e.g. only 60 images were used from the FUSC dataset, only 53 images were used from the Medetec dataset, and only 58 images were used from the SIH dataset. They experimented with a range of commonly used segmentation networks, with and without automated localisation, and with manually localised images. When tested only on the SwissWOU DFU images (10% of all patients), they found that U-Net was the best performing network, with an MCC of 0.85 and an IoU of 0.75. When tested with the SwissWOU systemic digital ulcers, Medetec, SIH, and FUSC images, they found U-Net to be the best performing network, with a mean MCC of 0.8725 and an mIoU of 0.7875.

Xing et al. [89] proposed an improved U-Net model for DFU wound segmentation. The first improvement is in the use of a coarse positioning module (CPM) which is used to automatically crop the target area of training images to the smallest outer rectangle using the ground truth mask. This method reduces background features prior to training, and was not used on test images. The second improvement is the use of an SVM in the output layer to improve the ability of the model to generalise. The SVM replaces the softmax layer in the U-Net, and was used due to its ability to take into account training error and model complexity. SVM can also achieve better classification results and deal with non-linearities on small datasets. For their experiments, they used

the FUSC dataset, from which 610 images were used training, 200 images were used for validation, and 200 images were used for testing. Their best performing model, using both the CPM and SVM, reported a DSC of 0.8902. Analysis of a selection of results indicated that the CPM, when used in isolation with U-Net, was able to improve detection of smaller regions and performed well in segmenting larger regions. When the SVM was tested in isolation with U-Net, its performance on smaller wounds was not as accurate as CPM, and was significantly worse when segmenting larger DFU wounds. The CPM and SVM methods combined with U-Net provided the best overall performance, but on visual inspection, the quality of smaller DFU wound regions was not as accurate as the CPM with U-Net approach. Overall, their proposed method outperformed U-Net, ResU-Net, LinkNet, and Attention U-Net.

Bose et al. [90] proposed D3MSU-Net for wound segmentation, based on the original U-Net architecture. They used dense dilated convolutions to vary the field of view for each network level. Deep multiscale supervision blocks were implemented to provide supervision to hidden layers. This involves calculating loss at the hidden layers in addition to normal loss calculation that occurs at the end of the model, followed by optimisation of the model on the aggregated loss value. They tested their model on a series of medical imaging datasets (x-ray, CT, and MRI), including the FUSC and Medetec wound datasets. For the FUSC and Medetec datasets, they reported DSC values of 0.9285 and 0.9637, respectively.

Chang et al. [91] trained five segmentation models (U-Net, DeeplabV3, PSPNet, FPN, and Mask R-CNN) with a ResNet-101 encoder using a dataset comprising 2893 PRU images. They found the DeepLabV3 model to have the best performance, with an F1-score of 0.9887, IoU of 0.9782, precision of 0.9888, recall of 0.9887, and accuracy of 0.9925. Although these results are promising, there are notable limitations to this work. The authors indicate that challenging examples were excluded from the dataset. These cases included images where wounds were dressed with ointment, covered with dressing, actively bleeding, obscured by hematoma or pus, or if the image was out of focus or acquired under poor lighting conditions. Additionally, their test set was small, comprising only 10% of the total dataset (289 images). They also excluded images where clinicians disagreed on labelling, and the exact composition of the dataset was not reported.

Curti et al. [92] created a private wound dataset entitled "Deepskin" and used semi-supervised techniques in gradual training and validation stages to train a U-Net model with an EfficientNet-B3 encoder pretrained using ImageNet weights. Semi-supervision was used to assist labelling of unlabelled images starting with 145 labelled images, with results being confirmed (although not quantified in the study) by specialists using a set of quality criteria. All images used to train the model were sourced from a single centre, which the authors note may hinder the model's ability to generalise. The dataset comprised 1564 images (1440 × 1080 pixels) from 474 patients, which were collected over a 2 year period. The training set comprised 1407 images, and the test set comprised 157 images. Wound locations included foot, leg, chest, arm, and head. The authors note that all images had background details removed and that wounds would always occupy the centre of the image. By removing challenging examples, they achieved 0.96 in DSC, precision, and recall. This work also noted the absence of a set of standardised criteria for wound and periwound area definition among specialists.

Kendrick et al. [93] proposed a modified FCN32 architecture with a VGG16 backbone. Their training strategy involved creating patches of the training images to help de-emphasise non-wound background features. They replaced ReLU with Leaky ReLU to assist feature learning and to reduce the occurrence of dead neurons. Excessive downsampling was avoided by removal of the final three max-pooling layers, helping to retain feature map size in the lowest part of the network. Figure 3 illustrates the overall network architecture. Using the DFUC 2022 dataset, they achieved a DSC of 0.7447 and an mIoU of 0.6467. To date, this work represents the current state-of-the-art result on the largest publicly available chronic wound segmentation dataset—DFUC 2022. A limitation of this work is the lack of data understanding concerning the DFUC 2022 dataset. Currently, the composition and class distribution has not yet been fully analysed, meaning that the dataset may be both imbalanced and biased in terms of train and test distribution for wound class, size, anatomical location, and other factors.

Liao et al. [94] proposed HarDNet-DFUS, a modified version of the HarDNet-MSEG architecture which enhances the backbone and replaces the decoder. Using the DFUC 2022 dataset, they achieved a DSC of 0.7287, placing them first in DFUC 2022. They enhanced the original HarDNet-MSEG by replacing each HarDBlk module in the encoder backbone with a new HarDBlkV2 module, and replacing RFB modules in the decoder with large window attention (Lawin transformer) which utilises an MLP decoder, an MLP-mixer, and spatial pyramid pooling (SPP) to capture multiscale features. To further increase accuracy, they adopted an ensemble strategy using fivefold cross-validation and test time augmentation (TTA). Additional augmented images were added to the test images when testing using the sub-models, with the average of their outputs used as the final prediction results.

Ramachandram et al. [95] developed a segmentation network capable of both wound segmentation (AutoTrace) and tissue type segmentation (AutoTissue) for use in a commercial mobile app. The AutoTrace segmentation model used a traditional auto-encoder design with depthwise separable

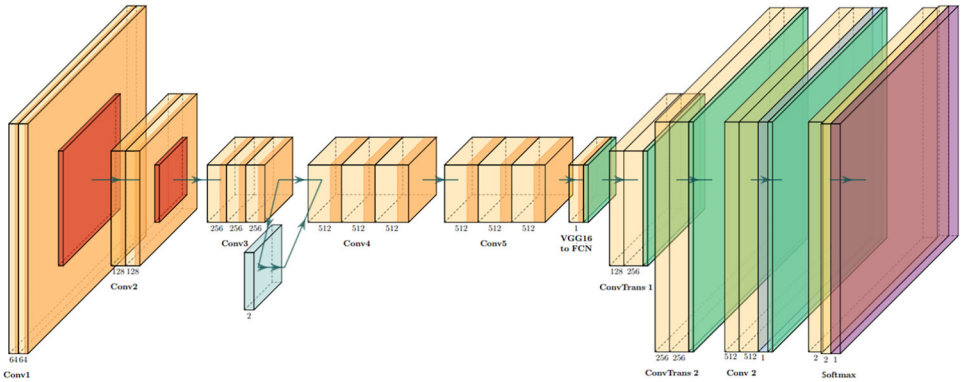


Fig. 3 Illustration of the DFU semantic segmentation network architecture proposed by Kendrick et al. [93]. Orange layers represent convolutions with Leaky ReLU activation, red layers indicate max-pooling, and light green layers indicate skip connections using modified squeeze and excite. In the decoding path, green represents dropout, yellow represents separable convolution with dilation and softmax layer. This model represents the current state-of-the-art performance on the largest publicly available chronic wound segmentation dataset (DFUC 2022)

convolutional layers, attention gates, and strided depthwise convolutions to downsample activations instead of using fixed max-pooling. Additive attention gates were placed at the end of each of the skip connections to regulate the flow of activations from prior layers. Attention coefficients can identify salient image regions and trim feature responses to preserve only the relevant activations. The decoder blocks consisted of bilinear upsampling followed by two depthwise separable convolution layers per block, reducing memory and computational requirements. The AutoTissue segmentation model used EfficientNetB0 as the encoder, with a four-block decoder, with each block comprising a single two-dimensional bilinear upsampling layer followed by two depthwise convolution layers. The AutoTrace model was trained using a private dataset of 467,000 images, and the AutoTissue model was trained using a second private dataset of 17,000 images. The images and clinical annotations were collected at hospitals across North America, providing them with a diverse range of inputs, including varied ethnic groups. However, the exact composition of the dataset was not disclosed in the study. They report an mIoU of 0.8644 for wound segmentation and 0.7192 for wound and tissue segmentation. A cohort of wound clinicians, by consensus, rated 91% (53/58) of the tissue segmentation results to be between fair and good for segmentation and tissue identification quality. Reporting of qualitative assessment of deep learning results is uncommon in wound related deep learning studies. The reporting of such qualitative measures in this study is useful; however, it is still very limited, with only 58 examples assessed. The reporting of the qualitative measures is also somewhat vague, i.e. a high percentage of ratings were found to be between fair and good. An additional limitation of this work was the small size of the test sets they used, with only 2000 image-label pairs used for testing the AutoTrace model, and 383 images for the AutoTissue model. This means that for the segmentation AutoTrace model, from a dataset total of 469,000 images (467,000 train + 2000 test), less than 0.44% of the total images were used for testing. Such small test sets may not be statistically significant and may not contain sufficiently varied examples when compared to the vast size of the corresponding training sets, meaning that evaluation metrics may not be reliable. This aspect may be especially pertinent in cases where the exact composition of the test set has not been quantified.

Marijanović et al. [96] developed three segmentation models for use with a robotic manipulator, RGB-D camera, and 3D scanner in the acquisition of wound images. They used the FUSC DFU dataset to train and test their models. A fixed-size overlapping sliding-window method was used to generate input images to the network, with each model using a different window size. Window sizes of 5, 7, and 9 were used for Model5, Model7, and Model9, respectively. The threshold for an input sub-image to be marked as wound was 50% for Model5 and Model7, and 25% for Model9. The core model architecture comprised of four fully connected hidden ReLU layers and one fully connected output layer with a sigmoid activation function. A total of 810 wound images were used for training (train = 648, val = 162) and 200 wound images were used for testing. Predictions from all three models were merged to form a final prediction mask using the AND logical operator. Finally, they performed post-processing using thresholding, noise removal, and region filling morphological operations. They reported a recall value of 0.77, a precision value of 0.72, and a DSC of 0.74.

Swerdlow et al. [97] trained a Mask R-CNN model with a ResNet-101 backbone for segmentation and classification of stage 1–4 PRU injuries. They used a dataset (eKare Inc. Data pressure injury wound data repository) of 969 pressure

injury images, with 848 images used for training, and 121 images used for testing. This work noted the lack of publicly available datasets, referencing pressure injury images available in the Medetec dataset. However, the study did not use these images in experiments due to degraded image quality. They reported a DSC of 0.92 for stage 1 injuries, 0.85 for stage 2 injuries, 0.93 for stage 3 injuries, and 0.91 for stage 4 injuries. The study protocol ensured that all images were taken with the same camera from approximately the same distance (40–65 cm) from the wound. The authors also note that wounds smaller than 2 × 2 cm were not included in the dataset.

Liu et al. [98] trained a U-Net segmentation model with a ResNet-101 backbone for use with automatic wound area measurement using a LiDAR camera. They used 528 images of PRU to train and test their model. They reported a mean DSC value of 0.8488, mIoU of 0.7773, mean precision of 0.8756, mean recall of 0.8639, and mean accuracy of 0.9807. However, this work notes that only the highest quality photographs were used, indicating that there may have been a lack of challenging examples. They initially collected 1038 PRU photographs, collected from the National Taiwan University Hospital. They then excluded images which were blurred, overexposed, underexposed, obscured, or contained other numerous non-wound objects or features. The total number of PRU photographs used in their experiments was 528, of which 327 were used for training and internal validation, and 201 used for external validation. All images were resized to 512 × 512 pixels prior to training. This study observed that in order for the model to become more robust, more challenging images would need to be used to train the model with, ideally exhibiting more complex non-wound background details. They also observed that PRU photographs could not always reveal drainage sinus and deep dead space especially when they were taken by nonprofessional first-line caregivers. These regions may show as black or very dark regions in the image.

Lan et al. [99] proposed FusionSegNet which performed segmentation as a means of improving binary classification of chronic wounds. They used the FUSC dataset for training and testing their segmentation model. Using pretraining on the AZH wound dataset they evaluated results of three segmentation networks using the FUSC validation set. All images were resized to 512 × 512 pixels for all experiments. They found that U-Net provided the highest metrics out of the three segmentation models they trained (Residual U-Net, MobileNetV2, and U-Net) and reported a precision value of 0.9009, a recall value of 0.9026, and a DSC value of 0.9010.

Oota et al. [100] proposed WSNET, a segmentation framework capable of (a) wound-domain adaptive pretraining on a large unlabelled wound dataset, and (b) a global–local architecture that utilises full image and its patches to learn fine-grained details from heterogeneous wounds. They used a new dataset of 2686 wounds (WOUNDSEG dataset), comprising examples of DFU, pressure trauma, venous, surgical, arterial, cellulitis, and others. Three classification backbones were first pretrained (DenseNet121, DenseNet169, and MobileNet) using images of DFU ($n = 19,773$), PRU ($n = 47,541$), surgical wounds ($n = 12,238$), trauma wounds ($n = 13,667$), and venous ulcers ($n = 32,492$). The classifiers were then frozen, and the decoder weights were fine-tuned over the wound segmentation dataset for four segmentation models. When tested on the WOUNDSEG dataset, they reported a DSC of 0.847. They also reported results of 0.956 DSC for the Medetec dataset (using U-Net with a DenseNet169 backbone), and 0.927 DSC for the AZH dataset (using LinkNet with a DenseNet121 backbone). This work highlights the importance of large-scale same-domain pretraining. Although these results are impressive, they are not reproducible as the vast majority of the dataset is private, and the pretrained weights have also not been shared publicly.

4 Instance segmentation

In this section, we discuss all the chronic wound segmentation papers that specifically reported on the use of instance segmentation methods. Instance segmentation focuses on identification of individual wounds per image, as opposed to semantic segmentation, which involves detection of wound pixels in an image. Our investigation found only four studies which specifically indicated that the experiments utilised instance segmentation. Instance segmentation is generally regarded as a more challenging task compared to semantic segmentation, as it requires localisation of wounds in addition to segmentation of wound regions. We note that there may be ambiguity as to what constitutes instance segmentation in the literature due to the models used in studies and the methods of evaluating those models. For example, Mask R-CNN is considered an instance segmentation network; however, studies that use this model architecture may not evaluate the results on an instance basis.

Wijesinghe et al. [105] developed a mobile app capable of severity stage classification of diabetic retinopathy and DFU wound segmentation. For the DFU segmentation task, they trained and tested a Mask R-CNN model using 400 DFU images sourced from a diabetic clinic in Sri Lanka. Using ten images for evaluation, they reported a mean average precision (mAP) of 0.87 at an IoU threshold of 0.5.

Gamage et al. [106] trained a Mask R-CNN network with a ResNet-101 backbone (pretrained with the MS COCO dataset) on neuropathic ulcers for instance segmentation. They used a dataset of 400 images, with 360 images used for training and 40 images used for testing. Their training strategy involved three stages: (1) during the first 30 epochs, only the head layers (the region proposal network, classifier

Table 1 Summary of deep learning experiments for chronic wound segmentation between 2015 and 2023

Publication	Year	Train	Val	Test	Split	DSC	mIoU	Dims	Model	IAM
Wang et al. [60]	2015	500	–	150	77/0/23	-	0.4730	480 × 640	ConvNet	HRC
Goyal et al. [62]	2017	420	60	225	60/9/32	0.7940	–	500 × 500	FCN-16 s	DC
Li et al. [63]	2018	760	–	190	80/0/20	–	0.8588	512 × 512	MobileNet*	C
Elmogy et al. [64]	2018	60	10	30	60/10/30	0.9300	–	1024 × 1024	3D DeepMedic	DC
Zapirain et al. [67]	2018	115	20	58	60/10/30	0.9200	–	1024 × 1024	3D DeepMedic	DC
Godeiro et al. [68]	2018	10	5	15	33/17/50	0.9425	–	640 × 480	UN	iP
Zahia et al. [69]	2018	17	4	5	65/15/20	0.9138	–	1020 × 1020	CNN*	NR
Pathompatai et al. [70]	2019	115	29	36	64/16/20	–	0.7816	600 × 400	UN	HRC
Li et al. [71]	2019	760	–	190	80/0/20	–	0.8647	–	MobileNet*	HRC
Cui et al. [72]	2019	392	–	53	88/0/12	0.8450	0.7610	240 × 320	UN	HRC
Ohura et al. [74]	2019	356	–	40	90/0/10	0.9470	–	256 × 256	UN \ VGG16	DC
Wagh et al. [75]	2020	95	–	19	83/0/17	0.7822	–	512 × 384	FCN	SP
Wagh et al. [75]	2020	263	–	53	83/0/17	0.8418	–	512 × 385	FCN	SP
Wagh et al. [75]	2020	1201	–	241	83/0/17	0.8554	–	512 × 386	DeepLabV3	SP
Ong et al. [77]	2020	467	–	116	80/0/20	–	0.8690	224 × 224	UN*	iP
Wang et al. [78]	2020	831	–	278	75/0/25	0.9047	–	224 × 224	MobileNetV2	DC, iP, C
Zahia et al. [81]	2020	175	–	35	83/0/17	0.8300	–	–	Mask R-CNN \ RN50	SP, iP, C
Niri et al. [82]	2021	5256	–	1530	77/0/23	0.7574	–	–	FCN32	SP, DC
Privalov et al. [101]	2021	295	–	35	89/0/11	0.7910	–	1024 × 1024	Mask R-CNN	HRC
Chairat et al. [83]	2021	132	28	28	70/15/15	–	0.8674	4896 × 3264	UN / ENB2	DC
Watanabe et al. [84]	2021	31	3	6	78/8/15	–	0.7900	256 × 256	RoleNet*	NR
Watanabe et al. [84]	2021	31	3	6	78/8/15	–	0.8120	256 × 256	UN	NR
Niri et al. [86]	2021	426	–	143	75/0/25	0.9304	0.8661	–	UN	SP
Monroy et al. [102]	2021	178	–	10	95/0/5	0.9100	–	320 × 320	UN	DC, NR
Mahbod et al. [80]	2022	810	200	200	67/17/17	0.8880	–	512 × 512	LN / ENB1 & UN / ENB2	C, DC, iP
Mahbod et al. [80]	2022	831	–	278	75/0/25	0.9207	–	224 × 224	LN / ENB1 & UN / ENB2	C, DC, iP

Table 1 continued

Publication	Year	Train	Val	Test	Split	DSC	mIoU	Dims	Model	IAM
Xing et al. [89]	2022	610	200	200	60/20/20	0.8902	–	512 × 512	UN	DC, iP
Bose et al. [90]	2022	152	–	8	95/0/5	0.9637	–	224 × 224	D3MSU-Net*	DC, iP
Bose et al. [90]	2022	831	–	278	75/0/25	0.9285	–	224 × 224	D3MSU-Net*	C
Chang et al. [91]	2022	2604	–	289	90/0/10	0.9887	0.9782	1000 × 750	DeepLabV3 \ RN101	NR, C
Curti et al. [92]	2022	1407	–	157	90/0/10	0.9600	–	–	UN \ ENB3	SP, DC, iP
Kendrick et al. [93]	2022	2000	–	2000	50/0/50	0.7447	0.6467	640 × 480	FCN32 \ VGG16*	DC
Liao et al. [94]	2022	2000	–	2000	50/0/50	0.7287	–	512 × 512	HarDNet-DFUS*	DC
Ramachandram et al. [95]	2022	467,000	–	2000	100/0/0	–	0.8644	–	UN*	SP
Foltynski et al. [103]	2022	475	90	41	78/15/7	0.9090	0.8390	512 × 512	UN	NR
Marijanović at al. [96]	2022	648	162	200	64/16/20	0.7400	–	224 × 224	CNN*	DC, iP
Swerdlow et al. [97]	2023	848	–	121	88/0/12	0.9025	–	1024 × 1024	Mask R-CNN	C
Liu et al. [98]	2023	327	–	201	62/0/38	0.8488	0.7773	512 × 512	UN / RN101	C
Lan et al. [99]	2023	803	200	200	67/17/17	0.9010	–	512 × 512	UN	iP
Oota et al. [100]	2023	125,711	–	8	100/0/0	0.9560	–	192 × 192	UN \ DN169	NR
Oota et al. [100]	2023	125,711	–	278	100/0/0	0.9270	–	192 × 192	LN \ DN121	C
Oota et al. [100]	2023	125,711	–	2686	98/0/2	0.8470	0.7130	192 × 192	LN \ DN121	DC, iP
Oliveira et al. [104]	2023	1,100	138	138	80/10/10	0.7540	–	384 × 384	VENet*	DC

Split is the percentage split of train, validation, and test sets, where zero denotes a fractional value < 1. DSC—Dice similarity coefficient, mIoU—mean intersection over union, Dims—dimensions, IAM—image acquisition method, UN—U-Net, RN—ResNet, EN—EfficientNet, LN—LinkNet, DN—DenseNet, *—modified or novel model architecture, C—camera, DC—digital camera, HRC—high-resolution camera, iP—iPhone, SP—smartphone, NR—not reported. Where possible, all dataset numbers are those reported prior to any preprocessing steps. Additionally, only publications that reported unambiguous train and test set numbers and DSC or mIoU values are summarised in this table

and mask heads) of the network are trained; (2) in the next 50 epochs, the upper 4+ layers are trained; and (3) in last 20 epochs all layers in the network are trained. They achieved 0.5084 mAP (for IoU = 0.5 to 0.95), 0.8632 average precision (AP) (for IoU = 0.5), and 0.6157 AP (for 0.75 IoU).

Privalov et al. [101] trained a Mask R-CNN network for DFU instance segmentation. They pretrained their model using the MS COCO dataset. They used a dataset of 295 wound images for training, and 35 images used for testing. They reported a DSC of 0.7910. They also performed inter- and intra-rater analyses using one-way analysis of variance (ANOVA) to analyse the variance between and within group means. This revealed no statistically significant differences for all raters for the network in the first round ($F = 1.424$ and $p > 0.228$) and the second round ($F = 0.9969$ and $p > 0.411$) for segmentation. The repeated measure analyses revealed no statistically significant differences in the quality of segmentation for the four medical experts ($F = 6.05$ and $p > 0.09$). However, they observed some intra-rater variability.

Evidence of the effect of reducing the size of the train and test sets on deep learning wound segmentation models was highlighted by Cao et al. [107]. They reported a series of baseline results using 4000 DFU images trained on a selection of segmentation networks. They then trained using their own instance segmentation model (a variant of Mask R-CNN) with a reduced dataset of 1426 images and observed a significant increase in mAP (0.6940 to 0.8570). However, it was not clear if the increases in performance metrics were due to the new model they created or a result of simply reducing the number of images used in their experiments. The study indicated that only DFU wound images with clear ulcer edges and non-blurry cases were used in the experiments with the proposed model, suggesting that challenging examples were excluded.

5 Meta-analysis

A total of 40 chronic wound deep learning segmentation papers were reviewed in this review, with a total of 43 experiments from the papers summarised in Table 1. Figure 4 summarises the number of publications in chronic wound deep learning research covered by this review and meta-analysis.

The total number of experiments that reported DSC values is 34, and the mean DSC value for all experiments is 0.8733. In terms of test sets, all experiments reporting DSC values can be divided into two distinct groups: (1) experiments that use between 5–289 test images; (2) experiments that use between 1530–2686 test images. A total of 38 experiments used test sets ranging from 5–289 images, and a total of five experiments used test sets ranging from 1530–2686 images. The mean DSC for experiments in the 5–289 test images range is 0.8872. The mean DSC for experiments in the 1530–2686 test images range is 0.7695. These mean DSC values indicate a strong correlation between test set size and reported DSC, with a difference of 0.1177 between those experiments that use $<$ 300 test images and those that use $>$ 1500 test images. Figure 5 shows the relationship between test set sizes and DSC values. The trend line demonstrates a negative correlation, i.e. smaller test sets resulted in higher DSC, and larger test sets resulted in lower DSC values.

The total number of experiments that reported mIoU values is 16, and the mean IoU value for all experiments was 0.7976. In terms of test sets, the experiments that reported mIoU values can be divided into two distinct groups: (1) experiments that use between 6–289 test images; (2) experiments that use between 2000–2686 test images. A total of 13 experiments used test sets ranging from 6–289 images, and a total of 3 experiments used test sets ranging from 2000–2686 images. The mIoU for experiments in the 6–289 test images range is 0.8106. The mIoU for experiments in the 2000–2686 test images range is 0.7414. As with the previous DSC analysis, these mIoU values indicate a negative correlation between test set size and reported mIoU, with a difference of 0.0692 between those experiments that use $<$ 300 test images and those that use $>=$ 2000 test images. Figure 6 shows the relationship between test set sizes and mIoU values. The trend line demonstrates that higher mIoU values correlate with smaller test sets, and that lower mIoU values correlate with larger test sets. Although this trend is less prominent when compared to the DSC results (see Fig. 5), there is still a clear correlation.

Of the 43 experiments summarised in Table 1, 18 experiments reported a DSC value of $>$ 0.9, all of which used test sets that comprised $<$ 300 images. The highest DSC value among all experiments that used $>=$ 1500 test set images is 0.8470. Only one experiment reported an mIoU value of $>$ 0.9, which used a test of $<$ 300 images, while there were no experiments which reported an mIoU $>$ 0.9 when $>$ 300 test set images were used.

The mean test set size for all experiments is 344. The mean test set size of the experiments that comprise $<$ 300 images is 120. The mean test set size of the experiments that comprise $>$ 1500 images is 2044. The total number of experiments that use $<$ 300 test images is 38. The total number of experiments that use $>$ 1500 test images is 5. The total number of experiments that reported DSC values is 34. The total number of experiments that reported mIoU values is 16. The total number of experiments that reported both DSC and mIoU is 7.

These findings show that, on average, experiments that use smaller test sets ($<$ 300 images) reported significantly higher DSC and mIoU values compared to experiments that use $>$ 1500 test set images. We observe that the R^2 values

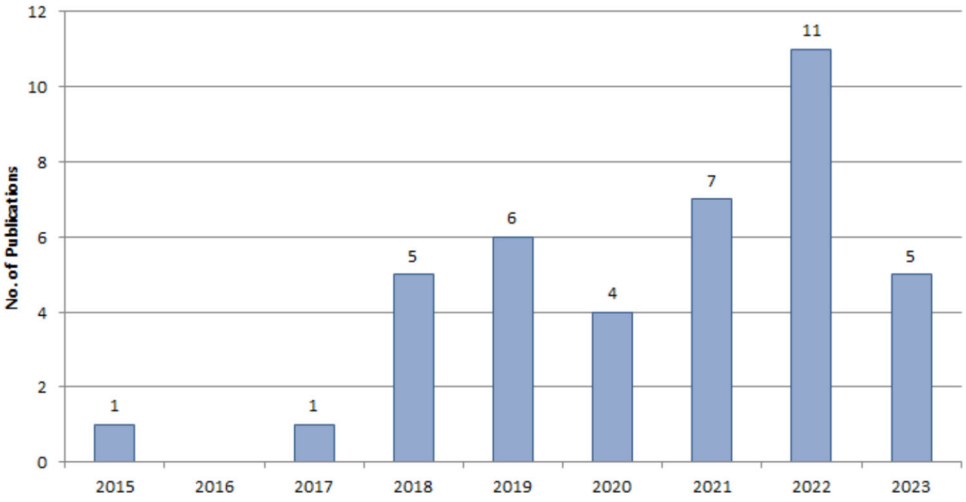


Fig. 4 Graph showing the number of deep learning chronic wound segmentation publications between 2015 and 2023. Note that the 2023 figure only includes papers published up to March 2023

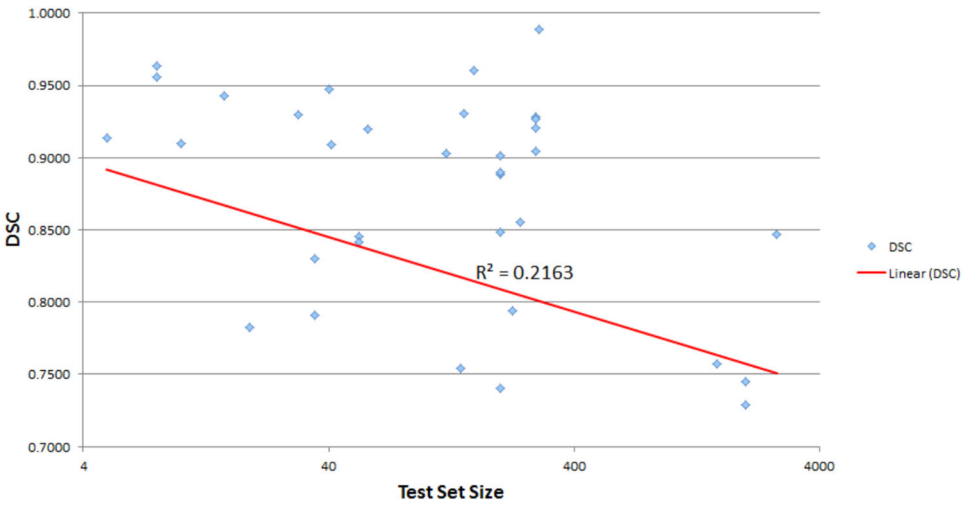


Fig. 5 Scatter chart showing the negative correlation between test set sizes used in the literature and corresponding DSC values

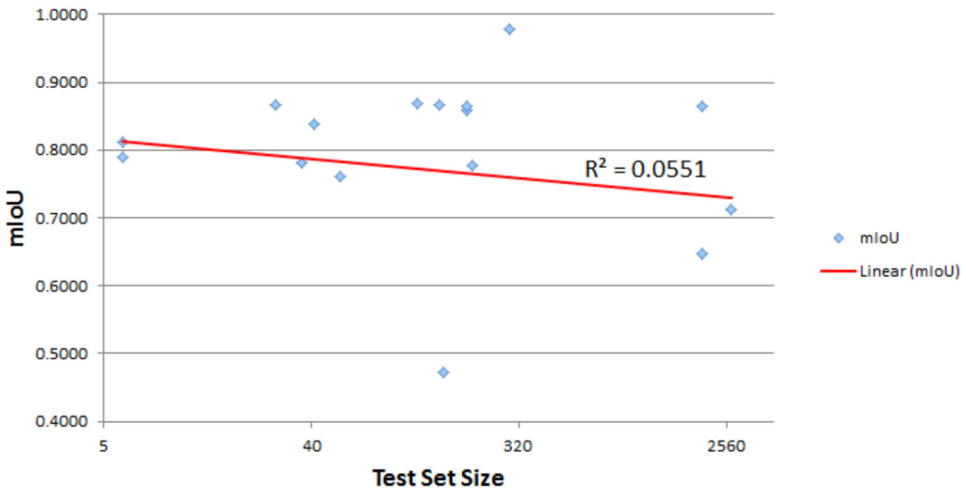


Fig. 6 Scatter chart showing the negative correlation between test set sizes used in the literature and corresponding mIoU values

for both graphs (Figs. 5 and 6) might be considered relatively low ($R^2 = 0.2163$ for DSC, $R^2 = 0.0551$ for mIoU) in some fields of study. However, R^2 values are domain dependent, and in the absence of similar studies in this field, we report these measures as baselines which can be compared against in future works. A possible limitation of this analysis is that the number of experiments that reported the use of test sets with < 300 images ($n = 38$) significantly outnumbers those experiments that reported the use of larger test sets ($n = 5$). Additionally, the total number of experiments present in the analysis may not yet be sufficient to form a comprehensive assessment. However, given that few experiments in the literature provide composition details of test sets (discussed further in Sect. 5.3), we suggest that the trends shown in Figs. 5 and 6 are at least partly indicative of issues prevalent in results reported in the literature.

For all experiments reported in Table 1, the smallest training set used is ten images, and the largest training set size is 467,000 images. The smallest validation set used is three images, and the largest validation set size is 200 images. The smallest test set used is five images, and the largest test set size is 2686 images. The mean total dataset size (train, validation, and test) for all experiments in Table 1 is 8712 images. Table 2 shows a summary of the total number of datasets used in the experiments reviewed in the literature.

The mean width and height of all images used in the experiments detailed in Table 1 is 585 and 506 pixels, respectively. The mean total number of pixels used in the experiments reported in Table 1 is 296,010 pixels per image.

A summary of the deep learning model architectures used in the experiments detailed in Table 1 is shown in Fig. 7. These figures clearly show that U-Net (20 experiments) is the most commonly used model architecture in chronic wound segmentation deep learning research, followed by FCN (5 experiments) and LinkNet (4 experiments).

5.1 Test sets

One of the main observations in the field of deep learning for wound segmentation is that most studies use very small test sets (< 300 images) and that test metrics, such as DSC and mIoU, often correspond to test set size (see Figs. 5 and 6). Studies that report results from smaller test sets, on average, show significantly higher test metrics, while studies that use larger test sets, on average, show significantly lower test metrics. The use of very small test sets in the majority of chronic wound deep learning studies means that those segmentation models may not be sufficiently challenged during testing. Such models are therefore unlikely to generalise well in real-world settings. Comparison of evaluation and test metrics across existing studies should therefore be regarded tentatively, especially in cases where small test sets have been used.

5.2 Qualitative assessment

Deep learning chronic wound segmentation methods have shown to provide high levels of accuracy in laboratory settings [108]. However, very few studies focus on expert qualitative assessment. This is understandable, given that obtaining access to clinical experts can be difficult. However, given that the aim of these studies is to work towards the development of systems that will be used in real-world settings, expert qualitative assessment should be regarded as a key factor in results analysis. Research groups who do not have access to clinical experts should consider collaboration with those groups who do in order to advance the field, and to promote the idea of human-in-the-loop in the development of chronic wound segmentation models.

A recent study reported very poor inter-rater agreement in determining the presence and regions of various tissue types, with results as low as 0.014 (Krippendorff alpha value) for epithelial tissue [48]. For the largest publicly available chronic wound dataset (DFUC 2022), we analysed the inter-reliability of the expert coders on $\approx 20\%$ of the data. For the delineation between experts, the mean DSC is 0.6982 ($sd = 0.2545$), with an mIoU of 0.5877 ($sd = 0.2670$), a recall of 0.9112 ($sd = 0.1907$), a precision of 0.6383 ($sd = 0.2846$), and an accuracy of 0.9869 ($sd = 0.0291$). The sd values from these results indicate significant variation in terms of mean DSC, mIoU, recall, and precision. Variability in expert labelling used for ground truth masks may mean that using only ground truth masks during testing could be insufficient when assessing the true accuracy of wound segmentation model prediction results. An example from the DFUC 2022 dataset showing expert rater delineation variability is shown in Fig. 8.

5.3 Data characterisation

Another notable observed limitation of current deep learning research in chronic wound segmentation is the lack of information regarding how well balanced datasets are when used in experiments. Data characterisation can be a vital component in deep learning research experiments [109]. Details of the exact composition of chronic wound datasets when used in deep learning segmentation experiments is often poorly understood and rarely reported. Examples of balancing may include any combination of the following:

1. Balance by wound class / tissue composition.
2. Balance by wound size / severity.
3. Balance by patient clinical data (e.g. ethnicity, age, etc.).
4. Balance by image quality (e.g. lighting or image sharpness).
5. Balance by feature similarity.

Table 2 Summary of datasets used in the experiments in the reviewed literature

Dataset	Experiments
Private	24
Medetec	11
DFUC 2022	10
FUSC	8
Internet scraped images	4
NYU Wound Database	3
AZH	3
WoundsDB	2
"multidisciplinary approach to chronic wounds" dataset	2

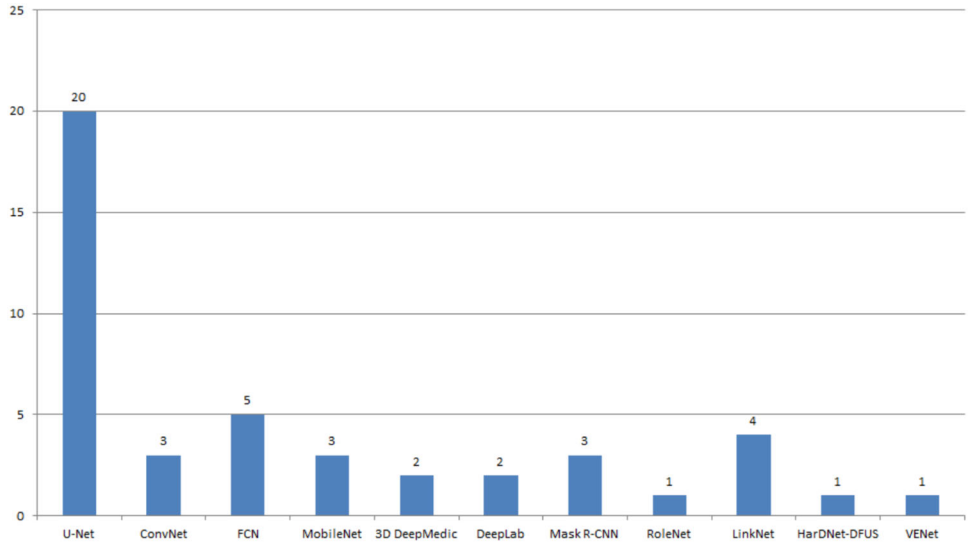


Fig. 7 Graph summarising the different model architectures used in the experiments in the reviewed literature

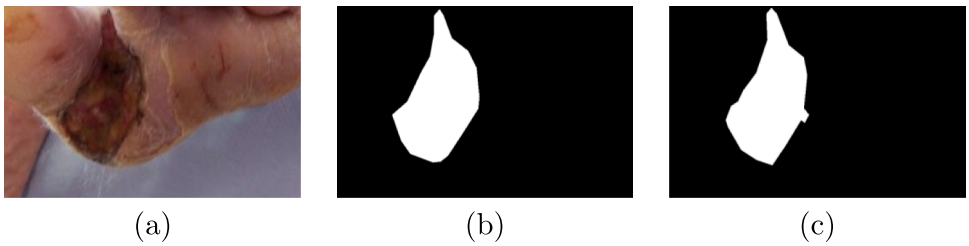


Fig. 8 Illustration showing delineation variability between two clinical experts—**a** shows the original wound image; **b** shows the delineation from expert A; **c** shows the delineation from expert B. Example taken from the DFUC 2022 dataset. Note that images are cropped for illustrative purposes

Studies that do not attempt to address at least a few of these possible imbalances may produce unreliable results. For example, a test set that includes only higher quality images means that the results may be biased as only easy to segment examples are used for inference. Another pertinent example is when test sets contain images that have high visual similarity with other images in either the train or test set [110]. Removal of duplicate images present within and across train and test sets is also an important factor [111]. Imbalances such as these are unlikely to allow for sufficient testing of the ability of a trained model to generalise, and resulting test metrics are likely to show over-estimates of a model's true ability. In studies where small test sets are used, these factors may be magnified considerably. These findings correlate with recent studies in other deep learning domains which note a sparsity of transparency and dataset characterisation [112, 113]. Thorough data understanding and transparency should be seen as a critical component of all deep learning research in order to obtain reliable results, which in turn will help to establish clinical confidence in model predictions.

5.4 Experiment design

Studies that use pretrained models with large general datasets such as Pascal VOC or ImageNet, often do not report on the effect of the pretraining, i.e. comparing models trained with and without pretraining. This is common among most wound segmentation papers, whereby the effect of an experiment component is usually not tested in isolation. Another example of this phenomenon is in the use of augmentation methods during training phases [114]. Reporting of results from studies which conduct experiments on individual augmentation methods in wound segmentation for different datasets would be highly useful for future research. Similar data have been published in other medical imaging domains [115], which should act as a motivator to make progress in the field of deep learning in chronic wound research.

5.5 Challenging cases

Several studies investigated in this review ($n = 6$: [87, 91, 97, 98, 103, 107]) noted that challenging examples had been removed prior to training and testing models. Such cases may include photographs with poor or inconsistent lighting, blurry features, small wounds, or wounds that are present on curvatures of the anatomy. In those studies that reported the removal of challenging examples, experiments had essentially been setup with ideal conditions in order to obtain the best results. We posit that reliable real-world wound segmentation systems should be designed to be highly robust and be able to handle such challenging cases, and that this aspect may be one of the key facets in moving the field forward, especially in terms of home monitoring of wounds where capture conditions are less likely to be controlled.

5.6 Reproducibility

Our findings show that not all published research works in wound segmentation are easily reproducible. Reproducibility and transparency in terms of source code sharing are vital in this domain to allow for other researchers to fully benefit from the findings of others, and to help build on existing methods. Model architectures may be reproducible, without publicly available source code, by use of model details recorded in publications. However, this highly depends on the quality of the details within each paper, and incurs a significant time commitment as model architectures have to be reconstructed from scratch. Of the 40 papers we reviewed, ten papers indicated that they had modified model architectures, of which two papers [90, 94] provided online repository links to their source code.

5.7 Chronic wound dataset availability

Sharing of chronic wound datasets has improved in recent years [116–118]; however, compared to other medical domains, the total number of publicly available datasets is still very small. One of the main reasons for this is the difficulty in accessing clinical experts who are able to provide sufficient ground truth labelling [119]. Additionally, General Data Protection Regulation (GDPR) compliance may also present barriers to data sharing. Scenarios that may prevent GDPR compliance include the presence of patient Personally Identifiable Information (PII) within dataset images and associated clinical data. Model explainability also falls under the remit of GDPR which poses additional ethical issues [120]. Debiasing of deep learning datasets may also fall under special categories of data when processing patient images for use in training deep learning segmentation models [121]. The lack of publicly available ethnically diverse datasets also poses a problem for chronic wounds segmentation research, especially those comprising darker skin tones [122]. Of those studies that claim to have used ethnically diverse wound datasets, the exact composition of those datasets was not reported, so it is unknown how well those models were able to generalise across different ethnic groups.

Due to the significant range of features found across different chronic wounds at different stages of development, availability of highly heterogeneous wound datasets should be seen as critical to the progress on scientific investigation in the field. Of the 43 experiments summarised in Table 1, 24 papers used private datasets. This includes datasets that were claimed to be publicly available but were either not found online or had been unsuccessfully requested. These findings indicate that the vast majority of deep learning studies in chronic wounds have not shared their datasets. The sparsity of public datasets and the abundance of private datasets represents a serious challenge for research in this domain and may be a significant limiting factor in research progress. A summary of publicly available chronic wound datasets is shown in Table 3.

6 Recommendations

This review highlights some of the major challenges present in current deep learning research in chronic wound segmentation. Based on our findings in the review and meta-analysis, we propose a number of recommendations for researchers working in this domain:

1. Future research works should focus on significantly increasing test set sizes to provide a more meaningful assessment of chronic wound segmentation that better

Table 3 Summary of publicly available chronic wound datasets

Publication	Year	Dataset	Resolution	Train	Test	Total
Thomas et al. [66]	2014	Medetec	Various	–	–	608
Yang et al. [123]	2016	YWHD	5184 × 3456	–	–	201
Alzubaidi et al. [124]	2020	Alzubaidi	Various	–	–	1036
Wang et al. [78]	2020	AZH*	224 × 224	831	278	1109
Kręcichwost et al. [125]	2021	WoundsDB*	4896 × 3264	–	–	188
Wang et al. [79]	2021	Medetec*	Various	152	8	160
Wang et al. [79]	2021	FUSC*	512 × 512	1010	200	1210
Kendrick et al. [93]	2022	DFUC 2022*	640 × 480	2000	2000	4000

YWHD—Yang Wound Healing Dataset; AZH—Advancing the Zenith of Healthcare; FUSC—Foot ulcer segmentation challenge; DFUC—diabetic foot ulcer challenge; *—dataset includes ground truth masks

reflects the large variety of wound features found in real-world settings. Taking into account the current publicly available chronic wound datasets, we recommend a train and test split of 50:50.

2. Researchers should share their source code with the research community so that others may more easily build on research progress. Exact details of environment setup should also be included as training environments can be highly dependent on specific library versions.
3. Researchers should attempt to integrate multiple publicly available chronic wound datasets into their experiments. This will help to gauge a better understanding of the ability of a given model, especially in cases where datasets from multiple sources are used as test sets.
4. Researchers should seek to collaborate in cases where access to clinical expertise is limited. This will help to increase the currently limited reporting of qualitative measures.
5. Rather than exclude challenging cases from experiments, researchers should include such examples and devise methods that are able to better accommodate these types of images in order to make models more robust.
6. Researchers should seek to better understand the nature of the data they are working with. This will help to improve the scope of understanding in the field and may reveal aspects that have not previously been appreciated or considered.

From the above suggestions, we note that simply increasing test set size may still not be sufficient when determining a model's true ability to generalise if the composition of the test set is not fully understood. However, in the absence of a means of fully understanding the composition of the test set, increasing the test set size may help to negate some of the issues associated with the use of small test sets especially if test sets from multiple sources are used.

In the long term, the most accurate overall assessment of the ability of a model to generalise would come from a combined analysis of dataset understanding, test metrics, and qualitative feedback obtained from clinical experts in wound care.

7 Conclusions

In this review and meta-analysis of chronic wound segmentation methods in deep learning, for the first time, we identify some of the major issues that present significant obstacles in this research domain. The most notable of these issues is the use of very small test sets in the vast majority of studies combined with a lack of data understanding. Models evaluated on such test sets are unlikely to perform well on out of distribution examples meaning that they will likely not generalise well on the wide range of chronic wound features found in real-world settings. This work represents the most substantial review of deep learning in chronic wound segmentation to date, and provides researchers with key insights into possible areas of research progress.

Author Contributions BC conceived the study and wrote the main manuscript. BC and MHY performed data analysis and interpretation. CK, NDR, JMP, and MHY provided project supervision and advice on analysis. NDR, JMP, and MHY provided project administration. All authors participated in contributing to text and the content of the manuscript, including revisions and edits. All authors approve of the content of the manuscript and agree to be held accountable for the work.

Data Availability No datasets were generated during the current study.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

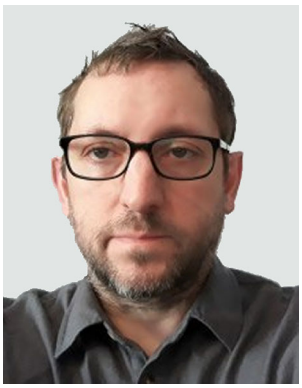
1. Eriksson, E., Liu, P., Schultz, G., Martins-Green, M., Tanaka, R., Weir, D., Gould, L., Armstrong, D., Gibbons, G., Wolcott, R., Olutoye, O., Kirsner, R., Gurtner, G.: Chronic wounds: treatment consensus. Wound Repair Regen. **30**(2), 156–171 (2022). https://doi.org/10.1111/wrr.12994
2. Moura, J., Rodrigues, J., Gonçalves, M., Amaral, C., Lima, M., Carvalho, E.: Imbalance in t-cell differentiation as a biomarker of chronic diabetic foot ulceration. Cell. Mol. Immunol. **16**(10), 833–834 (2019)
3. Petersen, B., Linde-Zwirble, W., Tan, T.-W., Rothenberg, G., Salgado, S., Bloom, J., Armstrong, D.: Higher rates of all-cause mortality and resource utilization during episodes-of-care for diabetic foot ulceration. Diabetes Res. Clin. Pract. (2022). https://doi.org/10.1016/j.diabres.2021.109182
4. Jenkins, D.A., Mohamed, S., Taylor, J.K., Peek, N., Veer, S.N.: Potential prognostic factors for delayed healing of common, non-traumatic skin ulcers: A scoping review. Int. Wound J. **16**(3), 800–812 (2019). https://doi.org/10.1111/iwj.13100
5. Zhang, P., Lu, J., Jing, Y., Tang, S., Zhu, D., Bi, Y.: Global epidemiology of diabetic foot ulceration: a systematic review and meta-analysis. Ann. Med. **49**(2), 106–116 (2017). https://doi.org/10.1080/07853890.2016.1231932
6. Willems, R.: Health economic considerations to effectively implement telemonitoring of diabetic foot ulcer. Lancet Region. Health Europe **32**, 100688 (2023). https://doi.org/10.1016/j.lanepe.2023.100688
7. Probst, S., Saini, C., Gschwind, G., Stefanelli, A., Bobbink, P., Pugliese, M.-T., Cekic, S., Pastor, D., Gethin, G.: Prevalence and incidence of venous leg ulcers—a systematic review and meta-analysis. Int. Wound J. **20**(9), 3906–3921 (2023). https://doi.org/10.1111/iwj.14272
8. Sardo, P.M.G., Teixeira, J.P.F., Machado, A.M.S.F., Oliveira, B.F., Alves, I.M.: A systematic review of prevalence and incidence of pressure ulcers/injuries in hospital emergency services. J. Tissue Viability **32**(2), 179–187 (2023). https://doi.org/10.1016/j.jtv.2023.02.001
9. Martinengo, L., Olsson, M., Bajpai, R., Soljak, M., Upton, Z., Schmidtchen, A., Car, J., Järbrink, K.: Prevalence of chronic wounds in the general population: systematic review and meta-analysis of observational studies. Ann. Epidemiol. **29**, 8–15 (2019). https://doi.org/10.1016/j.annepidem.2018.10.005
10. Martinengo, L., Yeo, N.J.Y., Markandran, K.D., Olsson, M., Kyaw, B.M., Car, L.T.: Digital health professions education on chronic wound management: A systematic review. Int. J. Nurs. Stud. **104**, 103512 (2020). https://doi.org/10.1016/j.ijnurstu.2019.103512
11. Lo, Z.J., Lim, X., Eng, D., Car, J., Hong, Q., Yong, E., Zhang, L., Chandrasekar, S., Tan, G.W.L., Chan, Y.M., Sim, S.C., Oei, C.W., Zhang, X., Dharmawan, A., Ng, Y.Z., Harding, K., Upton, Z., Yap, C.W., Heng, B.H.: Clinical and economic burden of wound care in the tropics: a 5-year institutional population health review. Int. Wound J. **17**(3), 790–803 (2020). https://doi.org/10.1111/iwj.13333
12. Popescu, V., Cauni, V., Petrutescu, M., Rustin, M., Bocai, R., Turculet, C., Doran, H., Patrascu, T., Lazar, A., Cretoiu, D., Varlas, V., Mastalier, B.: Chronic wound management: from gauze to homologous cellular matrix. Biomedicines **11**, 2457 (2023). https://doi.org/10.3390/biomedicines11092457
13. Sen, C.K.: Human wound and its burden: updated 2020 compendium of estimates. Adv. Wound Care **10**(5), 281–292 (2021). https://doi.org/10.1089/wound.2021.0026
14. Petrone, F., Giribono, A., Massini, L., Pietrangelo, L., Magnifico, I., Bracale, U., Di Marco, R., Bracale, R., Petronio Petronio, G.: Retrospective observational study on microbial contamination of ulcerative foot lesions in diabetic patients. Microbiol. Res. **12**, 1 (2021). https://doi.org/10.3390/microbiolres12040058
15. Rathur, H., Boulton, A.: The neuropathic diabetic foot. Nat. Clin. Practice Endocrinol. Metab. **3**, 14–25 (2007)
16. Bader, M.S.: Diabetic foot infection. Am. Fam. Phys. **78**(1), 71–79 (2008)
17. Franks, P., Barker, J., Collier, M., Gethin, G., Haesler, E., Jawien, A., Läuchli, S., Mosti, G., Probst, S., Weller, C.: Management of patients with venous leg ulcers: challenges and current best practice. J. Wound Care **25**, 1–67 (2016). https://doi.org/10.12968/jowc.2016.25.Sup6.S1
18. Mader, J., Haas, W., Aberer, F., Boulgaropoulos, B., Baumann, P.M., Pandis, M., Horvath, K., Aziz, F., Köhler, G., Pieber, T., Plank, J., Sourij, H.: Patients with healed diabetic foot ulcer represent a cohort at highest risk for future fatal events. Sci. Rep. (2019). https://doi.org/10.1038/s41598-019-46961-8
19. Xiong, X.-F., Wei, L., Xiao, Y., Han, Y.-C., Yang, J., Zhao, H., Yang, M., Sun, L.: Family history of diabetes is associated with diabetic foot complications in type 2 diabetes. Sci. Rep. **10**, 17056 (2020). https://doi.org/10.1038/s41598-020-74071-3
20. Costa, R., Cardoso, N., Procópio, R., Navarro, T., Dardik, A., Cisneros, L.: Diabetic foot ulcer carries high amputation and mortality rates, particularly in the presence of advanced age, peripheral artery disease and anemia. Diabetes Metab. Syndrome Clin. Res. Rev. **11**, 1 (2017). https://doi.org/10.1016/j.dsx.2017.04.008
21. Vainieri, E., Ahluwalia, R., Slim, H., Walton, D., Manu, C., Taori, S., Wilkins, J., Huang, D., Edmonds, M., Rashid, H., Kavarthapu, V., Vas, P.: Outcomes after emergency admission with a diabetic foot attack indicate a high rate of healing and limb salvage but increased mortality: 18-month follow-up study. Exp. Clin. Endocrinol. Diabetes (2020). https://doi.org/10.1055/a-1322-4811
22. Renner, R., Erfurt-Berge, C.: Depression and quality of life in patients with chronic wounds: ways to measure their influence and their effect on daily life. Chronic Wound Care Manag. Res. **4**, 143–151 (2017). https://doi.org/10.2147/CWCMR.S124917
23. Polikandrioti, M., Vasilopoulos, G., Koutelekos, I., Panoutsopoulos, G., Gerogianni, G., Alikari, V., Dousis, E., Zartaloudi, A.: Depression in diabetic foot ulcer: associated factors and the impact of perceived social support and anxiety on depression. Int. Wound J. **17**(4), 900–909 (2020). https://doi.org/10.1111/iwj.13348
24. Iversen, M.M., Tell, G.S., Espehaug, B., Midthjell, K., Graue, M., Rokne, B., Berge, L.I., Østbye, T.: Is depression a risk factor for diabetic foot ulcers? 11-years follow-up of the Nord-Trøndelag health study (hunt). J. Diabetes Complicat. **29**(1), 20–25 (2015). https://doi.org/10.1016/j.jdiacomp.2014.09.006
25. Iversen, M., Igland, J., Smith-Strøm, H., Østbye, T., Tell, G., Skeie, S., Cooper, J., Peyrot, M., Graue, M.: Effect of a telemedicine intervention for diabetes-related foot ulcers on health, well-being and quality of life: secondary outcomes from a cluster randomized controlled trial (diafoto). BMC Endocr. Disord. **20**, 1 (2020). https://doi.org/10.1186/s12902-020-00637-x
26. Boulton, A.J., Vileikyte, L., Ragnarson-Tennvall, G., Apelqvist, J.: The global burden of diabetic foot disease. The Lancet **366**(9498), 1719–1724 (2005). https://doi.org/10.1016/S0140-6736(05)67698-2

27. Van Netten, J., Clark, D., Lazzarini, P., Janda, M., Reed, L.: The validity and reliability of remote diabetic foot ulcer assessment using mobile phone images. Sci. Rep. **7**, 1 (2017). https://doi.org/10.1038/s41598-017-09828-4
28. Apelqvist, J., Larsson, J., Agardh, C.-D.: Long-term prognosis for diabetic patients with foot ulcers. J. Intern. Med. **233**(6), 485–491 (1993). https://doi.org/10.1111/j.1365-2796.1993.tb01003.x
29. Larsson, J., Agardh, C.-D., Apelqvist, J., Stenström, A.: Long term prognosis after healed amputation in patients with diabetes. Clin. Orthop. Relat. Res. **350**, 149–58 (1998). https://doi.org/10.1097/00003086-199805000-00021
30. Mcgurnaghan, S., Weir, A., Bishop, J., Kennedy, S., Blackbourn, L., Mcallister, D., Hutchinson, S., Caparrotta, T., Mellor, J., Jeyam, A., O'Reilly, J., Wild, S., Hatam, S., Höhn, A., Colombo, M., Robertson, C., Lone, N., Murray, J., Butterly, E., McCoubrey, J.: Risks of and risk factors for Covid-19 disease in people with diabetes: a cohort study of the total population of Scotland. Lancet Diabetes Endocrinol. **9**, 1 (2020). https://doi.org/10.1016/S2213-8587(20)30405-8
31. Sharma, P., Behl, T., Sharma, N., Singh, S., Grewal, A.S., Albarrati, A., Albratty, M., Meraya, A.M., Bungau, S.: Covid-19 and diabetes: association intensify risk factors for morbidity and mortality. Biomed. Pharmacother. **151**, 113089 (2022). https://doi.org/10.1016/j.biopha.2022.113089
32. Khunti, K., Valabhji, J., Misra, S.: Diabetes and the Covid-19 pandemic. Diabetologia (2022). https://doi.org/10.1007/s00125-022-05833-z
33. Cassidy, B., Stringer, G., Yap, M.H.: Mobile framework for cognitive assessment: Trail making test and reaction time test. In: 2014 IEEE International Conference on Computer and Information Technology, pp. 700–705 (2014) https://doi.org/10.1109/CIT.2014.164
34. Zaidan, A., et al.: A review on smartphone skin cancer diagnosis apps in evaluation and benchmarking: coherent taxonomy, open issues and recommendation pathway solution. Heal. Technol. **8**, 1 (2018). https://doi.org/10.1007/s12553-018-0223-9
35. Cassidy, B., Reeves, N.D., Pappachan, J.M., Ahmad, N., Haycocks, S., Gillespie, D., Yap, M.: A cloud-based deep learning framework for remote detection of diabetic foot ulcers. IEEE Pervasive Comput. **01**, 1–9 (2022). https://doi.org/10.1109/MPRV.2021.3135686
36. Reeves, N.D., Cassidy, B., Abbott, C.A., Yap, M.H.: Chapter 7 - novel technologies for detection and prevention of diabetic foot ulcers. In: Gefen, A. (ed.) The Science, Etiology and Mechanobiology of Diabetes and Its Complications, pp. 107–122. Academic Press (2021). https://doi.org/10.1016/B978-0-12-821070-3.00007-6 . https://www.sciencedirect.com/science/article/pii/B9780128210703000076
37. Yap, M.H., Cassidy, B., Byra, M., Liao, T.-Y., Yi, H., Galdran, A., Chen, Y.-H., Brüngel, R., Koitka, S., Friedrich, C.M., Lo, Y.-W., Yang, C.-H., Li, K., Lao, Q., Ballester, M.A.G., Carneiro, G., Ju, Y.-J., Huang, J.-D., Pappachan, J.M., Reeves, N.D., Chandrabalan, V., Dancey, D., Kendrick, C.: Diabetic foot ulcers segmentation challenge report: Benchmark and analysis. Med. Image Anal. **94**, 103153 (2024). https://doi.org/10.1016/j.media.2024.103153
38. Yammine, K., Estephan, M.: Telemedicine and diabetic foot ulcer outcomes. a meta-analysis of controlled trials. The Foot (2021) https://doi.org/10.1016/j.foot.2021.101872
39. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017)
40. Brinker, T.J., Hekler, A., Enk, A.H., Klode, J., Hauschild, A., Berking, C., Schilling, B., Haferkamp, S., Schadendorf, D., Fröhling, S., Utikal, J.S., von Kalle, C.: A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. Eur. J. Cancer **111**, 148–154 (2019). https://doi.org/10.1016/j.ejca.2019.02.005
41. Brinker, T.J., Hekler, A., Enk, A.H., Berking, C., Haferkamp, S., Hauschild, A., Weichenthal, M., Klode, J., Schadendorf, D., Holland-Letz, T., von Kalle, C., Fröhling, S., Schilling, B., Utikal, J.S.: Deep neural networks are superior to dermatologists in melanoma image classification. Eur. J. Cancer **119**, 11–17 (2019). https://doi.org/10.1016/j.ejca.2019.05.023
42. Fujisawa, Y., Otomo, Y., Ogata, Y., Nakamura, Y., Fujita, R., Ishitsuka, Y., Watanabe, R., Okiyama, N., Ohara, K., Fujimoto, M.: Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. Br. J. Dermatol. **180**(2), 373–381 (2019). https://doi.org/10.1111/bjd.16924
43. Pham, T.C., Hoang, V.D., Tran, C.T., Luu, M.S.K., Mai, D.A., Doucet, A., Luong, C.M.: Improving binary skin cancer classification based on best model selection method combined with optimizing full connected layers of deep cnn. In: 2020 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), 1–6 (2020). https://doi.org/10.1109/MAPR49794.2020.9237778
44. Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe, Y., Hamamoto, R.: The development of a skin cancer classification system for pigmented skin lesions using deep learning. Biomolecules **10**(8), 1123 (2020)
45. Haenssle, H., Winkler, J., Fink, C., Toberer, F., Enk, A., Stolz, W., Kränke, T., Hofmann-Wellenhof, R., Kittler, H., Tschandl, P., Rosendahl, C., Lallas, A., Blum, A., Abassi, M., Thomas, L., Tromme, I., Rosenberger, A., Bachelerie, M., Bajaj, S., Zukervar, P.: Skin lesions of face and scalp - classification by a market-approved convolutional neural network in comparison with 64 dermatologists. Eur. J. Cancer **144**, 192–199 (2021). https://doi.org/10.1016/j.ejca.2020.11.034
46. Sheehan, P., Jones, P., Caselli, A., Giurini, J., Veves, A.: Percent change in wound area of diabetic foot ulcers over a 4-week period is a robust predictor of complete healing in a 12-week prospective trial. Diabetes Care **26**, 1879–82 (2003). https://doi.org/10.2337/diacare.26.6.1879
47. Cassidy, B., Yap, M.H., Pappachan, J., Ahmad, N., Haycocks, S., O'Shea, C., Fernandez, C., Chacko, E., Jacob, K., Reeves, N.: Artificial intelligence for automated detection of diabetic foot ulcers: a real-world proof-of-concept clinical evaluation. Diabetes Res. Clin. Pract. (2023). https://doi.org/10.1016/j.diabres.2023.110951
48. Ramachandram, D., Ramirez-GarciaLuna, J.L., Fraser, R.D.J., Martínez-Jiménez, M.A., Arriaga-Caballero, J.E., Allport, J.: Fully automated wound tissue segmentation using deep learning on mobile devices: Cohort study. JMIR Mhealth Uhealth **10**(4), 36977 (2022). https://doi.org/10.2196/36977
49. Ma, J., Nakarmi, U., Yue Sik Kin, C., Sandino, C., Cheng, J., Syed, A., Wei, P., Pauly, J., Vasanawala, S.: Diagnostic image quality assessment and classification in medical imaging: opportunities and challenges, 2020, pp. 337–340 (2020) https://doi.org/10.1109/ISBI45749.2020.9098735
50. Pham, H.N., Koay, C.Y., Chakraborty, T., Gupta, S., Tan, B.L., Wu, H., Vardhan, A., Nguyen, Q.H., Palaparthi, N.R., Nguyen, B.P., H. Chua, M.C.: Lesion segmentation and automated melanoma detection using deep convolutional neural networks and xgboost. In: 2019 International Conference on System Science and Engineering (ICSSE), pp. 142–147 (2019). https://doi.org/10.1109/ICSSE.2019.8823129
51. Kadampur, M.A., Al Riyaee, S.: Skin cancer detection: applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. Inform. Med. Unlocked **18**, 100282 (2020). https://doi.org/10.1016/j.imu.2019.100282

52. Gouda, W., Sama, N.U., Al-Waakid, G., Humayun, M., Jhanjhi, N.Z.: Detection of skin cancer based on skin lesion images using deep learning. Healthcare **10**(7), 1 (2022). https://doi.org/10.3390/healthcare10071183
53. Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D., Wang, T.: Deep learning in medical ultrasound analysis: a review. Engineering **5**(2), 261–275 (2019). https://doi.org/10.1016/j.eng.2018.11.020
54. Wang, L., Pedersen, P., Agu, E., Strong, D., Tulu, B.: Area determination of diabetic foot ulcer images using a cascaded two-stage SVM-based classification. IEEE Trans. Biomed. Eng. **1**, 1 (2016). https://doi.org/10.1109/TBME.2016.2632522
55. Illath, K., S, S., Swaminathan, R.: Analysis of chronic wound images using factorization based segmentation and machine learning methods, pp. 74–78 (2017). https://doi.org/10.1145/3155077.3155092
56. Kendrick, C., Cassidy, B., Reeves, N., Pappachan, J., O'Shea, C., Chandrabalan, V., Yap, M.H.: Diabetic Foot Ulcer Grand Challenge 2022 Summary (2023). https://doi.org/10.1007/978-3-031-26354-5_10
57. Zimmermann, R.S., Siems, J.N.: Faster training of mask r-CNN by focusing on instance boundaries. Comput. Vis. Image Underst. **188**, 102795 (2019). https://doi.org/10.1016/j.cviu.2019.102795
58. Liu, H., Xiong, W., Zhang, Y.: Yolo-core: contour regression for efficient instance segmentation. Mach. Intell. Res. **20**, 716–728 (2023). https://doi.org/10.1007/s11633-022-1379-3
59. Bai, B., Tian, J., Luo, S., Wang, T., Lyu, S.: Yuseg: Yolo and unet is all you need for cell instance segmentation. In: Ma, J., Xie, R., Gupta, A., Almeida, J., Bader, G.D., Wang, B. (eds.), Proceedings of The Cell Segmentation Challenge in Multi-modality High-Resolution Microscopy Images. Proceedings of Machine Learning Research, vol. 212, pp. 1–15. PMLR (2023). https://proceedings.mlr.press/v212/bai23a.html
60. Wang, C., Yan, X., Smith, M., Kochhar, K., Rubin, M., Warren, S.M., *et al.*: A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. In: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, pp. 2415–2418 (2015). IEEE
61. Rother, C., Kolmogorov, V., Blake, A.: "grabcut": interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. **23**(3), 309–314 (2004). https://doi.org/10.1145/1015706.1015720
62. Goyal, M., Yap, M.H., Reeves, N.D., Rajbhandari, S., Spragg, J.: Fully convolutional networks for diabetic foot ulcer segmentation. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 618–623 (2017). https://doi.org/10.1109/SMC.2017.8122675
63. Li, F., Wang, C., Xiaohui, L., Peng, Y., Jin, S.: A composite model of wound segmentation based on traditional methods and deep neural networks. Comput. Intell. Neurosci. **2018**, 1–12 (2018). https://doi.org/10.1155/2018/4149103
64. Elmogy, M., García-Zapirain, B., Elmaghraby, A.S., El-Baz, A.: An automated classification framework for pressure ulcer tissues based on 3d convolutional neural network. In: 2018 24th International Conference on Pattern Recognition (ICPR), pp. 2356–2361 (2018). https://doi.org/10.1109/ICPR.2018.8546081
65. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017). https://doi.org/10.1016/j.media.2016.10.004
66. Thomas, S.: Medetec. http://www.medetec.co.uk/index.html. lastaccess:08/11/21 (2014)
67. Zapirain, B., Elmogy, M., El-Baz, A., Elmaghraby, A.: Classification of pressure ulcer tissues with 3d convolutional neural network. Med. Biol. Eng. Comput. **56**, 1–14 (2018). https://doi.org/10.1007/s11517-018-1835-y
68. Godeiro, V., Neto, J.S., Carvalho, B., Santana, B., Ferraz, J., Gama, R.: Chronic wound tissue classification using convolutional networks and color space reduction. In: 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6 (2018). https://doi.org/10.1109/MLSP.2018.8517026
69. Zahia, S., Sierra-Sosa, D., Garcia-Zapirain, B., Elmaghraby, A.: Tissue classification and segmentation of pressure injuries using convolutional neural networks. Comput. Methods Programs Biomed. **159**, 51–58 (2018). https://doi.org/10.1016/j.cmpb.2018.02.018
70. Pathompatai, C., Kanawong, R., Taeprasartsit, P.: Region-focus training: Boosting accuracy for deep-learning image segmentation. In: 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 319–323 (2019). https://doi.org/10.1109/JCSSE.2019.8864162
71. Li, F., Wang, C., Peng, Y., Yuan, Y., Jin, S.: Wound segmentation network based on location information enhancement. IEEE Access **7**, 87223–87232 (2019). https://doi.org/10.1109/ACCESS.2019.2925689
72. Cui, C., Thurnhofer-Hemsi, K., Soroushmehr, R., Mishra, A., Gryak, J., Domínguez, E., Najarian, K., López-Rubio, E.: Diabetic wound segmentation using convolutional neural networks. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1002–1005 (2019). https://doi.org/10.1109/EMBC.2019.8856665
73. Jafari, M.H., Karimi, N., Nasr-Esfahani, E., Samavi, S., Soroushmehr, S.M.R., Ward, K., Najarian, K.: Skin lesion segmentation in clinical images using deep learning. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 337–342 (2016). https://doi.org/10.1109/ICPR.2016.7899656
74. Ohura, N., Mitsuno, R., Sakisaka, M., Terabe, Y., Morishige, Y., Uchiyama, A., Okoshi, T., Shinji, I., Takushima, A.: Convolutional neural networks for wound detection: the role of artificial intelligence in wound care. J. Wound Care **28**(10), 13–24 (2019). https://doi.org/10.12968/jowc.2019.28.Sup10.S13
75. Wagh, A., Jain, S., Mukherjee, A., Agu, E., Pedersen, P.C., Strong, D., Tulu, B., Lindsay, C., Liu, Z.: Semantic segmentation of smartphone wound images: comparative analysis of AHRF and CNN-based approaches. IEEE Access **8**, 181590–181604 (2020). https://doi.org/10.1109/ACCESS.2020.3014175
76. Maninis, K.-K., Caelles, S., Pont-Tuset, J., Van Gool, L.: Deep extreme cut: From extreme points to object segmentation, pp. 616–625 (2018). https://doi.org/10.1109/CVPR.2018.00071
77. Ong, E.P., Tang Ka Yin, C., Lee, B.-H.: Efficient deep learning-based wound-bed segmentation for mobile applications. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1654–1657 (2020). https://doi.org/10.1109/EMBC44109.2020.9176299
78. Wang, C., Anisuzzaman, D.M., Williamson, V., Dhar, M.K., Rostami, B., Niezgoda, J., Gopalakrishnan, S., Yu, Z.: Fully automatic wound segmentation with deep convolutional neural networks. Sci. Rep. **10**(1), 1–9 (2020). https://doi.org/10.1038/s41598-020-78799-w
79. Wang, C., Mahbod, A., Ellinger, I., Galdran, A., Gopalakrishnan, S., Niezgoda, J., Yu, Z.: FUSeg: The Foot Ulcer Segmentation Challenge (2022). https://doi.org/10.48550/ARXIV.2201.00414 . https://arxiv.org/abs/2201.00414
80. Mahbod, A., Ecker, R., Ellinger, I.: Automatic foot ulcer segmentation using an ensemble of convolutional neural networks. arXiv:2109.01408 (2021)
81. Zahia, S., Zapirain, B., Elmaghraby, A.: Integrating 3d model representation for an accurate non-invasive assessment of pressure

injuries with deep learning. Sensors **20**, 1 (2020). https://doi.org/10.3390/s20102933
82. Niri, R., Hassan, D., Lucas, Y., Treuillet, S.: A Superpixel-Wise Fully Convolutional Neural Network Approach for Diabetic Foot Ulcer Tissue Classification, pp. 308–320 (2021). https://doi.org/10.1007/978-3-030-68763-2_23
83. Chairat, S., Dissaneewate, T., Wangkulangkul, P., Kongpanichakul, L., Chaichulee, S.: Non-contact chronic wound analysis using deep learning. In: 2021 13th Biomedical Engineering International Conference (BMEiCON), pp. 1–5 (2021).https://doi.org/10.1109/BMEiCON53485.2021.9745246
84. Watanabe, R., Shima, K., Horiuchi, T., Shimizu, T., Mukaeda, T., Shimatani, K.: A system for wound evaluation support using depth and image sensors. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 3709–3712 (2021). https://doi.org/10.1109/EMBC46164.2021.9629922
85. Ichioka, S., Sugama, J.: Multidisciplinary approach to chronic wounds. Shōrinsha, Japan (2011)
86. Niri, R., Gutierrez, E., Douzi, H., Lucas, Y., Treuillet, S., Castaneda, B., Hernandez, I.: Multi-view data augmentation to improve wound segmentation on 3d surface model by deep learning. IEEE Access **9**, 157628–157638 (2021). https://doi.org/10.1109/ACCESS.2021.3130784
87. Scebba, G., Zhang, J., Catanzaro, S., Mihai, C., Distler, O., Berli, M., Karlen, W.: Detect-and-segment: a deep learning approach to automate wound image segmentation. Inform. Med. Unlocked **29**, 100884 (2022). https://doi.org/10.1016/j.imu.2022.100884
88. Cassidy, B., Kendrick, C., Reeves, N., Pappachan, J., O'Shea, C., Armstrong, D., Yap, M.H.: Diabetic Foot Ulcer Grand Challenge 2021: Evaluation and Summary, pp. 90–105 (2022). https://doi.org/10.1007/978-3-030-94907-5_7
89. Xing, L., Li, L., Wang, Z., Ma, H.: An improved unet model for foot ulcer image segmentation. In: 2022 7th International Conference on Image, Vision and Computing (ICIVC), 393–397 (2022).https://doi.org/10.1109/ICIVC55077.2022.9886343
90. Bose, S., Sur Chowdhury, R., Das, R., Maulik, U.: Dense dilated deep multiscale supervised u-network for biomedical image segmentation. Comput. Biol. Med. **143**, 105274 (2022). https://doi.org/10.1016/j.compbiomed.2022.105274
91. Chang, C., Christian, M., Chang, D.H., Lai, F., Liu, T., Chen, Y., Chen, W.: Deep learning approach based on superpixel segmentation assisted labeling for automatic pressure ulcer diagnosis. PLoS ONE **17**, 0264139 (2022). https://doi.org/10.1371/journal.pone.0264139
92. Curti, N., Merli, Y., Zengarini, C., Giampieri, E., Merlotti, A., Dall'Olio, D., Marcelli, E., Bianchi, T., Castellani, G.: Effectiveness of semi-supervised active learning in automated wound image segmentation. Int. J. Mol. Sci. **24**, 706 (2022). https://doi.org/10.3390/ijms24010706
93. Kendrick, C., Cassidy, B., Pappachan, J.M., O'Shea, C., Fernandez, C.J., Chacko, E., Jacob, K., Reeves, N.D., Yap, M.H.: Translating clinical delineation of diabetic foot ulcers into machine interpretable segmentation. In: Yap, M.H., Kendrick, C., Brüngel, R. (eds.) Diabetic Foot Ulcers Grand Challenge, pp. 1–14. Springer, Cham (2025)
94. Liao, T.-Y., Yang, C.-H., Lo, Y.-W., Lai, K.-Y., Shen, P.-H., Lin, Y.-L.: HarDNet-DFUS: An Enhanced Harmonically-Connected Network for Diabetic Foot Ulcer Image Segmentation and Colonoscopy Polyp Segmentation (2022). https://doi.org/10.48550/ARXIV.2209.07313
95. Ramachandram, D., Ramírez-GarcíaLuna, J., Fraser, R., Martínez-Jiménez, M.A., Arriaga-Caballero, J., Allport, J.: Improving objective wound assessment: fully-automated wound tissue segmentation using deep learning on mobile devices. JMIR Mhealth Uhealth (2022). https://doi.org/10.2196/36977
96. Marijanović, D., Nyarko, E.K., Filko, D.: Wound detection by simple feedforward neural network. Electronics **11**, 3 (2022). https://doi.org/10.3390/electronics11030329
97. Swerdlow, M., Guler, O., Yaakov, R., Armstrong, D.G.: Simultaneous segmentation and classification of pressure injury image data using mask-r-cnn. Comput. Math. Methods Med. **2023**, 1–7 (2023). https://doi.org/10.1155/2023/3858997
98. Liu, T., Wang, H., Christian, M., Chang, C.-W., Lai, F., Tai, H.-C.: Automatic segmentation and measurement of pressure injuries using deep learning models and a lidar camera. Sci. Rep. **13**, 1 (2023). https://doi.org/10.1038/s41598-022-26812-9
99. Lan, T., Li, Z., Chen, J.: Fusionsegnet: fusing global foot features and local wound features to diagnose diabetic foot. Comput. Biol. Med. **152**, 106456 (2023). https://doi.org/10.1016/j.compbiomed.2022.106456
100. Oota, S.R., Rowtula, V., Mohammed, S., Liu, M., Gupta, M.: Wsnet: Towards an effective method for wound image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 3234–3243 (2023)
101. Privalov, M., Beisemann, N., El Barbari, J., Mandelka, E., Müller, M., Syrek, H., Grützner, P., Vetter, S.: Software-based method for automated segmentation and measurement of wounds on photographs using mask r-cnn: a validation study. J. Digit. Imaging **34**, 1–10 (2021). https://doi.org/10.1007/s10278-021-00490-x
102. Monroy, B., Bacca, J., Sanchez, K., Arguello, H., Castillo, S.: Two-step deep learning framework for chronic wounds detection and segmentation: A case study in Colombia. In: 2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), pp. 1–6 (2021). https://doi.org/10.1109/STSIVA53688.2021.9592008
103. Foltynski, P., Ładyżyński, P.: Internet service for wound area measurement using digital planimetry with adaptive calibration and image segmentation with deep convolutional neural networks. Biocybern. Biomed. Eng. (2022). https://doi.org/10.1016/j.bbe.2022.11.004
104. Oliveira, B., Torres, H., Morais, P., Veloso, F., Baptista, A., Fonseca, J., Vilaça, J.: A multi-task convolutional neural network for classification and segmentation of chronic venous disorders. Sci. Rep. **13**, 761 (2023). https://doi.org/10.1038/s41598-022-27089-8
105. Wijesinghe, I., Gamage, C., Perera, I., Chitraranjan, C.: A smart telemedicine system with deep learning to manage diabetic retinopathy and foot ulcers. In: 2019 Moratuwa Engineering Research Conference (MERCon), pp. 686–691 (2019) https://doi.org/10.1109/MERCon.2019.8818682
106. Gamage, C., Wijesinghe, I., Perera, I.: Instance-Based Segmentation for Boundary Detection of Neuropathic Ulcers Through Mask-RCNN, pp. 511–522 (2019). https://doi.org/10.1007/978-3-030-30493-5_49
107. Cao, C., Qiu, Y., Wang, Z., Ou, J., Wang, J., Hounye, A.H., Hou, M., Zhou, Q., Zhang, J.: Nested segmentation and multi-level classification of diabetic foot ulcer based on mask r-CNN. Multimedia Tools Appl. (2022). https://doi.org/10.1007/s11042-022-14101-6
108. McBride, C., Cassidy, B., Kendrick, C., Reeves, N., Pappachan, J., Yap, M.H.: Multi-colour space channel selection for improved chronic wound segmentation, pp. 1–5 (2024). https://doi.org/10.1109/ISBI56570.2024.10635155
109. Pewton, S., Cassidy, B., Kendrick, C., Yap, M.H.: Dermoscopic dark corner artifacts removal: friend or foe? Comput. Methods Programs Biomed. (2023). https://doi.org/10.1016/j.cmpb.2023.107986
110. Dipto, I., Cassidy, B., Kendrick, C., Reeves, N., Pappachan, J., Chandrabalan, V., Yap, M.H.: Quantifying the Effect of Image Similarity on Diabetic Foot Ulcer Classification, pp. 1–18 (2023). https://doi.org/10.1007/978-3-031-26354-5_1

111. Cassidy, B., Kendrick, C., Brodzicki, A., Jaworek-Korjakowska, J., Yap, M.H.: Analysis of the ISIC image datasets: usage, benchmarks and recommendations. Med. Image Anal. (2021). https://doi.org/10.1016/j.media.2021.102305
112. Daneshjou, R., Smith, M., Sun, M., Rotemberg, V., Zou, J.: Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review. JAMA Dermatol. (2021). https://doi.org/10.1001/jamadermatol.2021.3129
113. Wen, D., Khan, S.M., Xu, A.J., Ibrahim, H., Smith, L., Caballero, J., Zepeda, L., de Blas Perez, C., Denniston, A.K., Liu, X., Matin, R.N.: Characteristics of publicly available skin cancer image datasets: a systematic review. Lancet Digit. Health (2021). https://doi.org/10.1016/S2589-7500(21)00252-1
114. Okafor, N., Cassidy, B., O'Shea, C., Pappachan, J., Yap, M.H.: The Effect of Image Preprocessing Algorithms on Diabetic Foot Ulcer Classification, pp. 336–352 (2024). https://doi.org/10.1007/978-3-031-66958-3_25
115. Perez, F., Vasconcelos, C., Avila, S., Valle, E.: Data Augmentation for Skin Lesion Analysis, pp. 303–311 (2018).https://doi.org/10.1007/978-3-030-01201-4_33
116. Cassidy, B., Reeves, N.D., Pappachan, J.M., Gillespie, D., O'Shea, C., Rajbhandari, S., Maiya, A.G., Frank, E., Boulton, A.J.M., Armstrong, D.G., Najafi, B., Wu, J., Kochhar, R.S., Yap, M.H.: The DFUC 2020 dataset: analysis towards diabetic foot ulcer detection. touchREVIEWS Endocrinol **17**, 5–11 (2021). https://doi.org/10.17925/EE.2021.17.1.5
117. Yap, M.H., Cassidy, B., Pappachan, J.M., O'Shea, C., Gillespie, D., Reeves, N.D.: Analysis towards classification of infection and ischaemia of diabetic foot ulcers. In: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), pp. 1–4 (2021). https://doi.org/10.1109/BHI50953.2021.9508563
118. Yap, M.H., Kendrick, C., Reeves, N., Goyal, M., Pappachan, J., Cassidy, B.: Development of diabetic foot ulcer datasets: an overview, pp. 1–18 (2022). https://doi.org/10.1007/978-3-030-94907-5_1
119. Pappachan, J.M., Cassidy, B., Fernandez, C.J., Chandrabalan, V., Yap, M.H.: The role of artificial intelligence technology in the care of diabetic foot ulcers: the past, the present, and the future. World J. Diabetes **13**, 1131–1139 (2022). https://doi.org/10.4239/wjd.v13.i12.1131
120. Lorè, F., Basile, P., Appice, A., Gemmis, M., Malerba, D., Semeraro, G.: An AI framework to support decisions on GDPR compliance. J. Intell. Inf. Syst. **61**, 1–28 (2023). https://doi.org/10.1007/s10844-023-00782-4
121. Quezada-Tavarez, K., Dutkiewicz, L., Krack, N.: Voicing challenges: Gdpr and AI research. Open Res Europe **2**, 126 (2022). https://doi.org/10.12688/openreseurope.15145.1
122. Cassidy, B., McBride, C., Kendrick, C., Reeves, N.D., Pappachan, J.M., Fernandez, C.J., Chacko, E., Brüngel, R., Friedrich, C.M., Alotaibi, M., AlWabel, A.A., Alderwish, M., Lai, K.-Y., Yap, M.H.: An enhanced harmonic densely connected hybrid transformer network architecture for chronic wound segmentation utilising multi-colour space tensor merging. arXiv:2410.03359 (2024)
123. Yang, S., Park, J., Lee, H., Kim, S., Lee, B.-U., Chung, K.-Y., Oh, B.: Sequential change of wound calculated by image analysis using a color patch method during a secondary intention healing. PLoS ONE (2016). https://doi.org/10.1371/journal.pone.0163092
124. Alzubaidi, L., Fadhel, M.A., Oleiwi, S.R., Al-Shamma, O., Zhang, J.: Dfu_qutnet: diabetic foot ulcer classification using novel deep convolutional neural network. Multimedia Tools Appl. **79**(21–22), 15655–15677 (2020). https://doi.org/10.1007/s11042-019-07820-w
125. Kręcichwost, M., Czajkowska, J., Wijata, A., Juszczyk, J., Pyciński, B., Biesok, M., Rudzki, M., Majewski, J., Kostecki, J., Pietka, E.: Chronic wounds multimodal image database. Comput. Med. Imaging Graph. (2021). https://doi.org/10.1016/j.compmedimag.2020.101844



Bill Cassidy is a PhD student at Manchester Metropolitan University. He is currently studying multimodal medical image segmentation, focusing on chronic wounds and skin cancer. He has conducted the most extensive clinical evaluation and reliability studies to date in chronic wounds deep learning research.

Connah Kendrick is a lecturer in computer graphics at the Manchester Metropolitan University. Receiving his PhD in computer vision in 2019, focusing on facial landmarking with 3D positional data. Later worked as a postdoctoral research associate for a year before his lectureship. As a postdoctoral research associate, he developed both tools and deep learning AIs for multi-disciplinary teams. During his lecture role, he teaches java to apprenticeship students while continuing his research. His research expertise is computer vision, deep learning, and 3D data.

Neil D. Reeves is currently professor of Secure Health Technologies based in the Medical School at Lancaster University. His research is focused on digital health technologies and cyber security, diabetic foot ulcer prevention, digital technologies and medical devices in diabetic foot research, gait impairment in diabetes, and exercise-based interventions.

Joseph M. Pappachan is currently a consultant in endocrinology and metabolism and research lead at Countess of Chester Hospital NHS Trust. He is also an honorary professor with Manchester Metropolitan University, Manchester, UK. His current research interests include obesity, diabetic foot disease, neuroendocrinology, musculoskeletal metabolism, and parathyroid disease.

Moi Hoon Yap is professor of image and vision computing and the research lead of the Department of Computing and Mathematics, Manchester Metropolitan University. Moi Hoon has received research funding from the Royal Society, EU Funding, EPSRC, Innovate UK, Cancer Research UK, and industry partners. She is leading computer vision research in medical image analysis and biometrics, including face and gesture analysis.