Observational Study

Medicine

OPEN

Development of a pressure ulcer stage determination system for community healthcare providers using a vision transformer deep learning model

Young-Bok Cho, PhD[a], Hana Yoo, PhD[b,*]

Abstract

This study reports the first steps toward establishing a computer vision system to help caregivers of bedridden patients detect pressure ulcers (PUs) early. While many previous studies have focused on using convolutional neural networks (CNNs) to elevate stages, hardware constraints have presented challenges related to model training and overreliance on medical opinions. This study aimed to develop a tool to classify PU stages using a Vision Transformer model to process actual PU photos. To do so, we used a retrospective observational design involving the analysis of 395 images of different PU stages that were accurately labeled by nursing specialists and doctors from 3 hospitals. In the pressure ulcer cluster vision transformer (PUC-ViT) model classifies the PU stage with a mean ROC curve value of 0.936, indicating a model accuracy of 97.76% and F1 score of 95.46%. We found that a PUC-ViT model showed higher accuracy than conventional models incorporating CNNs, and both effectively reduced computational complexity and achieved low floating point operations per second. Furthermore, we used internet of things technologies to propose a model that allows anyone to analyze input images even at low computing power. Based on the high accuracy of our proposed model, we confirm that it enables community caregivers to detect PUs early, facilitating medical referral.

Abbreviations: AUC = area under the ROC curve, CNN = convolutional neural network, EM = experimental minimal, EM_fcnn = experimental minimal full convolution neural network, PU = pressure ulcer, PUC-ViT = pressure ulcer cluster vision transformer, sViT = semi ViT, ViT = vision transformer.

Keywords: deep learning, healthcare providers, pressure ulcer

1. Introduction

Pressure ulcers (PUs) are defined as localized damage to the skin and/or underlying tissue, resulting from prolonged pressure or pressure combined with shear.[1] PUs are among the most frequently used indicators for measuring healthcare quality worldwide. The most commonly used PU classifications include the guidelines of the European pressure ulcer advisory panel (EPUAP), national pressure injury advisory panel (NPIAP), and the pan Pacific pressure injury alliance (PPPIA), respectively. Stage 1 PU represents intact skin coupled with nonblanchable redness over a localized area, often bony prominence. Stage 2 PU includes partial thickness loss of the dermis, which presents as a shallow open ulcer with a red–pink wound bed without slough. Stage 3 PU includes full-thickness tissue loss. Subcutaneous fat may be visible; however, the underlying bone, tendon, or muscle tissues are not exposed. Stage 4 PU involves full-thickness tissue loss with exposed bones, tendons, or muscles. Slough or eschar may be present in some areas of the wound bed. Unstageable PU presents full-thickness tissue loss, in which the base of the ulcer is covered by slough (i.e., yellow, tan, gray, green, or brown) and/or eschar (i.e., tan, brown, or black) in the wound bed. The stage of suspected deep tissue injury can be characterized by purple or maroon localized areas of discolored intact skin or by the presence of blood-filled blisters caused by damage to the underlying soft tissue from pressure and/or shear.[1]

PU prevalence was found to be 27% in Italy,[2] 18.5% in Ireland,[3] and 3.7% to 18% in Australia.[4] Globally, the

This research was supported by the Daejeon University Research Grant (2022).

Written informed consent was waived because the retrospective datasets used were anonymous.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

This study was conducted in accordance with the guidelines of the declaration of Helsinki and all procedures were approved by the Institutional Review Board of Daejeon University (IRB No. 1040647-202310-HR-006-01).

[a] Department of Computer Education, Andong National University, andong, Gyeongsangbuk-do, Republic of Korea, [b] Department of Nursing, Daejeon University, Dong-gu, Daejeon, Republic of Korea.

** Correspondence: Hana Yoo, Department of Nursing, Daejeon University, 62, Daehak-ro, Dong-gu 34520, Daejeon, Republic of Korea (e-mail: hanayoo@dju.kr).*

How to cite this article: Cho Y-B, Yoo H. Development of a pressure ulcer stage determination system for community healthcare providers using a vision transformer deep learning model. Medicine 2025;104:7(e41530).

Received: 31 May 2024 / Received in final form: 4 December 2024 / Accepted: 27 January 2025

http://dx.doi.org/10.1097/MD.0000000000041530

prevalence of PU is estimated to be 11.6%, with a mean incidence of 6.3%.[5] These values suggest that PU prevalence and incidence can vary widely among countries and that estimates reflect differences in measurement tools, settings, and healthcare providers. Moreover, because humans live longer, the number of people affected by PU is expected to increase. Complex wounds such as PUs do not heal quickly and tend to worsen over time if not treated properly.[6] In addition, PUs generally increase the length of hospital stays, thereby intensifying resource demands and imposing high financial burdens, and can even cause lifelong difficulties for patients if repeated injuries occur in an affected area. Therefore, early detection of at-risk patients and a tailored management plan for PUs can reduce its incidence.[7] The ability of nurses to make a differential diagnosis in PU stage classification can be improved via training[8]; however, this relies on the experience and competence of healthcare professionals. Research on PU stage classification using deep learning methods is currently underway to reduce errors in PU stage discrimination by nurses and doctors.[9] Recently, such studies have primarily focused on hospitalized patients.[9] While it is important to determine the exact stage of a PU in the hospital, early PU detection and subsequent referral are also important for patients receiving home care in the community. Therefore, it is necessary to develop tools capable of helping nonexperts caring for bedridden patients and/or patients with mobility limitations living in the community to detect PUs early and accurately determine their stage.

Accordingly, the purpose of this study was to develop a deep learning-based PU stage classification model based on patient images and to confirm that this model is applicable for use cases involving community-based caregiver assessments of patients at home.

2. Methods

2.1. Research design

This was a retrospective observational study in which deep learning methods were applied to evaluate previously collected PU images.

2.2. Study participants

For this study, we selected 395 images out of 608 images from a dataset of images correctly diagnosed by experts. These images were provided by 3 hospitals in City D, South Korea, and were taken between 2020 and 2022 using a smartphone, with the original images in high dynamic range at 3840 × 2160 = 8.3 million pixels. The collected PU data was reviewed by an expert, and each image was labeled with a region of interest to distinguish the stage of PUs, resulting in a 112 × 112 × 3 dataset of 395 images, with significant differences in the distribution of each stage. Based on a previous research[10] indicating that deep learning model performance is affected by training data size, we preprocessed the selected 395 PU images into 112 × 112 × 3 and obtained 16,850 final images for model training by image data augmentation through geometric feature modification (Fig. 1).

2.3. Automatic lesion detection algorithm using a transformer model (pressure ulcer clustering-vision transformer)

In this study, the data analysis pipeline involved image preprocessing, feature extraction, feature selection, classification, and prediction based on images generated by users. Preprocessing tasks, such as extracting regions of interest from input images, noise filtering, resampling, and segmentation, can be selectively performed. Moreover, although many previous studies have focused on elevating stages using convolutional neural networks, various technical issues have arisen owing to the high hardware requirements for model training and the degree of reliance on medical opinions. Therefore, this paper proposes to enhance the performance of the proposed model by preprocessing image data based on previously acquired images, dividing the original images into smaller patches, and embedding them using a vision transformer (ViT) architecture commonly used in computer vision. The introduction of this model was followed by fine-tuning to extract image features and improve the performance of the proposed model.

2.3.1. Image data augmentation. The performance of deep learning models depends on how the model is optimized and the composition of the training dataset. It is well known that more generalized datasets generate better performance, but data characterizing PUs is sparse. Moreover, if the distribution of the training data is skewed toward certain strata, the model can overfit – that is, it can learn only about specific components but not about others, which can prevent a generalized learning model from being obtained. Various methods are used to solve this problem, including dropout protocols, batch normalization, data augmentation, and generalization, but the best way is to train using a large dataset. Here, to address the problem of overfitting, we augmented PU data with image data by transforming the geometric features of images. To perform this augmentation of the training data, we set the left and right parallel shift pixel range of images from −30 to 30, the rotation angle range from −90 to 90 degrees, and the scaling range from 0.5 to 1.5. Flipping was also applied to produce new datapoints with characteristics that were similar to those of the original dataset. Finally, resolution was set to 3024, 3452, 4608, 4032, 3648, and 2736 pixels to enlarge the image. This was randomly determined at the time of training.

Next, PU images were categorized into 5 stages based on PU progression. This ranged from stage 1 PUs to stage 4 PUs, with stage 5 disease associated with suspected deep tissue damage. In the PU dataset, stages 1 and 2 were the most common, comprising 75% of the dataset, whereas stage 4 was sparser, accounting for only 1.1% of the total data, indicating significant bias. To address the interclass imbalance caused by data augmentation, we generated data frames that were similar to the original data while keeping the number of samples per class the same. We then measured data balance via a 10-fold hierarchical cross-validation of the training and testing phases using a random split (90:10) of the total dataset. To determine whether the image data generated using each augmentation technique affected model performance, we ran experiments with 4 different baselines: training with 20% of the training data converted to the generated data, training with 50% of the validation data converted to the generated data, training with 50% of the training data converted to the generated data, and training with a mixture of 50% training data and 50% validation data. For each cross-validation test, we used a data ratio of 8:1:1 between the training, validation, and test datasets (Fig. 2).

2.3.2. Data preprocessing. In a typical study on medical image segmentation using semisupervised learning, the labeled training dataset is commonly designated as T, the unlabeled training dataset is U, and a separate set is allocated for testing purposes. For the labeled training and testing sets, we represent a batch of labeled data as $(X_l, Y_{gt}) \in L, (X_t, Y_{gt}) \in T$, along with its corresponding ground truth. In the unlabeled training set, we denote a batch of raw data as $(X_u) \in U$, where $X \in \mathbb{R}^{h \times w}$ represents a 2D grayscale image. Furthermore, Y_p is the dense map predicted by the segmentation model $f(\theta) : X \rightarrow Y_p$, with θ representing the parameter set of the model f. Y_p can be considered a pseudo labeled batch of data used to retrain models using unlabeled data $(X_u, Y_p) \in U$. Final evaluation results are calculated based on the differences between Y_p and Y_{gt} of T. The training objective of Pressure Ulcer Clustering-ViT (PUC-ViT) is

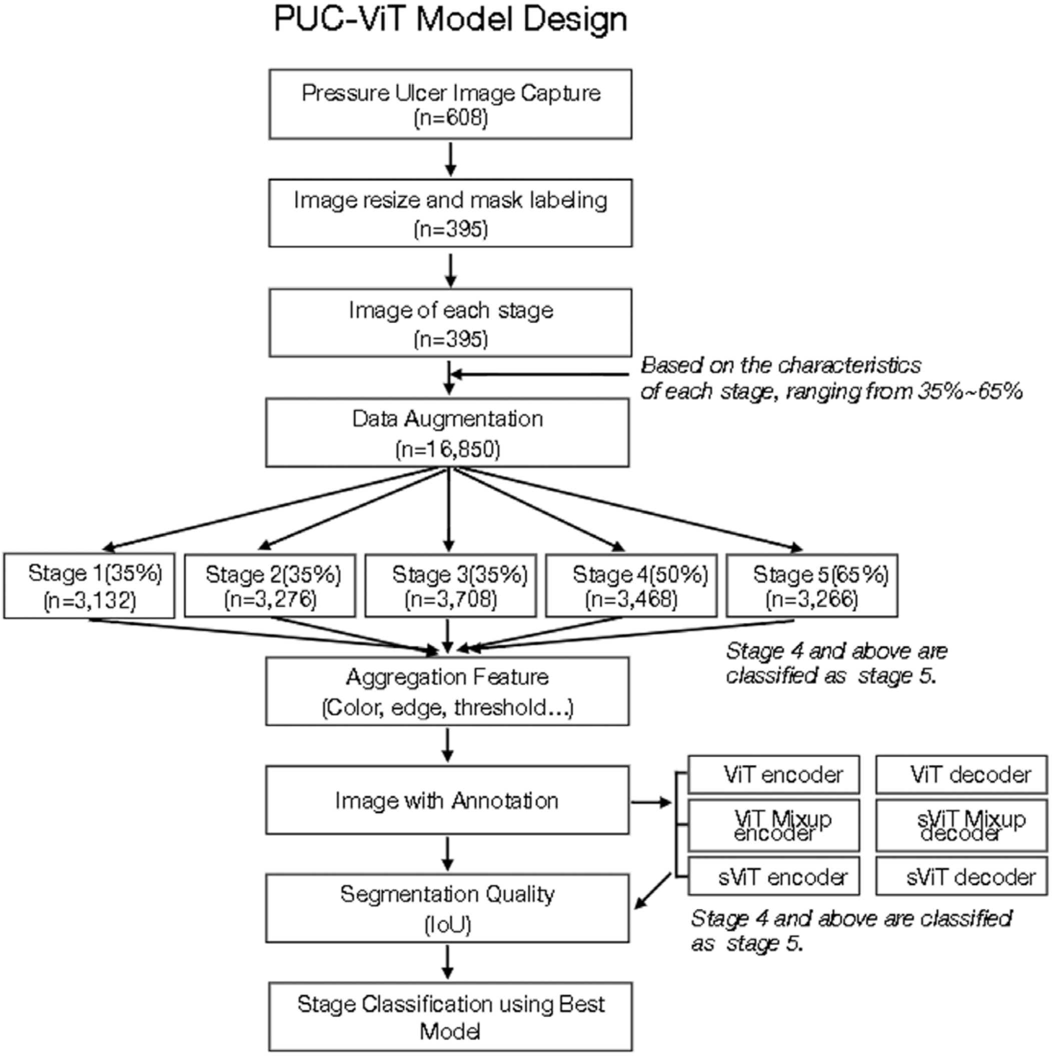
PUC-ViT Model Design
Pressure Ulcer Image Capture (n=608)
Image resize and mask labeling (n=395)
Image of each stage (n=395)
Based on the characteristics of each stage, ranging from 35%~65%
Data Augmentation (n=16,850)
Stage 1(35%) (n=3,132)
Stage 2(35%) (n=3,276)
Stage 3(35%) (n=3,708)
Stage 4(50%) (n=3,468)
Stage 5(65%) (n=3,266)
Stage 4 and above are classified as stage 5.
Aggregation Feature (Color, edge, threshold...)
Image with Annotation
ViT encoder
ViT decoder
ViT Mixup encoder
sViT Mixup decoder
sViT encoder
sViT decoder
Segmentation Quality (IoU)
Stage 4 and above are classified as stage 5.
Stage Classification using Best Model

Figure 1. Dataset collection and deep learning modeling of PUC-ViT design. PUC-ViT = pressure ulcer clustering-vision transformer.

to minimize the supervision loss L_{super} and the semivision loss L_{semi} of semi ViT (sViT), along with the supervision loss L_{super} of EM_{CNN}. The losses L_{super} and L_{semi} differ only in their use of labeled or unlabeled data (i.e., X_l or X_u, respectively).

The framework of PUC-ViT, including an adversarial training stage and a MixUp-based interpolation consistency training stage, as well as the corresponding loss functions involved, are briefly illustrated in Figure 3.

The motivation behind PUC-ViT primarily stems from adversarial training, as previously proposed.[11] Moreover, previous discussions on consistency training[12,13] have aimed to harness the capabilities of vision transformers. The training process of the proposed PUC-ViT model consists of 3 steps. These steps were designed to include interpolation consistency training for sViT and 2 adversarial training steps for sViT and EM_{CNN} (Fig. 3). In addition, we computed 2 types of supervision loss and 2 types of semisupervision loss.

PUC-ViT is primarily motivated by adversarial training[14] and consistency training[15] to explore the power of vision transformers. To achieve this, 3 different training stages, including 1 MixUp-based interpolation consistency training stage for PUC-ViT and 2 separate adversarial training stages for sViT($f_{ViT}(\theta) : X \rightarrow Y_p$) and the CNN-based evaluation model (EM_{CNN}, $f_{CNN}(\theta) : f_{ViT}(X;\theta) \rightarrow Y_{quality}$), were then compared with each other in an iterative manner.

Next, the ViT-based model takes the input images X and infers the corresponding segmentation feature map Y_p. Moreover, it does so while incorporating a suggestive loss from the EM_{CNN} (also known as the discriminator). Specifically, EM_{CNN} takes the inference of sViT Y_p and its corresponding input image(s) X and infers a quality score (i.e., $Y_{quality}$). In PUC-ViT, the discriminator is designed to classify the quality of Y_p using the set $Y_{quality} \in (0, 1)$. This is considered a binary classification task, in which 1: high-quality inference and/or images with annotations and 0: low-quality inference, including unannotated images. Next, the architecture of EM_{CNN} was used with ResNet50. In either of the 2 adversarial training stages (i.e., sViT and EM_{CNN}), 1 model was trained and the other model was evaluated (Fig. 3).

The objective of EM_{CNN} is to differentiate between the inferences made by sViT using annotated inputs to predict the values made using unannotated input. Conversely, sViT is incentivized to produce predictions from unannotated inputs that match the quality of those generated from annotated inputs. During training, sViT was found to improve the quality of its inferences by incorporating feedback from EM_{CNN}; we therefore aimed to achieve consistent quality regardless of the annotations. Meanwhile, the EM_{CNN} is trained to distinguish between the results obtained from annotated and unannotated inputs by leveraging known input indices with labels. This collaborative process enabled both models to attain post-training proficiency in segmentation and discrimination.

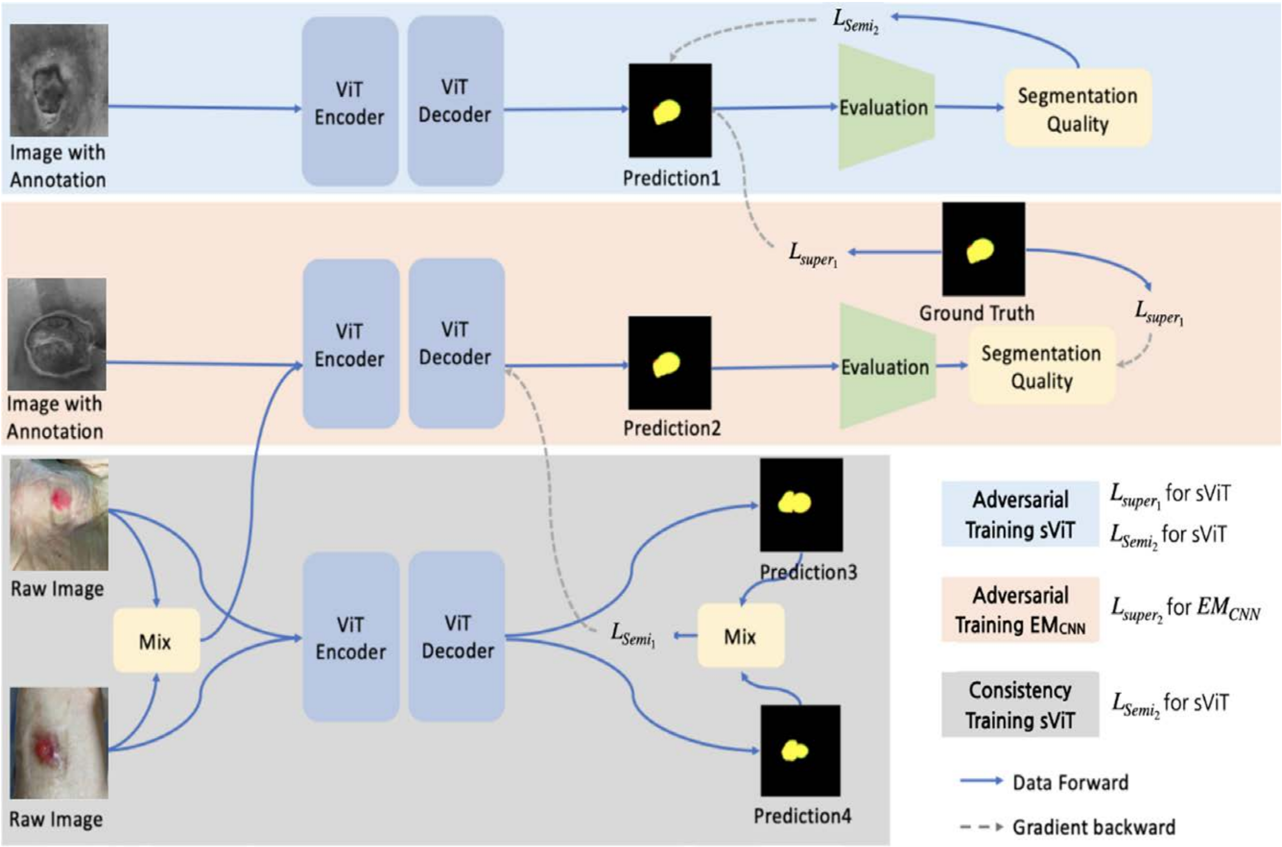

Figure 2. A MixUp-based interpolation consistency and adversarial vision transformer used to construct a semisupervised machine learning model.

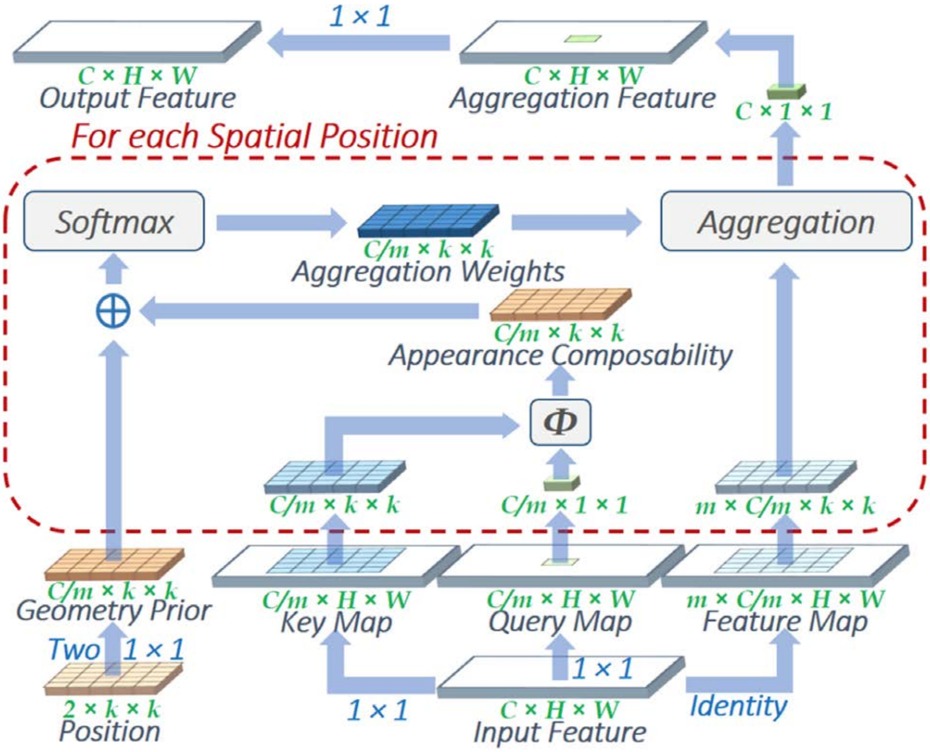

Figure 3. Diagram showing local relation layers.

The proposed PUC-ViT model takes input images of size 224 × 224 × 3 and first preprocesses the data to generate the original images of size 112 × 112 × 3. The generation of original images is aimed at minimizing information loss when resizing images because resizing may lead to information loss. In addition, the patch size used was 16 × 16 pixels, which is consistent with the size used in sViT B-16, which also aims to minimize information loss.

When the original image was input, a preprocessed bed sore image was passed through the sViT model. The ViT feature extractor then divides the image into predefined patch units and converts each patch into a 1-dimensional embedding dimension. Next, these embeddings were augmented with class tokens and positional embeddings and then passed to the encoder of the final sViT transformer. The features of the input image were determined through the sViT encoder layers, and the output features from the encoder were concatenated with the color and depth of the preprocessed image and its corresponding features. Subsequently, these concatenated features are input into the final layer, the classifier, in which each model classifier classifies the stages of the PU Given Mixup operation[16] that the above process is known as the PUC-ViT model.

Equation 1 illustrates pixel-wise interpolation, in which the sViT and/or full ViT (fViT) models are compelled to generate consistent predictions for the interpolated points of the unlabeled images. The objective of the consistency training stage is to ensure that the reliable and consistent segmentation of image points can be interpolated from existing points. This relationship can be summarized using Equation 2.

$$Mix_{\lambda}\left(a,b\right)=\lambda\times a+\left(1-\lambda\right)\times b \tag{1}$$

$$f_{ViT}\left(Mix_{\lambda}\left(X_{u1},X_{u2}\right);\theta\right)\approx Mix_{\lambda}\left(f_{ViT}\left(X_{u1};\theta\right),f_{ViT}\left(X_{u2};\hat{\theta}\right)\right) \tag{2}$$

The training objective is to minimize the sum of the supervision loss L_{super} and semivision loss L_{semi} of the 3 proposed training stages. Moreover, L_{super} or L_{semi} also depend on whether the data are annotated. In general, λ is known as a major ramp-up function[17] because $\lambda=e^{-5\times\left(1-t_{iteration/tmaxiteration}\right)^{2}}$, where t represents the iteration. Here, we updated λ every 150 iterations, which allowed us to make the training process shift the focus from L_{super} to the initialization of annotations with L_{semi} to better learn from the features of the raw images. Next, L_{super1} was used to train sViT with labeled data $(X_{l},Y_{gt})\in L$ using 1 of the adversarial training stages (Equation 3). By applying CE for the evaluation score of inference, that is, $f_{ViT}\left(X_{u}\right)$ from EM f_{CNN}, the sViT produces reasonable segmentation maps supervised by experimental minimal (EM).

$$L_{sup1}=CE\left(Y_{gt},f_{ViT}\left(X_{l};\theta\right)\right)+Dice\left(Y_{gt},f_{ViT}\left(X_{l};\theta\right)\right)$$

$$L_{sup2}=BCE\left(f_{CNN}\left(f_{ViT}\left(X_{l}/X_{u};\theta_{ViT}\right);\theta_{CNN}\right),1/0\right)$$

$$\begin{aligned}L_{semi1}=CE\,(&f_{ViT}\left(Mix_{\lambda}\left(X_{u1},X_{u2}\right);\theta\right),\\&f_{ViT}\left(X_{l};\theta\right),Mix_{\lambda}\left(f_{ViT}\left(X_{u1};\theta\right),f_{ViT}\left(X_{u2};\theta\right)\right))\end{aligned}$$

$$L_{semi2}=-CE\left(f_{CNN}\left(f_{ViT}X_{u};\theta_{ViT}\right),X_{u};\theta_{CNN}\right),0 \tag{3}$$

2.3.3. Data sets and preprocess of images. Subsequently, the PU images were transformed into image data suitable for training the EM_{CNN} model. No relabeling or exclusion was performed. To determine the relevant optimal values, the pretrained sViT model, which accepts images of size 224 × 224 × 3, was used as the base model. Finally, to minimize data loss during preprocessing, spectrogram images of size 112 × 112 × 3 were generated using Mixup (Equation 3).

2.3.4. Model architecture. Next, the ViT feature extractor divides the image into predetermined patch units and converts each patch into 1-dimensional embedding dimensions. These embedding patches were augmented with class tokens and positional embeddings before being passed to the encoder. For image classification, the ViT encoder processes the input and extracts features. These extracted features are then learned via the encoder layers before being concatenated through the encoder. The concatenated features are then input into the classifier during the final stage of the model, where the PU stages are classified. In this study, we used a transformer model to appropriately classify the stages of the bed sore images. To do so, the input image was first divided into multiple patches and then fed into a CNN (ResNet) to extract the feature maps. These feature maps were then flattened and input into the transformer encoder, followed by the attachment of a classifier for training. Because of the limited dataset included, this study also aims to improve model performance using transformer encoder blocks based on the swine transformer. Thus, during analysis, all images were initially split into patches, which were then sequentially processed through Swin Transformer blocks. During this process, Patch Merging was employed to increase the patch size. For example, if an image of size 32 × 32 is divided into 4 × 4 patches, thereby resulting in 64 patches, merging them into 16 patches would require merging them into 8 × 8 patches. Figure 3 illustrates the transformer encoder block proposed in this study.

2.4. Ethical considerations

This study was conducted in accordance with the guidelines of the declaration of Helsinki and all procedures were approved by the Institutional Review Board of Daejeon University (IRB No. 1040647-202310-HR-006-01). Written informed consent was waived because the retrospective datasets used were anonymous.

3. Results

3.1. PU classification

In this study, a classifier using eval linear and self-attention as inputs was employed for PU classification, following both multiclassification and multilabel classification training. The encoder and multilayer perception (MLP) header used for the PU image classification are defined in Equation 4.

$$z_{j}{}'\ =MSA\left(LA\left(z_{l-1}\right)\right)+z_{l-1},l=1,2,3...L \tag{4}$$

$$z_{j}{}'=MLP\left(LN\left(z_{1}'\right)\right)+z_{1}'$$

In Equation 4, z_{l}' is used for multihead attention and z_{l} is used for MLP. For the MLP head, $\mu=LN\left(z_{L}^{0}\right)$ was used to obtain the value of y. Inattention, query, key, and value were required parameters for each head, and embedded tensors were implemented by rearrangement to divide the embedding dimension for each head. In addition, z_{L}^{0} was found to represent the final output of the last step, during which the 0th vector signifies the final output. Furthermore, when the MLP head is used, the predicted value of the final stage of the PU image is the model output.

In this study, we used feature maps obtained through a CNN as input sequences, instead of using raw image patches. Subsequently, the patch size (P) is set to 1 because this eliminates the need to crop feature maps at the patch level; instead, they are directly flattened and projected into the dimensions (D) of the transformer. Next, for fine-tuning and achieving higher resolution, the value of N changes according to the resolution of the input images, whereas P remains fixed for experimentation. Fine-tuning involves removing the pretrained prediction head and initializing it with zeros as $D\times K$. Next, to extract lesions from the input images and make predictions, we used embeddings containing the positional information of the patches. Thus, the input images were sliced into patches of previous size using embeddings with positional information. These patches were then treated as sequences and fed into the transformer encoder. To ensure that positional information was not lost, position embeddings were added to the classification token and patch embedding pairs as inputs to the transformer encoder for training.

Next, we verified the theoretical foundations of our experiments. To do so, we sampled 1 epoch's worth of Mixup points – to simulate training – from a downsampled version of each training dataset. We then computed the minimum distance between each Mixup point and the points from classes other than the 2 mixed classes. Subsequently, we computed the distances for both the training and test datasets to determine whether good training but poor test performance could be attributed to test data conflicting with mixed training points.

Table 1 shows the results of experiments with 3 data augmentation methods and shows a significant improvement relative to the naive ViT model. Although Mixup showed an improvement over the naive model, PUC-ViT demonstrated the best performance. In this paper, we also report that further performance enhancement was achieved by removing some instances from the sViT. In addition, by mixing various stages, the model could extract instance features. Performing instance-level grouping using most instances in stage 3 as the backbone yielded good results. After considering all these comparisons, we chose the ViTMixup and sViT as the PUC-ViT models to be used in the future.

Table 2 displays the intersection over union (IoU) results for predicting PU areas in images across different stages for each ViT. The experimental findings indicated that when experimenting at 90% IoU, the average discovery of each stage was higher than when the IoU was 100%.

The initial PU dataset was augmented via data augmentation and transform learning was then used to avoid overfitting. Because the training dataset was small, we kept the feature extraction in the pretrained model, removed the existing classifier, and fine-tuned the fully connected (FC) layers of the final classifier. We trained models by cropping the training data to 112 × 112 × 3 and adding 1 × 5 FC nodes to each model to match the input and output numbers of the final FC layer nodes. To train the models, we used Adam optimization with an initial learning rate of 0.0, set the initial learning rate to 0.001, and trained for 100 iterations. To prevent overfitting as training progressed, we applied learning rate reduction and early termination algorithms. For the learning rate reduction algorithm, the reduction rate was set to 1%. The early termination algorithm tracks the validation loss of the loss function and terminates learning if there is no reduction for 20 epochs.

Table 3 shows the similarity between enhanced images. First, we used the area under the curve classification to evaluate the overall performance of the model; this was highest for stage 4 (88.27 ± 1.06), followed by stage 5 (87.09 ± 2.47). In addition, stages 1, 2, 3, and 5 were all found to show significant differences at $P < .001$, stage 4 was significant at $P < .05$, thereby confirming the reliability of the augmented data.

3.2. Model test

Next, to train the deep learning model, we used the Human and Machine (HAM10000) dataset provided by the International Skin Imaging Collaboration for pretraining. HAM10000 consists of 10,015 images of 7 skin diseases and has been reported to outperform classification datasets such as ResNet50. For the deep learning model trained with HAM10000, the best learning rate corresponds to the highest accuracy achieved by a baseline CNN model pretrained with PU datasets. However, a learning rate with higher sensitivity is selected if the accuracy is equal. In this study, the performance metrics of each deep learning model (i.e., ViT, ViTMixup, sViT, and the proposed model PUC-ViT) have been compared in Table 4, and their performance metrics have been cross-validated in Table 5. Cross-validation of each model with the baseline CNN model showed that sensitivity was higher for both ViTMixup and PUC-ViT, and accuracy increased for ViT, sViT, and PUC-ViT. Of all models, we observed the largest effect for PUC-ViT. Finally, we also found that F1 scores increased for ViTMixup, sViT, and PUC-ViT, with ViT showing the highest scores (Table 5).

In this study, we applied data augmentation to the PU image training data – a dataset that is small in size and shows unbalanced characteristics between classes – to increase dataset size and balance data between classes. Next, we pretrained using the HAM10000 dataset and a CNN model and tested the classification performance of the ViT, ViTMixup, and sViT models using the augmented data. Our results showed the following accuracy and sensitivity values for the respective models: ViT (87.9%, 64.6%), ViTMixup (82.8%, 75.9%), sViT (93.9%, 73.6%), and PUC-ViT (95.6%, 76.5%). Thus, our data show that the PUC-ViT model has the highest performance (Table 5).

4. Discussion

Using deep learning, we experimented with data augmentation using the CNN, ViT, and sViT models to determine their impact on model accuracy. Our aim was to investigate the possibility of accurately identifying lesions and extracting detailed features of lesions via data augmentation. Our results showed that by blending PU images of different stages, the resulting model can effectively extract lesion features.

The PUC-ViT model proposed in this paper consists of sampling blending points at each epoch, blending them with the adjacent classes, and performing continuous blending through whole-image augmentation to minimize the conflict between training points and test data. Using the trained model to classify unlabeled images via deep learning, we found that the prediction

Table 1

Direct comparison of semisupervised frameworks for a pressure ulcer test set.

Comparison	Original	Augmentation	Original and augmentation
ViT naive	11.433	12.936	17.716
ViTMixup	11.897	15.076	22.133
Semi ViT	14.312	18.701	24.175

ViT = vision transformer.

Table 2

Mean IoU results for a test set under different assumed ratios of label/total data during model training.

Category	10%	30%	50%	70%	90%	100%
Stage 1 + ViT	0.735	0.833	0.856	0.856	0.859	0.865
Stage 2 + ViT	0.716	0.814	0.845	0.843	0.839	0.854
Stage 3 + ViT	0.774	0.831	0.855	0.858	0.893	0.861
Stage 4 + ViT	0.757	0.829	0.854	0.853	0.881	0.858
Stage 5 + ViT	0.784	0.830	0.860	0.856	0.859	0.863
PUC-ViT Average	0.7532	0.8274	0.854	0.8532	0.8662	0.8602

IoU = intersection over union, PUC-ViT = pressure ulcer cluster vision transformer, ViT = vision transformer.

Table 3

Comparison of image enhancement performance at each stage.

Stage	Classification AUC	Segmentation task (dice)	*P*-value
Stage 1	86.23 ± 2.31	85.60 ± 0.04	<.001
Stage 2	83.48 ± 2.37	81.25 ± 0.11	<.001
Stage 3	85.97 ± 1.77	82.09 ± 0.07	<.001
Stage 4	88.27 ± 1.06	82.61 ± 0.06	<.05
Stage 5	87.09 ± 2.47	84.04 ± 0.17	<.001

AUC = area under the ROC curve.

accuracy was low at 88.3% before adjusting the hyperparameters. However, after making some changes, such as using ViT to fine-tune the hyperparameters, segmenting the training data, performing data augmentation, and validating the effectiveness of the pretrained weights, the performance improved significantly on the validation data, reaching a prediction accuracy of 95.6%. We also found that the PUC-ViT model had an average ROC curve of 0.937 when predicting PU classes (Fig. 4). However, the accuracy was slightly lower in stage 2; we interpreted that this finding was due to the lower boundary of stage 1.

Deep learning has many possible applications for addressing PUs. Previous studies have reported high prediction rates obtained via deep learning analyses of various risk factors related to PUs that have been used to predict the incidence of PUs in hospitalized patients.[18–20] Similar to our present study, a previous study used a CNN model to extract features from images of PUs, assessing the wounds based on infection status and the amount of necrotic tissue. This data was then used to propose a management method for PUs.[21] Another study attempted to improve the accuracy of PU classification by comparing CNN and you only look once v5 (YOLOv5) models,[22] and there also exists a systematic literature review that analyzed 90 studies that employed machine learning models to guide caring for patients with PU.[7]

A meta-analysis of 14 studies found that machine learning models show outstanding performance when predicting pressure injury, although further studies are still needed to validate this finding and confirm the clinical value of these models with respect to pressure injury development and management.[23] At present, researchers are comparing various deep learning models to increase the accuracy of pressure ulcer identification and classification. The results of such studies are models that can be used in hospitals to improve diagnostic accuracy and increase the prediction rate by collaborating with medical professionals when assessing the risk of PU. However, our study focused on enabling community caregivers to detect PUs more quickly in-home care patients and facilitating early hospitals referrals. In-home caregivers are sometimes nonmedical personnel, and their ability to perform manual evaluation using PU assessment tools may be limited, so it is difficult to expect early PU detection.

Table 4

Model comparison using a pretrained CNN.

Model	Accuracy	Precision	Recall	F1 score
CNN (baseline)	91.34	90.43	77.98	87.33
ViT	91.99	90.19	86.24	87.40
ViTMixup	92.58	94.00	86.24	92.03
SemiViT	93.99	90.35	94.50	93.71
PUC-ViT	97.76	96.02	90.83	95.46

CNN = convolutional neural network, PUC-ViT = pressure ulcer cluster vision transformer, ViT = vision transformer.

Table 5

Cross-validation results for each model.

Model	Accuracy	Sensitivity	Precision	F1-score
CNN (baseline)	0.883 ± 0.009	0.696 ± 0.024	0.831 ± 0.008	0.726 ± 0.024
ViT	0.879 ± 0.008*	0.646 ± 0.025*	0.931 ± 0.013*	0.772 ± 0.016*
ViTMixup	0.828 ± 0.012	0.759 ± 0.026*	0.940 ± 0.021*	0.763 ± 0.018
SemiViT	0.939 ± 0.013	0.736 ± 0.030	0.905 ± 0.013	0.764 ± 0.026*
PUC-ViT	0.956 ± 0.015	0.765 ± 0.021*	0.982 ± 0.032*	0.770 ± 0.019*

CNN = convolutional neural network, PUC-ViT = pressure ulcer cluster vision transformer, ViT = vision transformer.
*P <.05.

The protocol proposed here uses vision transformer (ViT)-based image classification to support prescreening for medical opinions using smart devices that are widely available in-home hospices or accompanying home care services. This is important because PUs are medical conditions that require immediate attention but can be difficult for patients to recognize on their own. Accordingly, the difficulty in accurately diagnosing PUs can lead to challenges in wound treatment as the condition progresses.

Therefore, by leveraging advanced AI technology, this paper proposes a protocol involving the preprocessing of various images to prepare a dataset for training a PU classification model. The application proposed in this study is a smart device app designed to address hardware limitations while increasing classification accuracy by adjusting the hyperparameters of the dataset used for training the model.[24] The approach outlined here aims to make this application available for general use in the future, enabling individuals to receive timely medical opinions from experts and ensuring appropriate treatment based on medical recommendations.

However, there are some limitations to this study. First, the photos used in the analysis were collected by different people using their own smartphone cameras (i.e., using different camera models, albeit at the same resolution). However, to account for cases in which images with different resolutions are used as input, data augmentation was performed by adjusting the resolution. Next, the similarity of the generated image to the original image was compared, and the proposed model showed a similarity of 81.25 to 85.60. Although the results reported here are promising, the proposed method has not been fully validated for different smartphone cameras. Second, we base our PU stage determination model on 2D photos that do not include PU depth information. However, we expect that the accuracy of this model can be further improved if the PU stage determination is performed using 3D PU images containing depth information. Finally, in this study, we divided PU into 4 stages. In fact, for community-based users, making accurate decisions for PUs at stages 1, 2, and 3 is more important compared to making decisions involving other stages. Therefore, to increase the accuracy of stage

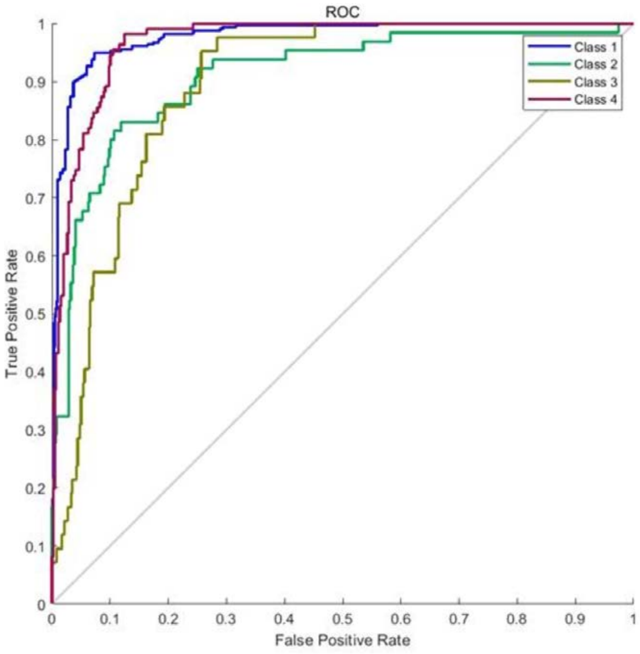

Figure 4. ROC curve for classification of pressure ulcer lesions.

determination depending on the user, it is necessary to simplify or reclassify PU stages.

In the future, the PU staging tool developed here should be implemented as an app to address the needs of community care clients and confirm its usefulness. Using this PU staging tool, we hope to further improve the quality of community-based care.

Author contributions

Conceptualization: Hana Yoo.
Data curation: Young-Bok Cho.
Formal analysis: Young-Bok Cho.
Funding acquisition: Hana Yoo.
Project administration: Hana Yoo.
Resources: Hana Yoo.
Software: Young-Bok Cho.
Supervision: Hana Yoo.
Visualization: Young-Bok Cho.
Writing – original draft: Hana Yoo.

References

[1] European Pressure Ulcer Advisory Panel, National Pressure Injury Advisory Panel, and Pan Pacific Pressure Injury Alliance. Prevention and treatment of pressure ulcers/injuries: clinical practice guideline: the international Guideline 2019. EPUAP, NPIAP, PPPIA, 2019. https://www.biosanas.com.br/uploads/outros/artigos_cientificos/146/3cf6b27eb06aa7587bd832e6f0306955.pdf. Accessed December 1, 2022.

[2] Capon A, Pavoni N, Mastromattei A, Di Lallo D. Pressure ulcer risk in long-term units: prevalence and associated factors. J Adv Nurs. 2007;58:263–72.

[3] Gallagher P, Barry P, Hartigan I, McCluskey P, O'Connor K, O'Connor M. Prevalence of pressure ulcers in three university teaching hospitals in Ireland. J Tissue Viability. 2008;17:103–9.

[4] Fulbrook P, Miles S, Coyer F. Prevalence of pressure injury in adults presenting to the emergency department by ambulance. Aust Crit Care. 2019;32:509–14.

[5] Al Mutairi KB, Hendrie D. Global incidence and prevalence of pressure injuries in public hospitals: a systematic review. Wound Med. 2018;22:23–31.

[6] Paolini G, Sorotos M, Firmani G, Gravili G, Ceci D, Santanelli di Pompeo F. Santanelli di Pompeo F. Low-vacuum negative pressure wound therapy protocol for complex wounds with exposed vessels. J Wound Care. 2022;31:78–85.

[7] Dweekat OY, Lam SS, McGrath L. Machine learning techniques, applications, and potential future opportunities in pressure injuries (bedsores) management: a systematic review. Int J Environ Res Public Health. 2023;20:796.

[8] Kim G, Park M, Kim K. The effect of pressure injury training for nurses: a systematic review and meta-analysis. Adv Skin Wound Care. 2020;33:1–11.

[9] Qu C, Luo W, Zeng Z, et al. The predictive effect of different machine learning algorithms for pressure injuries in hospitalized patients: a network meta-analyses. Heliyon. 2022;8:e11361.

[10] Cho YB. XAI personalized recommendation algorithm using ViT and K-means. J Electr Eng Technol. 2024;19:4495–503.

[11] Zhao Z, Zeng Z, Xu K, Chen C, Guan C. Deeply supervised active learning from strong and weak labelers for biomedical image segmentation. IEEE J Biomed Health Inform. 2021;25:3744–51.

[12] Basak H, Yin Z. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023:19786–19797.

[13] Xu Z, Wang Y, Lu D, et al. All-around real label supervision: cyclic prototype consistency learning for semi-supervised medical image segmentation. IEEE J Biomed Health Inform. 2022;26:3174–84.

[14] Kalina B, Cho Y. Improving adversarial domain adaptation with mixup regularization. J Inform Commun Convergence Engineering. 2023;21:139–44.

[15] Bui PN, Le DT, Bum J, Kim S, Song SJ, Choo H. Semi-supervised learning with fact-forcing for medical image segmentation. IEEE Access. 2023;11:99413–25.

[16] Chidambaram M, Wang X, Hu Y, Wu C, Ge R. Towards understanding the data dependency of mixup-style training. *arXiv preprint arXiv:2110.07647*. 2021.

[17] Wang Z, Zhang H, Liu Y.. Weakly-supervised self-ensembling vision transformer for MRI cardiac segmentation. In: 2023 IEEE Conference on Artificial Intelligence (CAI). IEEE. pp. 2023;101–2.

[18] Walther F, Heinrich L, Schmitt J, Eberlein-Gonska M, Roessler M. Prediction of inpatient pressure ulcers based on routine healthcare data using machine learning methodology. Sci Rep. 2022; 12:5044.

[19] Šín P, Hokynková A, Marie N, Andrea P, Krč R, Podroužek J. Machine learning-based pressure ulcer prediction in modular critical care data. Diagnostics (Basel). 2022;12:850.

[20] Kim M, Kim TH, Kim D, et al. In-advance prediction of pressure ulcers via deep-learning-based robust missing value imputation on real-time intensive care variables. J Clin Med. 2023;13:36.

[21] Liu TJ, Christian M, Chu YC, et al. A pressure ulcers assessment system for diagnosis and decision making using convolutional neural networks. J Formos Med Assoc. 2022;121:2227–36.

[22] Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M. YOLO-based deep learning model for pressure ulcer detection and classification. Healthcare (Basel). 2023;11:1222.

[23] Pei J, Guo X, Tao H, et al. Machine learning-based prediction models for pressure injury: a systematic review and meta-analysis. Int Wound J. 2023;20:4328–39.

[24] Seo JB, Lee JS, Yoo H, Cho YB. Deep learning-based pressure ulcer image object detection study. In: 2022 Korean Society of Computer Information Conference. KCSI. pp. 2022;311–2.