

Original Paper

# Convolutional Neural Network Models for Visual Classification of Pressure Ulcer Stages: Cross-Sectional Study

Changbin Lei<sup>1</sup>, MSN; Yan Jiang<sup>2</sup>, PhD; Ke Xu<sup>3</sup>, PhD; Shanshan Liu<sup>3</sup>, MSN; Hua Cao<sup>4</sup>, BSN; Cong Wang<sup>3</sup>, MSN

<sup>1</sup>Trauma Center, West China Hospital, West China School of Nursing, Sichuan University, Chengdu, China

<sup>2</sup>Nursing Department, Evidence-Based Nursing Center, West China Hospital, Sichuan University, Chengdu, China

<sup>3</sup>Evidence-Based Nursing Center, West China Hospital, Sichuan University, Chengdu, China

<sup>4</sup>Neurosurgery Department, West China Hospital, Sichuan University, Chengdu, China

## Corresponding Author:

Yan Jiang, PhD

Nursing Department, Evidence-Based Nursing Center

West China Hospital, Sichuan University

37# Guoxue Alley

Chengdu, 610041

China

Phone: 86 18980601511

Email: [hxhljy2018@163.com](mailto:hxhljy2018@163.com)

## Abstract

**Background:** Pressure injuries (PIs) pose a negative health impact and a substantial economic burden on patients and society. Accurate staging is crucial for treating PIs. Owing to the diversity in the clinical manifestations of PIs and the lack of objective biochemical and pathological examinations, accurate staging of PIs is a major challenge. The deep learning algorithm, which uses convolutional neural networks (CNNs), has demonstrated exceptional classification performance in the intricate domain of skin diseases and wounds and has the potential to improve the staging accuracy of PIs.

**Objective:** We explored the potential of applying AlexNet, VGGNet16, ResNet18, and DenseNet121 to PI staging, aiming to provide an effective tool to assist in staging.

**Methods:** PI images from patients—including those with stage I, stage II, stage III, stage IV, unstageable, and suspected deep tissue injury (SDTI)—were collected at a tertiary hospital in China. Additionally, we augmented the PI data by cropping and flipping the PI images 9 times. The collected images were then divided into training, validation, and test sets at a ratio of 8:1:1. We subsequently trained them via AlexNet, VGGNet16, ResNet18, and DenseNet121 to develop staging models.

**Results:** We collected 853 raw PI images with the following distributions across stages: stage I (n=148), stage II (n=121), stage III (n=216), stage IV (n=110), unstageable (n=128), and SDTI (n=130). A total of 7677 images were obtained after data augmentation. Among all the CNN models, DenseNet121 demonstrated the highest overall accuracy of 93.71%. The classification performances of AlexNet, VGGNet16, and ResNet18 exhibited overall accuracies of 87.74%, 82.42%, and 92.42%, respectively.

**Conclusions:** The CNN-based models demonstrated strong classification ability for PI images, which might promote highly efficient, intelligent PI staging methods. In the future, the models can be compared with nurses with different levels of experience to further verify the clinical application effect.

*JMIR Med Inform* 2025;13:e62774; doi: [10.2196/62774](https://doi.org/10.2196/62774)

**Keywords:** pressure ulcer; deep learning; artificial intelligence; neural network; CNN; machine learning; image; imaging; classification; ulcer; sore; pressure; wound; skin

## Introduction

Pressure injuries (PIs), also known as pressure ulcers, are prevalent in health care settings and result from sustained pressure or shear forces on bony prominences and soft tissues [1]. A meta-analysis encompassing 2,579,049 hospitalized adult patients revealed a 12.8% prevalence of PIs, and the estimated annual total cost of managing PIs increased approximately several times even in high-income countries [2,3]. Accurate staging of PIs is essential for determining appropriate treatment protocols and predicting clinical outcomes. Traditional staging relies on subjective evaluation by health care providers, often wound nurses, who use guidelines such as the *Prevention and Treatment of Pressure Ulcers/Injuries: Clinical Practice Guideline, International Guideline 2019*, and divide the PI stage into 6 stages—namely, stages I, II, III, and IV, unstageable, and suspected deep tissue injury (SDTI)—based on visual assessment, such as the color and texture characteristics of the wound bed, wound edges, and skin around the wound [4].

However, visual assessment conducted by medical staff is subject to considerable interrater and intrarater variability [5]. Moreover, the staging results are significantly influenced by the level of wound knowledge and the clinical experience of the evaluators. It has been reported that only 23%-58% of medical staff correctly classify PIs [6]. Therefore, accurate staging of PIs is still a great challenge, and a universal, reliable, and more objective staging system is urgently needed.

Convolutional neural networks (CNNs) offer an opportunity to increase the objectivity and accuracy of PI staging. CNNs, a subset of deep learning (DL) models, have shown remarkable potential in medical image analysis [7]. CNNs automatically extract and learn hierarchical features from grid-like data, such as images, and have achieved performance levels comparable to or surpassing those of human experts in various medical imaging domains [8]. In dermatological and wound imaging, CNNs have demonstrated promising results, matching or even exceeding the diagnostic accuracy of dermatologists in classifying skin cancer and other skin lesions [9]. On the basis of the good classification performance of CNNs in medical images and the increased clinical need for automated staging of PIs, in recent years, the use of CNNs to learn PI images has gradually developed, but there are also areas for improvement. For example, the use of a singular model had an accuracy rate below 90%, alongside images sourced from outdated databases and of poor quality, and the labeling of PI images often relies on the subjective judgment of clinical experts, so the quality and consistency of the images were uneven in these studies,

which limit the evidence supporting the model's widespread applicability [10-12]. Therefore, further exploration is needed in this field to prove the effectiveness of CNNs.

In this study, we trained AlexNet, VGGNet16, ResNet18, and DenseNet121 to classify PI images with the aim of contributing to artificial intelligence-driven wound care knowledge and informing the development of advanced clinical tools.

## Methods

### Setting and Samples

Our study was conducted at a tertiary hospital in Chengdu, China, from March 1, 2022, to September 30, 2024.

The inclusion criteria for patients were as follows: (1) PIs were diagnosed by 2 wound therapists simultaneously on the basis of the *Prevention and Treatment of Pressure Ulcers/Injuries: Clinical Practice Guideline, International Guideline 2019* [4] and (2) the patients provided informed consent.

The exclusion criteria for patients were as follows: (1) PIs included other skin diseases, such as incontinence dermatitis; (2) PIs were covered by dressings or tattoos; (3) exudation was excessive, resulting in obvious reflection; (4) mucosal PIs; (5) the patient's vital signs remained too unstable to exchange positions; and (6) the patients were in isolation wards.

The inclusion criteria for images were as follows: (1) images were taken by the investigator, and (2) images fully presented the wound triangle, namely, the wound bed, wound edge, and surrounding wound tissue.

The exclusion criteria for images were as follows: (1) images containing nonwounded and skin tissues such as clothing and sheets and (2) images that were blurred and overexposed.

### Dataset

We recruited wound therapists who were awarded an international ostomy wound certificate to act as PI staging evaluators. Five wound therapists were included in this study. The basic information is provided in Table 1. To test the consistency of the 5 wound therapists' assessments, we sent them 30 PI images from the NPIAP (National Pressure Injury Advisory Panel) website in the form of a questionnaire [13]. The Fleiss kappa coefficient was 0.856,  $P < .001$ , indicating strong staging consistency among them [14]. The staging results are shown in Tables 2 and 3.

**Table 1.** Basic information of the 5 wound therapists.

Order	Gender	Age (years)	Working years	Department
A	Female	42	20	Respiratory ICU <sup>a</sup>
B	Female	36	15	Neurology
C	Female	40	18	General ICU
D	Female	43	25	Neurology
E	Female	36	15	ICU

<sup>a</sup>ICU: intensive care unit.**Table 2.** Pressure injury staging results of the 5 wound therapists.

Order	Staging results of the 5 wound therapists				
	A	B	C	D	E
1	Stage III	Stage III	Stage III	Stage III	Stage IV
2	Stage I	Stage I	Stage I	Stage I	Stage I
3	Stage II	Stage III	Stage II	Stage II	Stage II
4	Unstageable	Unstageable	Stage IV	Unstageable	Unstageable
5	Unstageable	Unstageable	Unstageable	Unstageable	Unstageable
6	Stage IV	Stage IV	Stage IV	Stage IV	Stage IV
7	SDTI <sup>a</sup>	SDTI	SDTI	SDTI	SDTI
8	Stage I	Stage I	Stage I	Unstageable	Stage I
9	Stage IV	Stage IV	Stage IV	Stage IV	Stage IV
10	Unstageable	Stage IV	Stage IV	Stage IV	Stage IV
11	Stage III	Stage III	Stage III	Stage III	Stage III
12	Stage III	Stage III	Stage III	Stage III	Stage III
13	Stage II	Stage III	Stage II	Stage II	Stage II
14	SDTI	SDTI	SDTI	SDTI	SDTI
15	Stage IV	Stage IV	Stage IV	Stage IV	Stage IV
16	SDTI	SDTI	SDTI	SDTI	SDTI
17	Stage I	Stage I	Stage I	Stage I	Stage I
18	Stage II	Stage III	Stage II	Stage II	Stage II
19	Stage III	Stage III	Stage III	Stage III	Stage III
20	Stage IV	Stage SDTI	Stage IV	Stage IV	Stage IV
21	Stage I	Stage I	Stage I	Stage I	Stage I
22	Unstageable	Unstageable	Unstageable	Unstageable	Unstageable
23	SDTI	SDTI	SDTI	SDTI	SDTI
24	Stage II	Stage II	Stage II	Stage II	Stage II
25	Stage I	Stage I	Stage I	Stage I	Stage I
26	Stage II	Stage II	Stage II	Stage II	Stage II
27	Unstageable	Unstageable	Unstageable	Unstageable	Unstageable
28	Stage IV	Stage IV	Stage IV	Stage IV	Stage IV
29	Stage III	Stage III	Stage III	Stage III	Stage III
30	SDTI	SDTI	SDTI	I	SDTI

<sup>a</sup>SDTI: suspected deep tissue injury.**Table 3.** Coefficient of internal consistency among the 5 wound therapists. The sample data contain 30 valid participants and 5 raters.

	Asymptotic				
	Kappa	SE	Z	P value	Asymptotic 95% CI
Overall	0.856	0.026	32.948	<.001	0.805-0.906

## The PI Staging Assessment Process and Shooting Process

Before taking the shots, the researcher and the 2 wound therapists communicated with the patients and their family members to obtain consent. After that, the 2 therapists simultaneously evaluated the staging of the pressure ulcers via the *Prevention and Treatment of Pressure Ulcers/Injuries: Clinical Practice Guideline, International Guideline 2019*. If there were any objections to the results, we informed the third wound therapist on duty, and they would negotiate the staging.

The following types of shooting equipment were used: (1) Fuji XT-4 was used for shooting, with 21.6 million effective pixels, and Fuji XF 60 mm F2.4 R Macro and (2) a gray card. The specifications of the shooting mode are as follows: (1)

antishake plus automatic mode; (2) parameter setting: shutter time: 60/1 s; (3) sensitivity: 3600-6000; (4) aperture: F6-8; (5) focal length: 60 mm; and (6) white balance: automatic white balance.

When photographs were taken, the camera was positioned parallel to the wound, with a focus on both the wound itself and the surrounding skin ([Figure 1](#)). Additionally, a gray card was used to minimize any interference caused by natural light. Following each shot, the images were carefully examined. If any image was deemed unclear, the researcher retracted the image. All of the original images were saved in JPG format and given a specific name consisting of a stage followed by a corresponding number. Finally, the images were transferred onto a computer for further analysis and storage.

**Figure 1.** Shooting process.

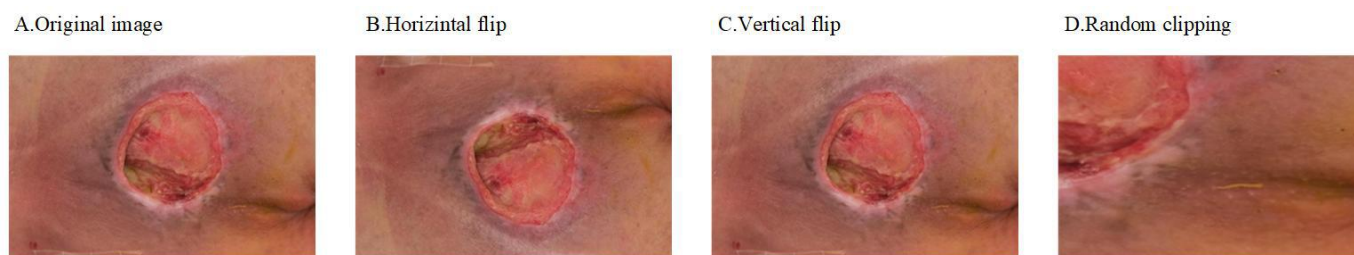


## Image Augmentation

Horizontal flipping, vertical flipping, and random clipping were used for data augmentation to expand the training dataset and improve the generalization ability of the models ([Figure 2](#)) [15]. Taking random cropping as an example,

the size of the original image was reset to 512×512×3, and after random cropping, the size was 256×256×3. The size of the augmented image was fixed again to 224×224×3 before being input into the networks so that the network model could recognize them.

**Figure 2.** Image augmentation: (A) original image, (B) horizontal flip, (C) vertical flip, and (D) random clipping.



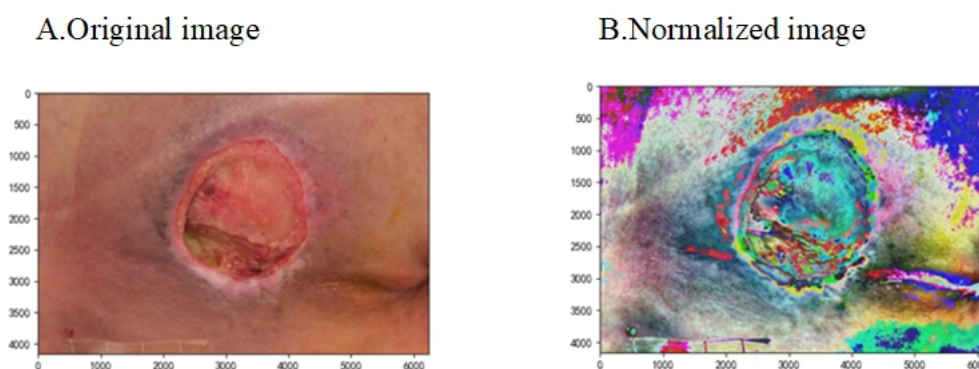
### Image Normalization

The PI images are RGB color patterns (Figure 3), and the number of pixels is [0, 256] [16]. To reduce the adverse effects caused by singular sample data and speed up model training, we limit the number of pixels to [-1, 1] in this study via 2 image normalization calculation formulas, where  $\bar{x}$  is

the mean of the number of pixels,  $N$  represents the number of pixels, and  $x$  represents the pixel value of each pixel [17]. The calculation formula is as follows:

$$(1)s = \left( \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^{\frac{1}{2}}$$

**Figure 3.** Image normalization: (A) original image and (B) normalized image.

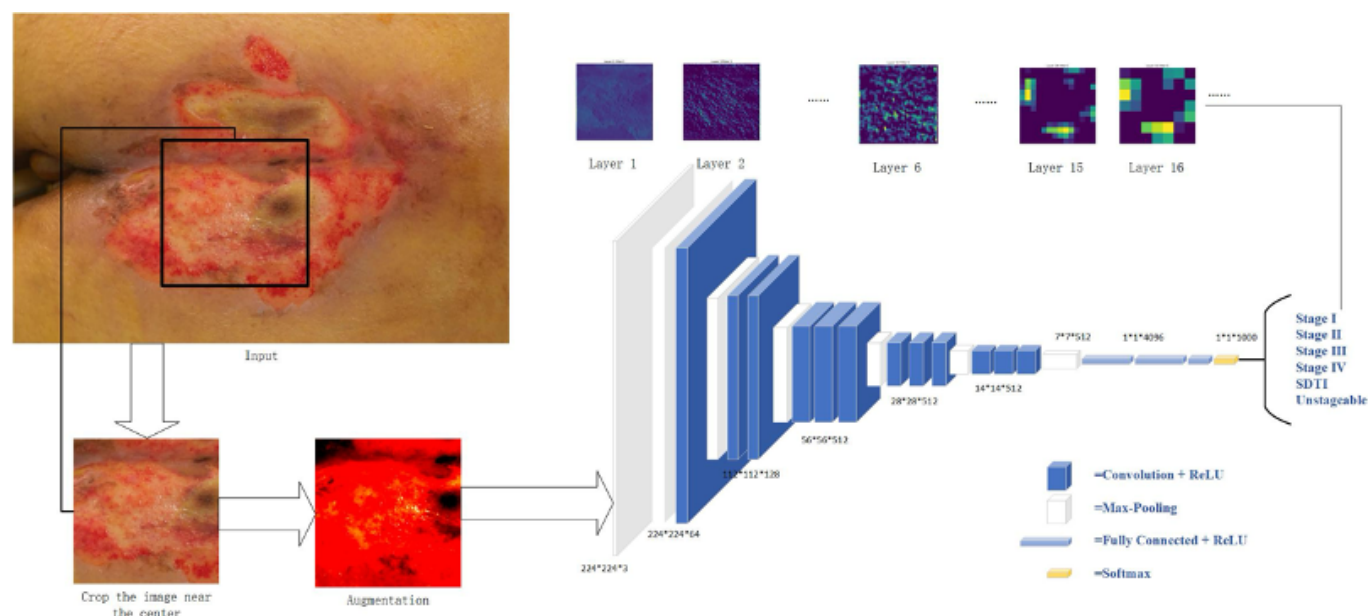


The Adam optimizer was used, and the loss function was the cross-entropy loss function (cross-entropy loss function), which is commonly used in multiple classification tasks. The data batch (batch size) was set to 8, and the initial learning rate was set to 0.0001. Every network was trained for 30 epochs.

This study was based on the 64-bit Windows 11 computer system and the Ubuntu 16.04 operating system for image training in the PyTorch 2.0.1 framework using the Python 3.9.19 language with CUDA 11.8 and NVIDIA GeForce RTX 4060 Laptop (8G).

The process of model training was divided into training, validation, and testing. The process is shown in Figure 4. First, the PI image was cut into the main wound area, and the cut image was treated as a uniform pixel, both of which were cut to 224 pixels. In the training set, the CNN training model is input; the performance of the model on the validation set is evaluated; and the state and convergence of the model are tested to adjust the hyperparameters. In the test set, the model outputs corresponding prediction results through a series of convolutions, nonlinear activations, pooling, etc, to evaluate the generalization ability of the model.



**Figure 4.** The network workflow diagrams. SDTI: suspected deep tissue injury.

In our study, we used AlexNet, VGGNet 16, ResNet 18, and DenseNet121 to train the images. AlexNet, a pioneering CNN in image classification, consists of 8 layers: 5 convolutional layers with varying filter numbers and 3 fully connected layers. It employs ReLU activations, max pooling, and dropout for regularization. The introduction of ReLU and dropout layers in AlexNet reduced training times and prevented overfitting, whereas its deep architecture allowed for the learning of complex features, enhancing classification accuracy [18].

VGGNet 16 is a 16-layer deep network consisting of 13 convolutional layers and 3 fully connected layers, all of which use 3x3 filters. It features a consistent architecture with convolutional layers followed by max pooling layers, which simplifies optimization. This network is effective for capturing grid-like patterns in images and allows for the learning of more abstract and complex features due to its depth [19].

ResNet 18 is a shallow network with 18 layers organized in a residual learning framework. It consists of 4 blocks of residual units, each with convolutional layers, batch normalization, ReLU activation, and max pooling layers. The key innovation is the introduction of residual connections, which allow the network to learn residual mappings, making it easier to train deeper networks by effectively flowing gradients through the network [20].

DenseNet121 connects each layer to every other layer in a feedforward manner, using features from all preceding layers as inputs and its own feature map for all subsequent layers. This dense connectivity reduces the number of parameters, making DenseNet121 more parameter-efficient than other CNNs of similar depth. It helps mitigate the vanishing gradient problem and enhances feature propagation, leading to improved performance and efficiency [21].

In summary, the 4 CNN models were selected for their historical performance in image classification tasks, the

uniqueness of their architectural design, and their significant contributions to the field of DL (Figure 4).

## Evaluation of Performance

We evaluated the performance from a single image result. The diagnostic performance was measured by accuracy (ACC), precision (Pre), recall (Rec), and the  $F_1$ -score. To calculate the above metrics, we defined an abnormal result as positive and a normal result as negative.

$$(1) ACC = \frac{TP+TN}{TP+FP+FN+TN}$$

$$(2) Pre = \frac{TP}{TP+FP}$$

$$(3) Rec = \frac{TP}{TP+FN}$$

$$(4) F1\text{-score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

## Ethical Considerations

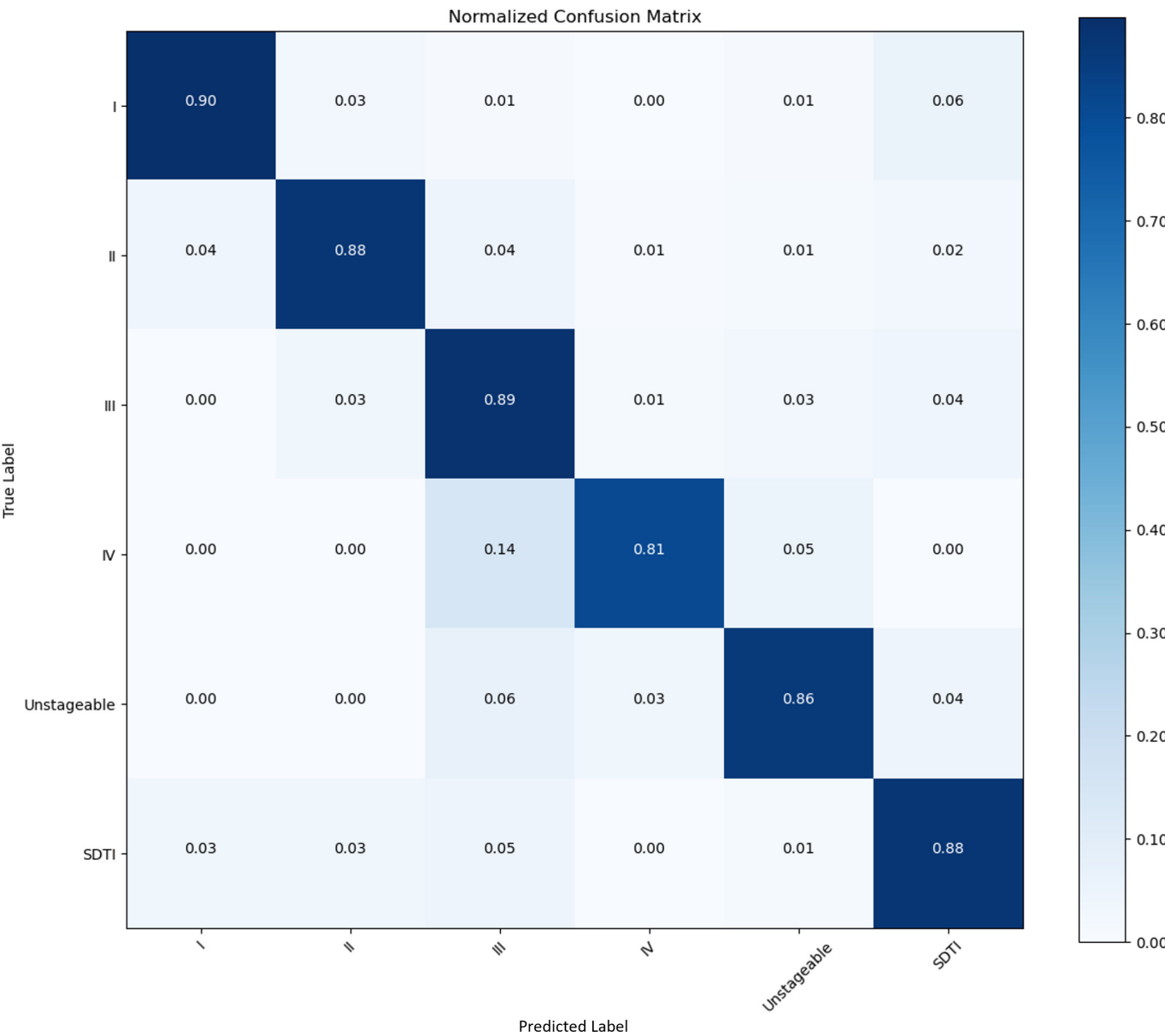
This study was approved by the Biomedical Ethics Committee of the West China Hospital of Sichuan University (#1053). In this study, the informed consent process was strictly carried out. Participation in the study was completely voluntary, and patients could refuse to participate or withdraw at any time during any phase of the study without discrimination or retaliation and without affecting their medical treatment and benefits. If participants decided to withdraw from the study, they would contact us. Patient privacy was strictly protected, and all data obtained were only used for this study. Patients did not have to pay any fees to participate in the study.

Results

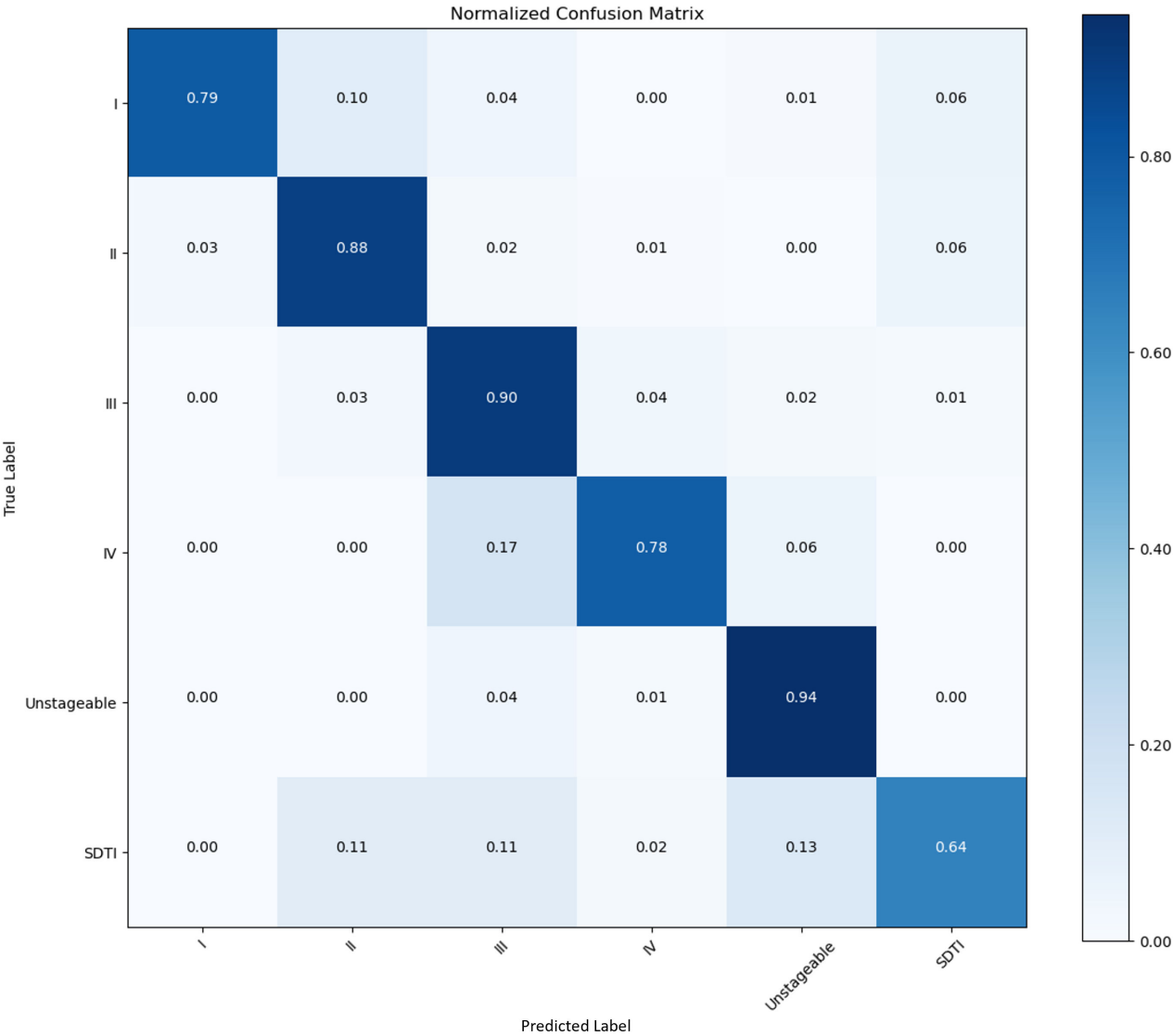
We collected 853 raw PI images in this study. We set the images into a training set, a validation set, and a test set at a ratio of 8:1:1. After augmentation, 7677 images were used. A confusion matrix is a numerical table used to display the performance results of a classification model on test data with

known target labels. It serves as a visual representation of how the model makes predictions on the test dataset. On the basis of the results obtained from each network validation, we plotted the normalized confusion matrix, with the true labels on the vertical axis and the predicted labels on the horizontal axis (Figures 5-8). Table 4 provides a summary of the accuracy, precision, recall, and  $F_1$ -scores for the 4 CNNs.

**Figure 5.** Performance of the CNNs in the identification and classification of PI images. Confusion matrix showing the accuracy and precision of 87.74% and 97.48%, respectively. CNN: convolutional neural network; PI: pressure injury; SDTI: suspected deep tissue injury.

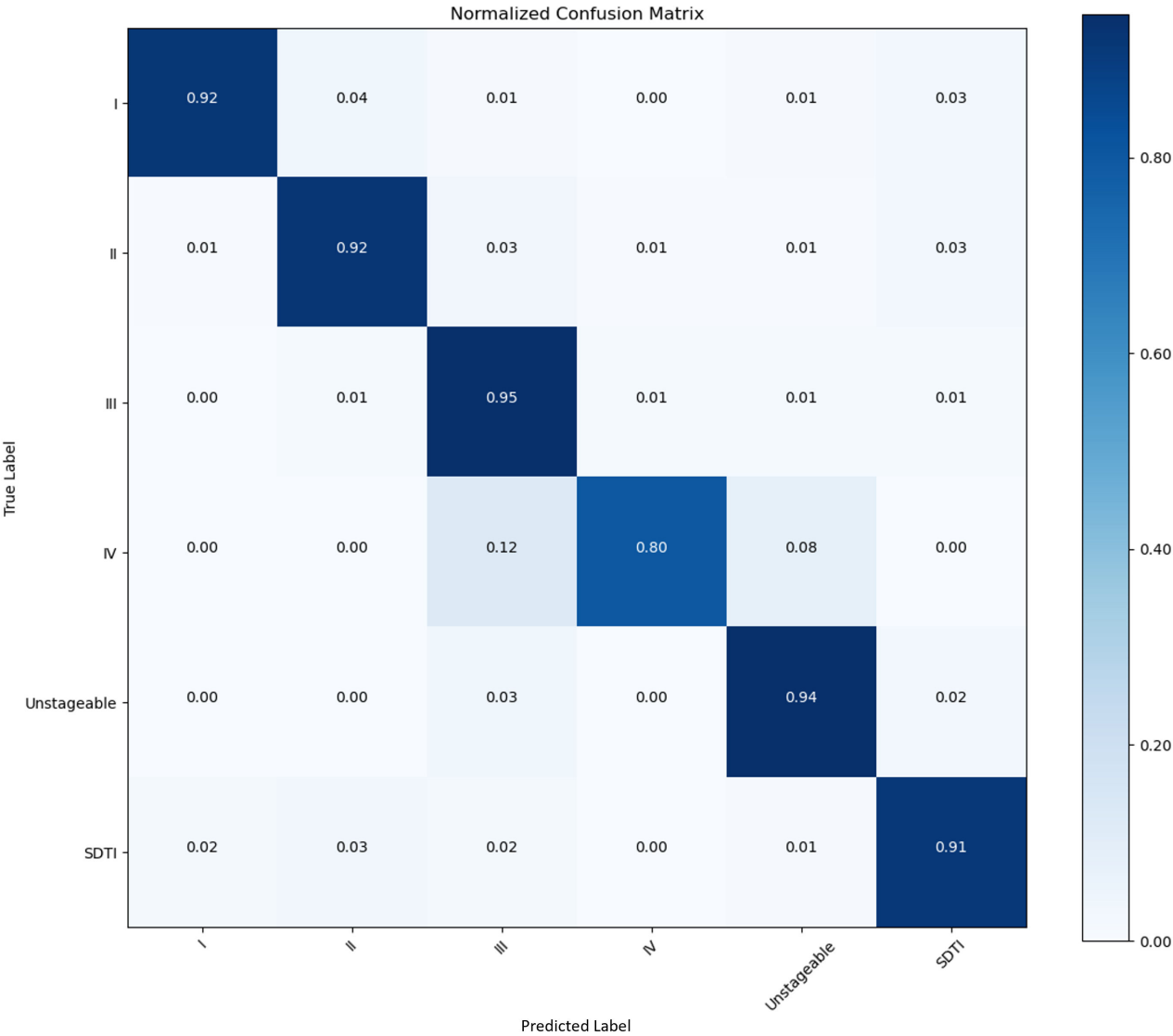


**Figure 6.** Performance of the CNNs in the identification and classification of PI images. Confusion matrix showing the accuracy and precision of 82.42% and 92.4%, respectively. CNN: convolutional neural network; PI: pressure injury; SDTI: suspected deep tissue injury.

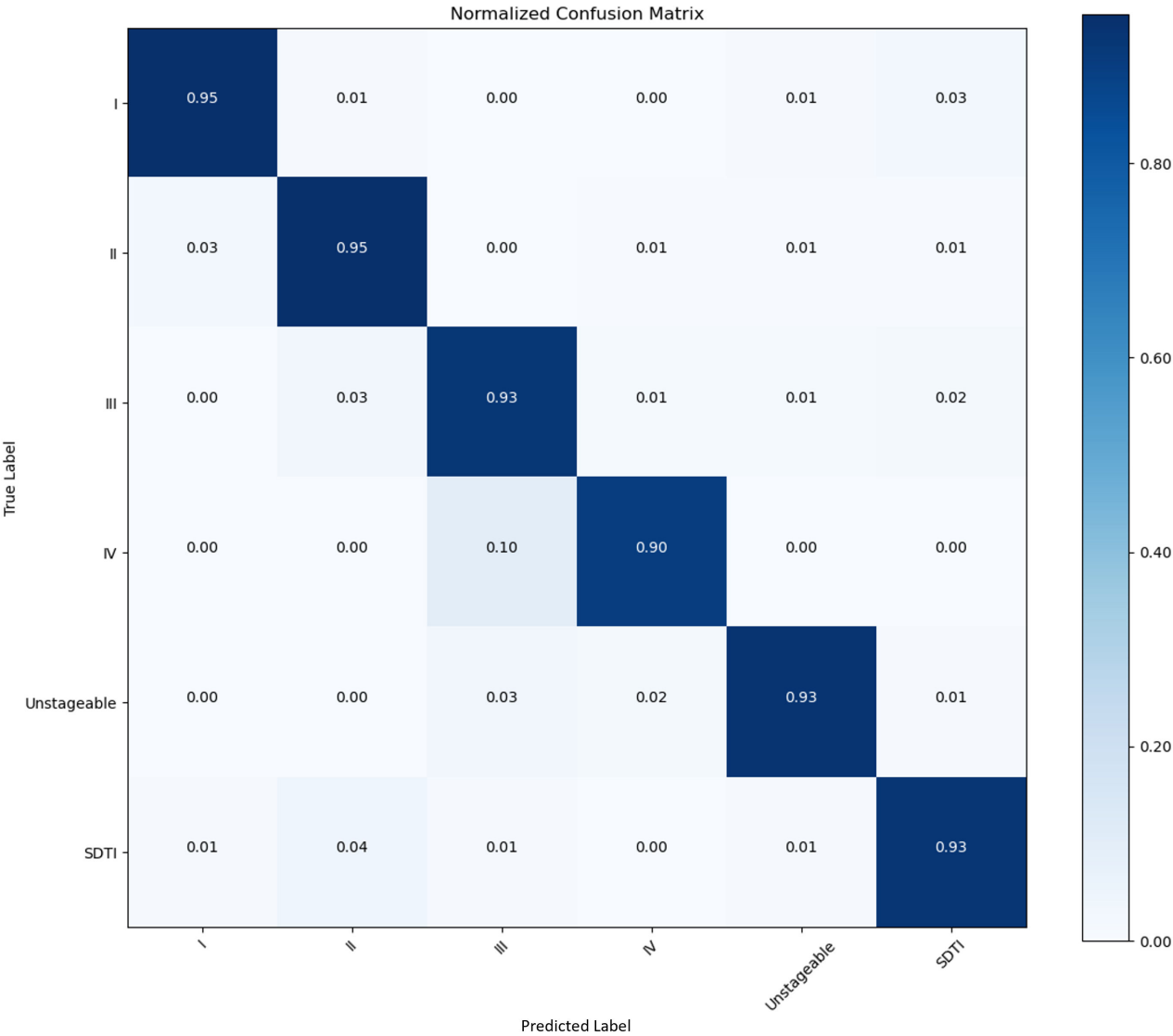




**Figure 7.** Performance of the CNNs in the identification and classification of PI images. Confusion matrix showing the accuracy and precision of 92.42% and 98.43%, respectively. CNN: convolutional neural network; PI: pressure injury; SDTI: suspected deep tissue injury.



**Figure 8.** Performance of the CNNs in the identification and classification of PI images. Confusion matrix showing the accuracy and precision of 93.71% and 98.72%, respectively. CNN: convolutional neural network; PI: pressure injury; SDTI: suspected deep tissue injury.



**Table 4.** The model’s overall classification performance.

Modes and PI <sup>a</sup> stages		Accuracy (%)	Precision (%)	Recall(%)	F <sub>1</sub> -score (%)
AlexNet	Stage I	92.76	98.25	91.43	94.71
	Stage II	86.36	97.60	92.76	95.12
	Stage III	73.91	95.53	86.36	90.71
	Stage IV	89.89	99.00	73.91	84.63
	Unstageable	79.38	98.29	89.89	93.90
	SDTI <sup>b</sup>	91.43	96.24	79.38	87.00
	Overall	87.74	97.48	85.62	94.71
VGGNet 16	Stage I	96.19	99.19	96.19	97.66
	Stage II	80.26	93.78	93.78	86.49
	Stage III	82.47	94.36	94.36	88.01
	Stage IV	76.40	98.50	98.50	75.24
	Unstageable	81.44	96.17	96.17	85.15
	SDTI	96.19	96.38	96.38	88.28

Modes and PI <sup>a</sup> stages		Accuracy (%)	Precision (%)	Recall(%)	F <sub>1</sub> -score (%)
ResNet 18	Overall	82.42	96.40	95.90	97.66
	Stage I	94.08	99.41	97.14	98.26
	Stage II	90.91	98.06	94.08	96.02
	Stage III	86.96	97.03	90.91	93.84
	Stage IV	92.13	99.50	86.96	92.80
	Unstageable	88.66	98.69	92.13	95.29
	SDTI	97.14	97.91	88.66	93.95
	Overall	92.42	98.43	91.64	98.26
DenseNet121	Stage I	93.42	99.03	95.24	97.09
	Stage II	96.10	97.88	93.42	95.59
	Stage III	78.26	98.70	96.10	97.38
	Stage IV	94.38	99.17	78.26	87.48
	Unstageable	91.75	99.06	94.38	96.66
	SDTI	95.24	98.47	91.75	94.99
	Overall	93.71	98.72	91.53	97.09

<sup>a</sup>PI: pressure injury.  
<sup>b</sup>SDTI: suspected deep tissue injury.

Discussion

Principal Findings

Among all the CNN models, DenseNet121 demonstrated the highest overall accuracy of 93.71%. The classification performances of AlexNet, VGGNet16, and ResNet18 exhibited overall accuracies of 87.74%, 82.42%, and 92.42%, respectively.

PIs are a global health issue, and effective treatment requires early and accurate classification and prevention. The staging of the PI is usually a subjective evaluation by medical professionals via standard systems, but it can differ due to variations in staff experience, training, and wound characteristics [22].

CNNs are feedforward neural networks with convolutional computations and deep structures. They capture local image features regardless of position through their core convolutional operation. This allows CNN-based models to potentially reduce assessment disparities among medical personnel. This approach promises precise classification of PIs, ensuring consistent patient care and valuable advice post discharge, which improves quality of life and health care resource efficiency [23].

However, many factors need to be considered when a reliable staging model is constructed (eg, the quality and quantity of images matter for model training). Several previous studies were based on retrospectively collected images in which other types of wounds were included, and the shooting equipment may not have been updated with lower image pixels, which affects the reliability of supervised learning in DL [24-26]. Some researchers have attempted

to classify PIs via neural network models on the basis of the clinical staging system; however, most of these studies have relied on public datasets containing a limited number of images per grade, typically only a few dozen [25,26]. It is recommended that more than 150 training images should be used per grade to achieve reasonable classification accuracy when resources are limited [27].

Therefore, it is necessary to construct a large dataset of high quality with reliable labeling of each grade. In contrast, in this study, PI images were collected prospectively by professional cameras, and a gray card and executable shooting standard were developed to ensure that the quality of each image was clear and that the wound characteristics could be clearly displayed. Moreover, the staging results of each image were assessed by 2 wound therapists simultaneously. The wound therapists included were all qualified and had more than 15 years of working experience, so their assessment results were convincing, which contributed to the good comparability of the images and the reliable sample set. Additionally, we used a professional digital camera and a gray card, which were rarely used in previous studies, to ensure the quality of the images, particularly in accurately representing the original color of PIs [28].

Compared with other studies, this study still has advantages in terms of classification accuracy. Ay et al [24] used the European Pressure Ulcer Advisory Panel staging system and included PIs from stage 1 to stage 4. They trained 1091 images from a public dataset of PIs called the PIID (Public Injury Images Dataset) and 15 images from Google via DenseNet121, InceptionV3, MobileNetV2, ResNet152, ResNet50, and VGG16. The results indicated that the average accuracy of the 6 algorithms for each stage during pretraining

ranged from 54.84% to 77.42% [24]. Kim et al [29] set SE-ResNext101 to train 2614 images from 493 participants. The accuracy of the model was 0.793 over the internal testing set and 0.717 over the external testing set [29]. In our study, DenseNet121 exhibited the highest accuracy (93.71%), precision (98.72%), recall (91.53%), and  $F_1$ -score (97.09%), possibly because of the following reasons. (1) DenseNet121 improves the backpropagation of gradients due to the dense connection mode, making the network easier to train. Moreover, it can reduce the gradient disappearance problem caused by the transmission of input information and gradient information between many layers. (2) The number of parameters is reduced. (3) Low-dimensional features are preserved. In a standard convolutional network, the final output is only used to extract the highest-level features [30]. It is also important to consider the algorithm and image quality. In our study, 5 wound therapists, all of whom were qualified and had been engaged in wound management for at least 15 years, were recruited for grading. Therefore, the results of image staging were relatively accurate, and the classification of each label learned by CNNs was also convincing. Overall, DenseNet121 has become a popular choice in the field of computer vision because of its effective training of deep networks, high performance, generalizability, and insights into the learning process. These advantages make them powerful tools for image recognition and classification tasks. However, DenseNet121 processes a large number of layers, is relatively time-consuming, and consumes considerable computing power. DenseNet121 is more suitable for our present small sample. In regard to large-scale datasets, from the perspective of computing power and time factors, ResNet-18 or a higher level, such as ResNet34 or ResNet50, may be a better choice [31].

To improve the model's adaptability to skin color variations, future research should increase the sample size

and expand collection areas. Additionally, implementing the model in clinical settings requires effective communication due to the perception of opacity in DL methods. Accountability is crucial in the medical field, as errors can have legal implications.

## Limitations

The images obtained from data segmentation in this study contain some normal skin tissue. The next step is to make the wound more prominent or to make a judgment after segmenting the image.

Clinical wound management is influenced by blood and fluid seepage. Future studies should integrate DL with 3D imaging, thermal imaging, and fluid seepage assessment for a more comprehensive wound assessment.

The study also faced challenges due to limited prospective image data and reliance on a single device, potentially leading to biases and affecting model accuracy. This study did not compare the model's performance with that of wound therapists or nurses in discriminating pressure ulcers; instead, it focused on internal validation and algorithm comparison.

## Conclusions

Staging models that use CNNs as their foundation exhibit robust classification capabilities. However, further research is needed to validate the reliability of these observed results. As such, we intend to gather an extensive array of additional imagery and undertake a comparative analysis between the staging outcomes generated by the model and those achieved by frontline clinical nurses. This comparative assessment will allow us to identify any potential discrepancies and disparities between the two, thereby affording us valuable insights for refining the model's performance and suggesting effective strategies for enhancing the skills and capabilities of nurses.

## Acknowledgments

The authors would like to thank the 5 wound therapists for their assistance in assessing staging and Duoxiang Zhao from the Pittsburgh Institute, Sichuan University who finished part of the image processing. This study is funded by the 1·3·5 projects for artificial intelligence (ZYAI24029) at West China Hospital, Sichuan University.

## Data Availability

The raw data supporting the conclusions of this paper will be made available by the authors without undue reservation.

## Conflicts of Interest

None declared.

## References

1. Edsberg LE, Black JM, Goldberg M, McNichol L, Moore L, Sieggreen M. Revised national pressure ulcer advisory panel pressure injury staging system: revised pressure injury staging system. *J Wound Ostomy Continence Nurs*. 2016;43(6):585-597. [doi: [10.1097/WON.0000000000000281](https://doi.org/10.1097/WON.0000000000000281)] [Medline: [27749790](https://pubmed.ncbi.nlm.nih.gov/27749790/)]
2. Li Z, Lin F, Thalib L, Chaboyer W. Global prevalence and incidence of pressure injuries in hospitalised adult patients: a systematic review and meta-analysis. *Int J Nurs Stud*. May 2020;105:103546. [doi: [10.1016/j.ijnurstu.2020.103546](https://doi.org/10.1016/j.ijnurstu.2020.103546)] [Medline: [32113142](https://pubmed.ncbi.nlm.nih.gov/32113142/)]
3. Bennett G, Dealey C, Posnett J. The cost of pressure ulcers in the UK. *Age Ageing*. May 2004;33(3):230-235. [doi: [10.1093/ageing/afh086](https://doi.org/10.1093/ageing/afh086)] [Medline: [15082426](https://pubmed.ncbi.nlm.nih.gov/15082426/)]

4. Prevention and Treatment of Pressure Ulcers/Injuries: Clinical Practice Guideline. 3rd ed. European Pressure Ulcer Advisory Panel, National Pressure Injury Advisory Panel, and Pan Pacific Pressure Injury Alliance; 2019. ISBN: 978-0-6480097-8-8
5. Tschannen D, McKay M, Steven M. Improving pressure ulcer staging accuracy through a nursing student experiential intervention. *J Nurs Educ*. May 1, 2016;55(5):266-270. [doi: [10.3928/01484834-20160414-05](https://doi.org/10.3928/01484834-20160414-05)] [Medline: [27115453](https://pubmed.ncbi.nlm.nih.gov/27115453/)]
6. Young DL, Shen JJ, Estocado N, Landers MR. Financial impact of improved pressure ulcer staging in the acute hospital with use of a new tool, the NE1 Wound Assessment Tool. *Adv Skin Wound Care*. Apr 2012;25(4):158-166. [doi: [10.1097/01.ASW.0000413597.20438.d2](https://doi.org/10.1097/01.ASW.0000413597.20438.d2)] [Medline: [22441048](https://pubmed.ncbi.nlm.nih.gov/22441048/)]
7. Zahia S, Garcia Zapirain MB, Sevillano X, González A, Kim PJ, Elmaghraby A. Pressure injury image analysis with machine learning techniques: a systematic review on previous and possible future methods. *Artif Intell Med*. Jan 2020;102:101742. [doi: [10.1016/j.artmed.2019.101742](https://doi.org/10.1016/j.artmed.2019.101742)] [Medline: [31980110](https://pubmed.ncbi.nlm.nih.gov/31980110/)]
8. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imaging Radiat Sci*. Dec 2019;50(4):477-487. [doi: [10.1016/j.jmir.2019.09.005](https://doi.org/10.1016/j.jmir.2019.09.005)] [Medline: [31601480](https://pubmed.ncbi.nlm.nih.gov/31601480/)]
9. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature New Biol*. Feb 2, 2017;542(7639):115-118. [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)] [Medline: [28117445](https://pubmed.ncbi.nlm.nih.gov/28117445/)]
10. Chang Y, Kim JH, Shin HW, Ha C, Lee SY, Go T. Diagnosis of pressure ulcer stage using on-device AI. *Appl Sci (Basel)*. 2024;14(16):7124. [doi: [10.3390/app14167124](https://doi.org/10.3390/app14167124)]
11. Swerdlow M, Guler O, Yaakov R, Armstrong DG. Simultaneous segmentation and classification of pressure injury image data using Mask-RCNN. *Comput Math Methods Med*. 2023;2023:3858997. [doi: [10.1155/2023/3858997](https://doi.org/10.1155/2023/3858997)] [Medline: [36778787](https://pubmed.ncbi.nlm.nih.gov/36778787/)]
12. Seo S, Kang J, Eom IH, et al. Visual classification of pressure injury stages for nurses: a deep learning model applying modern convolutional neural networks. *J Adv Nurs*. Aug 2023;79(8):3047-3056. [doi: [10.1111/jan.15584](https://doi.org/10.1111/jan.15584)] [Medline: [36752192](https://pubmed.ncbi.nlm.nih.gov/36752192/)]
13. National pressure injury advisory panel. URL: <https://npiap.com/> [Accessed 2025-03-11]
14. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76(5):378-382. [doi: [10.1037/h0031619](https://doi.org/10.1037/h0031619)]
15. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. *J Med Imaging Radiat Oncol*. Aug 2021;65(5):545-563. [doi: [10.1111/1754-9485.13261](https://doi.org/10.1111/1754-9485.13261)] [Medline: [34145766](https://pubmed.ncbi.nlm.nih.gov/34145766/)]
16. Kasajima I. Measuring plant colors. *Plant Biotechnol (Tokyo)*. 2019;36(2):63-75. [doi: [10.5511/plantbiotechnology.19.0322a](https://doi.org/10.5511/plantbiotechnology.19.0322a)] [Medline: [31768106](https://pubmed.ncbi.nlm.nih.gov/31768106/)]
17. Huang L, Qin J, Zhou Y, Zhu F, Liu L, Shao L. Normalization techniques in training DNNs: methodology, analysis and application. *IEEE Trans Pattern Anal Mach Intell*. Aug 2023;45(8):10173-10196. [doi: [10.1109/TPAMI.2023.3250241](https://doi.org/10.1109/TPAMI.2023.3250241)] [Medline: [37027763](https://pubmed.ncbi.nlm.nih.gov/37027763/)]
18. Hosny KM, Kassem MA, Fouad MM. Classification of skin lesions into seven classes using transfer learning with AlexNet. *J Digit Imaging*. Oct 2020;33(5):1325-1334. [doi: [10.1007/s10278-020-00371-9](https://doi.org/10.1007/s10278-020-00371-9)] [Medline: [32607904](https://pubmed.ncbi.nlm.nih.gov/32607904/)]
19. Khan A, Khan A, Ullah M, Alam MM, Bangash JI, Suud MM. A computational classification method of breast cancer images using the VGGNet model. *Front Comput Neurosci*. 2022;16:1001803. [doi: [10.3389/fncom.2022.1001803](https://doi.org/10.3389/fncom.2022.1001803)] [Medline: [36405784](https://pubmed.ncbi.nlm.nih.gov/36405784/)]
20. Nayak T, Chadaga K, Sampathila N, et al. Deep learning based detection of monkeypox virus using skin lesion images. *Med Nov Technol Devices*. Jun 2023;18:100243. [doi: [10.1016/j.medntd.2023.100243](https://doi.org/10.1016/j.medntd.2023.100243)] [Medline: [37293134](https://pubmed.ncbi.nlm.nih.gov/37293134/)]
21. Wu J, Hu W, Wen Y, Tu W, Liu X. Skin lesion classification using densely connected convolutional networks with attention residual learning. *Sensors (Basel)*. Dec 10, 2020;20(24):7080. [doi: [10.3390/s20247080](https://doi.org/10.3390/s20247080)]
22. Sayar S, Aşkın Ceran M, Demir A. Determining the pressure injury and staging knowledge of nurses at a hospital in Turkey. *J Tissue Viability*. Nov 2022;31(4):735-740. [doi: [10.1016/j.jtv.2022.08.004](https://doi.org/10.1016/j.jtv.2022.08.004)] [Medline: [36041891](https://pubmed.ncbi.nlm.nih.gov/36041891/)]
23. Jiang M, Ma Y, Guo S, et al. Using machine learning technologies in pressure injury management: systematic review. *JMIR Med Inform*. Mar 10, 2021;9(3):e25704. [doi: [10.2196/25704](https://doi.org/10.2196/25704)] [Medline: [33688846](https://pubmed.ncbi.nlm.nih.gov/33688846/)]
24. Ay B, Tasar B, Utlu Z, Ay K, Aydin G. Deep transfer learning-based visual classification of pressure injuries stages. *Neural Comput Applic*. Sep 2022;34(18):16157-16168. [doi: [10.1007/s00521-022-07274-6](https://doi.org/10.1007/s00521-022-07274-6)]
25. Fergus P, Chalmers C, Henderson W, Roberts D, Waraich A. Pressure ulcer categorization using deep learning: a clinical trial to evaluate model performance. *arXiv*. Preprint posted online on Mar 7, 2022. [doi: [10.48550/arXiv.2203.06248](https://doi.org/10.48550/arXiv.2203.06248)]
26. Lau CH, Yu KHO, Yip TF, et al. An artificial intelligence-enabled smartphone app for real-time pressure injury assessment. *Front Med Technol*. 2022;4:905074. [doi: [10.3389/fmedt.2022.905074](https://doi.org/10.3389/fmedt.2022.905074)] [Medline: [36212608](https://pubmed.ncbi.nlm.nih.gov/36212608/)]

27. Shahinfar S, Meek P, Falzon G. "How many images do I need?" Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. *Ecol Inform.* May 2020;57:101085. [doi: [10.1016/j.ecoinf.2020.101085](https://doi.org/10.1016/j.ecoinf.2020.101085)]
28. Sampaio CS, Atria PJ, Hirata R, Jorquera G. Variability of color matching with different digital photography techniques and a gray reference card. *J Prosthet Dent.* Feb 2019;121(2):333-339. [doi: [10.1016/j.prosdent.2018.03.009](https://doi.org/10.1016/j.prosdent.2018.03.009)] [Medline: [30093117](https://pubmed.ncbi.nlm.nih.gov/30093117/)]
29. Kim J, Lee C, Choi S, et al. Augmented decision-making in wound care: evaluating the clinical utility of a deep-learning model for pressure injury staging. *Int J Med Inform.* Dec 2023;180:105266. [doi: [10.1016/j.jmedinf.2023.105266](https://doi.org/10.1016/j.jmedinf.2023.105266)] [Medline: [37866277](https://pubmed.ncbi.nlm.nih.gov/37866277/)]
30. Liao T, Li L, Ouyang R, et al. Classification of asymmetry in mammography via the DenseNet convolutional neural network. *Eur J Radiol Open.* Dec 2023;11:100502. [doi: [10.1016/j.ejro.2023.100502](https://doi.org/10.1016/j.ejro.2023.100502)] [Medline: [37448557](https://pubmed.ncbi.nlm.nih.gov/37448557/)]
31. Sun Y, Xue B, Zhang M, Yen GG. Completely automated CNN architecture design based on blocks. *IEEE Trans Neural Netw Learn Syst.* Apr 2020;31(4):1242-1254. [doi: [10.1109/TNNLS.2019.2919608](https://doi.org/10.1109/TNNLS.2019.2919608)] [Medline: [31247572](https://pubmed.ncbi.nlm.nih.gov/31247572/)]

## Abbreviations

**CNN:** convolutional neural network  
**DL:** deep learning  
**NPIAP:** National Pressure Injury Advisory Panel  
**PI:** pressure injury  
**PIID:** Public Injury Images Dataset  
**SDTI:** suspected deep tissue injury

*Edited by Christian Lovis, Qingyu Chen; peer-reviewed by C J Prabhakar, Yunchao Tang; submitted 03.06.2024; final revised version received 21.10.2024; accepted 11.01.2025; published 25.03.2025*

### *Please cite as:*

Lei C, Jiang Y, Xu K, Liu S, Cao H, Wang C

*Convolutional Neural Network Models for Visual Classification of Pressure Ulcer Stages: Cross-Sectional Study*

*JMIR Med Inform* 2025;13:e62774

URL: <https://medinform.jmir.org/2025/1/e62774>

doi: [10.2196/62774](https://doi.org/10.2196/62774)

© Changbin Lei, Yan Jiang, Ke Xu, Shanshan Liu, Hua Cao, Cong Wang. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.