International Journal of Scientific Engineering and Science
Volume 10, Issue 1, pp. 94-99, 2026. ISSN (Online): 2456-7361

# A Study on Multi-Classification Algorithms for Chronic Wounds Using the SARNet Model

Xu Tianxin[1], Wan Nan[2], Tang Anqi[3], Wu Yao[4], Zhu Jin[5], Zhao Zeda[6], Ye Wengang[7]
[1, 2, 3, 5, 6, 7]School of Medical Informatics, Wannan Medical College, Wuhu, Anhui, China,241002
[4]School of Clinical Medicine, Wannan Medical College, Wuhu, Anhui, China,241002

***Abstract*****—***Against the backdrop of rapid internet development and globalization, integrating artificial intelligence with healthcare represents an inevitable trend. This project focuses on deep learning research for chronic wound image segmentation and multi-classification algorithms. It employs the SARNet model trained on the ImageNet dataset and implements a series of operations to reduce computational load and accelerate image processing. After achieving satisfactory classification results, an enhanced ESARNet model is developed through ensemble learning, further improving classification accuracy.*

***Keywords*****—** *Chronic Wounds, Deep Learning, SARNet.*

## I. INTRODUCTION

Chronic wounds refer to injuries that fail to achieve anatomical and functional integrity through normal, orderly, and timely repair processes due to various internal or external factors. They primarily include diabetic foot ulcers, vascular ulcers, pressure injuries, and acute wounds that fail to heal[1].Effective wound management is a crucial measure for promoting chronic wound healing and reducing healthcare expenditures[2]. Complex pathogenesis, coupled with multiple underlying conditions, often leads to complications such as pain, infection, functional impairment, and even amputation, profoundly impacting the quality of life and mental health of 40 million people worldwide[3].

## II. RESEARCH BACKGROUND

### A. Introduction to Chronic Wounds

Due to changes in lifestyle habits and the increasing prevalence of aging, various chronic diseases have become major threats to health, with chronic wounds being one of them[4]. Chronic wounds are those that fail to heal through the normal healing process due to various causes[5-6].The International Wound Healing Society classifies chronic wounds into four major categories: diabetic foot ulcers (DFU), pressure injuries (PI), venous leg ulcers (VLU), and arterial ulcers. Additionally, non-healing surgical incisions and traumatic wounds constitute significant sources of chronic wounds in China[7]. Epidemiological studies over the past three decades reveal that the predominant type of chronic wounds in China has shifted from those caused by various traumas to those primarily resulting from chronic diseases[8-9]. A study indicates[10] that the incidence of PI among hospitalized adults worldwide is 12.8%, with ≥ stage II PI accounting for 56.5% of the total PI incidence.

### B. Introduction to Deep Learning

Deep learning forms the foundation of artificial intelligence and represents its most mature development, encompassing mathematical models such as neural networks, decision trees, support vector machines, clustering algorithms, component analysis, and Bayesian classifiers[11]. In the field of chronic wound care, machine learning is primarily applied to image capture, electronic medical record (EMR) integration, consolidated wound data management, telemedicine, risk prediction, and clinical decision support systems (CDSS)[12].

Effective wound management is a crucial measure for promoting chronic wound healing and reducing healthcare expenditures. Existing chronic wound management models struggle to provide dynamic, comprehensive, and in-depth assessments of chronic wounds. Meanwhile, deep learning, leveraging its powerful data processing capabilities, has achieved significant results in the field of chronic wound care.

### C. Current State of Research

Current algorithms for chronic wound classification predominantly employ feature generation + SVM or end-to-end CNN-based approaches. These methods are designed to distinguish normal skin from a single wound or to grade a single wound, rendering them unsuitable for diagnosing actual chronic wound patients. A small number of models targeting multi-type wound classification also yield unsatisfactory results. Training a model capable of multi-classification for chronic wounds is therefore of significant importance[13].

Klosenik et al[14]. generated a set of features for support vector machines (SVMs) by analyzing color thresholds through multidimensional histograms to separate wound from non-wound areas. Mohammad et al[15]. proposed using a modified Hue-Saturation-Value (HSV) model to generate a Red-Yellow-Kuro-White (RYKW) probability map of the input image. This map then guides image segmentation using optimal thresholds or region growing. Lee et al[16]. employed Euclidean distances from the mean calculated in wound regions as feature vectors, utilizing gradient vector flow to detect wound region contours. Like many computer vision systems, artificially crafted features are susceptible to skin pigmentation, lighting variations, and image resolution. They rely on manually tuned parameters and features crafted based on prior experience, which cannot guarantee optimal results. Due to human limitations, it is

impossible to exhaustively account for all special cases, resulting in poor robustness for atypical scenarios.

## III. Research Process

### A. Dataset Selection

The dataset used in this study is the publicly available chronic wound dataset from ImageNet, comprising 1,000 images of chronic wounds. To utilize additional chronic wound data for self-supervised pre-training, we additionally utilized a publicly available multi-classification dataset of chronic wound images from ImageNet, comprising 2,023 images categorized into eight wound types: diabetic foot, burns, normal skin, pressure ulcers, skin lacerations, surgical wounds, trauma, and venous wounds. The inclusion of this dataset significantly enhances the self-supervised learning potential during pre-training while balancing wound sample diversity across categories. This prevents abnormal segmentation errors caused by sample scarcity in specific wound types. A ResNet50 network serves as the feature extractor, connected to two parallel attention mechanisms from Danet (Jun Fu et al., 2019), followed by retraining on the small wound image dataset. The dataset segmentation results are shown in table I.

TABLE I. Dataset Partitioning Results.

| | 10% | 20% | 30% | 40% | 50% |
|---|---|---|---|---|---|
| Test set | 110 | 110 | 110 | 110 | 110 |
| Training set | 74 | 144 | 216 | 289 | 360 |
| Validation set | 18 | 36 | 54 | 72 | 90 |

### B. Data Preprocessing

Given the limited dataset, random transformations such as rotation and flipping are applied to both original images and labeled images in the training set during training. This generates similar yet distinct training images. Randomly altering training samples not only expands the training set but also reduces the model's reliance on certain irrelevant image attributes during training, thereby improving the model's generalization ability.

Additionally, to accelerate model convergence, batch normalization is applied to the input data. During each iteration, the mini-batch samples undergo a mathematical transformation: subtracting the batch mean, dividing by the batch standard deviation, and then multiplying by the batch scaling factor and batch offset. The mathematical formula for batch normalization is as follows:

$$BN(X) = \gamma \otimes \frac{X - \hat{\mu}_B}{\hat{\sigma}_B} + \beta \quad (1)$$

The parameters γ and β represent the tensile parameter and the offset parameter, respectively.

In the data processing for image recognition, we implemented several data augmentation techniques. Primarily, we employed the Mixup method (Hongyi Zhang et al., 2017) and the Mosaic method (Alexey et al., 2020). Both approaches have been demonstrated to be effective in YOLOv4 and YOLOX.

### C. Data Distortion Strategy

In the preprocessing phase of self-supervised learning, four primary image transformation methods are applied: rotation and flipping, grayscale conversion, color transformation, and cropping. These have all been proven effective for contrastive learning gains. Specifically, the flipping and rotation process comprises three steps: random vertical flipping, random horizontal flipping, and 90-degree rotation. Since these all involve overall positional changes to the image, they are grouped into a single method module. Gray-scale conversion transforms color images into monochrome versions. Prior to experimentation, this was considered the most promising distortion strategy for enhancing segmentation performance. Monochrome images bear a certain similarity to segmented mask images, enabling the learning of more texture and position-related information without interference from wound coloration. Color transformation employs four methods: altering contrast, brightness, hue, and saturation. Cropping involves randomly extracting image blocks covering 0.5–0.8 of the original area, filling the cropped-out sections with black. Since black pixels are inherently easier to segment, filling the cropped gaps with black does not affect final feature extraction. The results obtained by randomly applying these four image transformations are shown in Figure 1.

Fig. 1. Transform the preprocessing results.

### D. Construction of Classification Models

Given that most current research on chronic wound classification relies on single-wound classification using wound image data, with limited multi-classification studies exhibiting low accuracy, we employed a convolutional neural network model named SARNet based on a polar self-attention mechanism, as shown in Figure 2.

The SARNet model employs a multi-branch topology network architecture, enabling each branch to utilize distinct convolutional kernels to achieve different receptive fields and learn deeper features. The model takes chronic wound images as input. The trained network consists of five convolutional blocks, one self-attention layer, and one fully connected layer. Each convolutional block incorporates a 3×3 convolutional kernel alongside a residual branch and a 1×1 convolutional branch. This multi-branch architecture enables the model to

International Journal of Scientific Engineering and Science
Volume 10, Issue 1, pp. 94-99, 2026. ISSN (Online): 2456-7361

preserve rich gradient information while achieving high classification performance. Down sampling begins at the first layer using a stride-2 convolution. An auto-attention module is introduced between the fourth and fifth convolutional layers. By modeling both spatial and channel dimensions, the model maintains strong performance on fine-grained, pixel-level tasks. The final stage outputs classification results via a fully connected layer. Five models with varying layer counts form an integrated system, with their outputs voted to select the most probable classification. This framework effectively classifies wound images into burn, surgical, venous ulcer, pressure injury, diabetic foot ulcer, and normal skin categories.

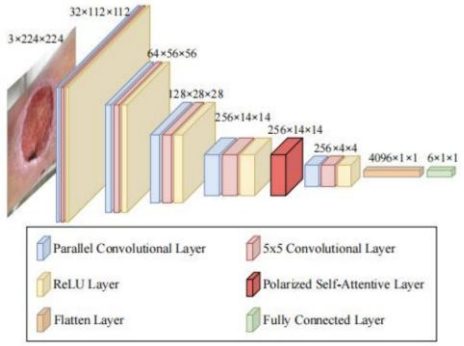


Fig. 2. SARNet Architecture

After obtaining satisfactory classification results, we constructed an ensemble enhancement framework using SARNet models with varying numbers of layers. By employing the Bagging method to combine model outputs and vote for the final result, we improved the classification accuracy by 0.55%. This framework is termed ESARNet.The ESARNet model architecture is shown in Figure 3.

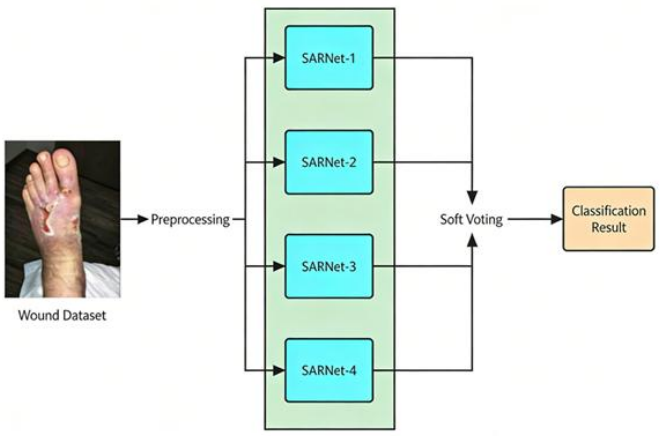


Fig. 3. ESARNet Model Architecture

## *E. Training of Classification Models*

*a.* Experimental Environment

- Pycharm-2019.2
- Python-3.7
- Torch-1.4.0
- Numpy,Sklearn,Visdom
- RTX3060 6G
- Visualization Tool Visdom

*b.* Training Process

- Task Type: Multi-class classification, Logistic regression
- Model: SARNet
- Input Dimensions: 256×256×3
- batch_size: 32
- Iteration count: 2000×14, ImageNet iterated 3000×43 times
- Loss function: Cross-entropy, CrossEntropyLoss()
- Optimizer: Adam
- Training visualization: visdom
- Performance evaluation: precision, recall, AUC, F1, accuracy, etc.

The cross-entropy loss function is defined as follows, with the code shown in Figure 4.

$$L(\hat{Y},Y)=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}y_{i,j}\cdot\log\hat{y}_{i,j} \quad (2)$$

```
def cross_entropy(y_true, y_pred):
    return -np.mean(y_true * np.log(y_pred + 10e-6))
```

Fig. 4. Cross-Entropy Loss Function

*c.* Training code

Partial training code is shown in Figures 5 and 6.

```
    return avg_loss, accuracy

# 验证模型
def validate(model, dataloader, criterion, config):
    """验证模型"""
    model.eval()
    total_loss = 0.0
    total_correct = 0
    total_samples = 0

    all_preds = []
    all_labels = []

    with torch.no_grad():
        for inputs, labels in dataloader:
            inputs, labels = inputs.to(config.device), labels.to(config.device)

            outputs = model(inputs)
            loss = criterion(outputs.logits, labels)

            total_loss += loss.item() * inputs.size(0)
            _, predicted = torch.max(outputs.logits, 1)
            total_correct += (predicted == labels).sum().item()
            total_samples += inputs.size(0)

            all_preds.extend(predicted.cpu().numpy())
            all_labels.extend(labels.cpu().numpy())

    avg_loss = total_loss / total_samples
    accuracy = total_correct / total_samples

    return avg_loss, accuracy, all_preds, all_labels
```

Fig. 5. Partial training code

```
# 绘制训练曲线
def plot_training_curves(train_losses, val_losses, train_accs, val_accs, config):
    """绘制训练曲线"""
    plt.figure(figsize=(12, 5))

    # 损失曲线
    plt.subplot(1, 2, 1)
    plt.plot(train_losses, label='训练损失')
    plt.plot(val_losses, label='验证损失')
    plt.xlabel('Epoch')
    plt.ylabel('损失')
    plt.title('训练和验证损失')
    plt.legend()
    plt.grid(True)

    # 准确率曲线
    plt.subplot(1, 2, 2)
    plt.plot(train_accs, label='训练准确率')
    plt.plot(val_accs, label='验证准确率')
    plt.xlabel('Epoch')
    plt.ylabel('准确率')
    plt.title('训练和验证准确率')
    plt.legend()
    plt.grid(True)

    plt.tight_layout()
    plt.savefig(os.path.join(config.result_dir, f"training_curves_{config.model_id}.png"), dpi=300, bbox_inches='tight')
    plt.close()

    print(f"训练曲线已保存到: {os.path.join(config.result_dir, f'training_curves_{config.model_id}.png')}")

# 绘制混淆矩阵
def plot_confusion_matrix(conf_matrix, class_names, config):
    """绘制混淆矩阵"""
    plt.figure(figsize=(10, 8))
    sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues',
                xticklabels=class_names, yticklabels=class_names)
    plt.xlabel('预测标签')
    plt.ylabel('真实标签')
```

Fig. 6. Code for plotting a portion of the training curve

International Journal of Scientific Engineering and Science
Volume 10, Issue 1, pp. 94-99, 2026. ISSN (Online): 2456-7361

## IV. Analysis and Optimization of Experimental Results

### A. Evaluation Criteria

To evaluate classification performance, we employ accuracy as a metric. Accuracy represents the ratio of correctly predicted data to the total data volume. To assess the model's multi-class classification capability for chronic wounds, we additionally utilize precision, recall, and F1 score as performance indicators. Precision reflects the model's ability to predict specific categories accurately; recall indicates how often the model detects a specific category; and F1 score represents the harmonic mean of precision and recall, providing an estimate of both values. If we denote TP as true positive samples, FP as false positive samples, TN as true negative samples, and FN as false negative samples, the metric equations can be defined as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

$$precision = \frac{TP}{TP + FP} \quad (4)$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - \mathrm{Score} = 2 \times \frac{\mathrm{Re}call \times \mathrm{Pr}ecision}{\mathrm{Re}call + \mathrm{Pr}ecision} \quad (6)$$

### B. Experimental results

We employed SARNet to classify chronic wounds into six categories with an accuracy of 80.87%. Table II lists the abbreviations for classification labels, while Table III presents the results of this six-category classification. The wound set represents all wound types and was used to evaluate the model's performance in a binary classification task distinguishing wounds from normal skin. Additionally, we performed six-class classification on this dataset. The classification results are shown in Table 3, where B, S, V, P, D, and N represent abbreviations for burn, post-operative, venous ulcer, pressure injury, diabetic foot ulcer, and normal skin, respectively.

TABLE II. Category Tag Abbreviations

| Wound | Abbreviation |
|---|---|
| Burn | B |
| Postoperative wound | S |
| Venous ulcer | V |
| Pressure injury | P |
| Diabetic foot ulcer | D |
| Normal skin | N |
| Wound Collection | W |

TABLE III. Model 5-Class and 6-Class Classification Results

| Num of Classes | Classes | Test Accuracy |
|---|---|---|
| 2-classes | WN | 97.58% |
| 5-classes | BDNPS | 84.24% |
| | BDNPV | 82.78% |
| | BDNSV | 83.45% |
| | BDPSV | 75.94% |
| | BNPSV | 83.99% |
| | DNPSV | 84.71% |
| 6-classes | BDNPSV | 80.87% |

SARNet achieved an accuracy of 80.87% for six-class classification. To date, the highest recorded accuracy for wound classification stands at 75.64%, achieved using a VGG16 network on the AZH dataset. Consequently, our accuracy represents an improvement of 5.78%. In contrast, the BDPSV model for five-class classification without normal skin achieved only 75.94% accuracy, falling below the results of other classifications.

### C. Comparison of Results

We compared the accuracy of SARNet, other machine learning models, and deep learning models for the six-category classification of chronic wounds, as shown in Table 4. It is evident that GrowNet achieves only 57.15% accuracy in six-class classification, indicating its unsuitability for multi-class classification of high-resolution images. Meanwhile, machine learning algorithms like Adaboost, XGBoost, and RandomForest also demonstrate suboptimal performance when handling multi-class classification of wounds with relatively similar images. In contrast, multi-branch convolutional neural networks such as SARNet and ResNet meet our expectations.

TABLE IV. Classification results of different models

| Model Algorithm | Wound Type | Classification Accuracy |
|---|---|---|
| GrowNet | BDNPSV | 57.15% |
| Adaboost | | 72.36% |
| XGBoost | | 74.55% |
| Random Forest | | 71.35% |
| ResNet50 | | 75.10% |
| SARNet | | 80.87% |
| ESARNet | | 81.42% |

Additionally, we compared the classification performance of Mixup and attention mechanisms against single models and multi-model ensembles, as shown in Tables V and VI and Figure 7. The multi-branch model architecture without mixup and attention mechanisms achieved only 69.34% classification accuracy. This indicates that mixup and attention mechanisms boosted the model's accuracy by 11.53%, providing ample evidence of their importance. Mixup alleviates the issue of model overfitting when data is scarce, while attention mechanisms enhance the model's focus on local features. Notably, ESARNet employs an ensemble approach for model enhancement. We combine five SARNet models with varying layer counts into a single model, then aggregate their outputs through voting to select the most probable classification result. Experiments demonstrate that the ensemble model achieves a 0.55% higher accuracy than the original model. We also evaluated the impact of Mixup and attention mechanisms on ESARNet classification, as shown in Table VI.

TABLE V. The Impact of Mix-Up and Attention Mechanisms on Model Accuracy

| Classifier | Classes | Test Accuracy |
|---|---|---|
| SARNet without mixup or attention | 6 | 69.34% |
| SARNet without mixup | | 77.24% |
| SARNet without attention | | 71.92% |
| SARNet | | 80.87% |

TABLE VI. The Impact of Multi-Model Ensemble on Model Accuracy

| Classifier | Classes | Test Accuracy |
|---|---|---|
| ESARNet without mixup or attention | 6 | 70.03% |
| ESARNet without mixup | | 77.78% |
| ESARNet without attention | | 72.31% |
| ESARNet | | 81.42% |

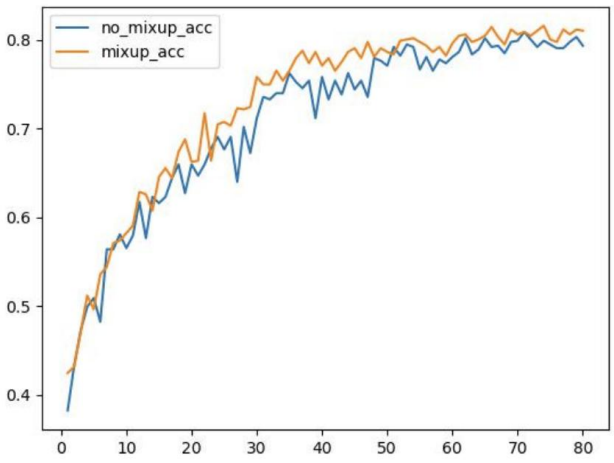


Fig. 7. The Impact of Mixup on Model Accuracy

### D. *Some charts in training*

Figures 8, 9, 10, and 11 respectively illustrate the accuracy across different categories, recall rates across different categories, F1 scores at different levels, and the loss curve evolution between the training and validation sets during the training process. These curves provide a more intuitive way to visualize and analyze the relationships between different indicators across categories, as well as the superiority of the model.

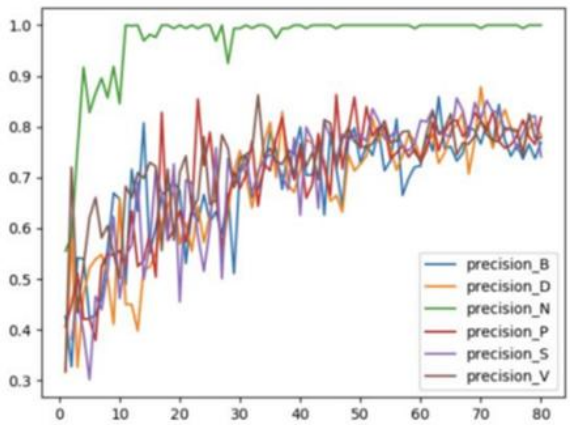


Fig. 8. Accuracy of Different Categories

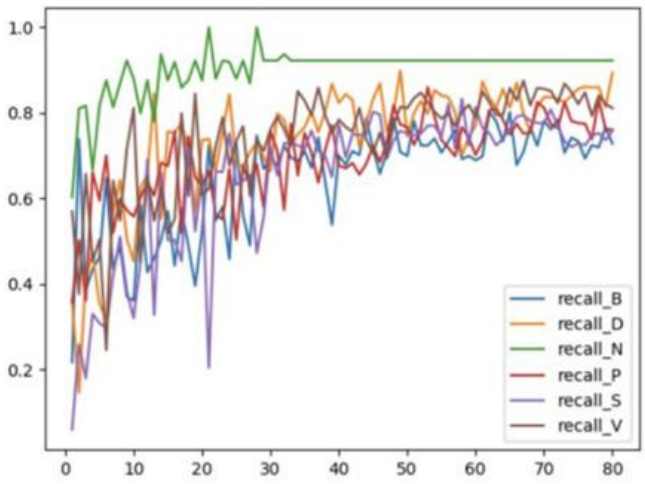


Fig. 9. Recall rates for different categories

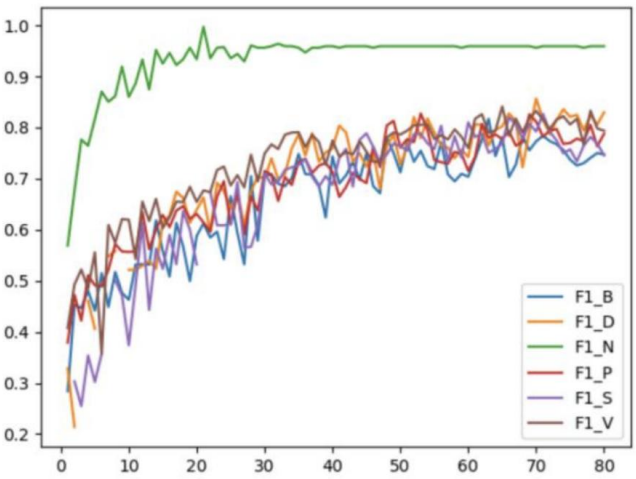


Fig. 10. F1 Grades by Level

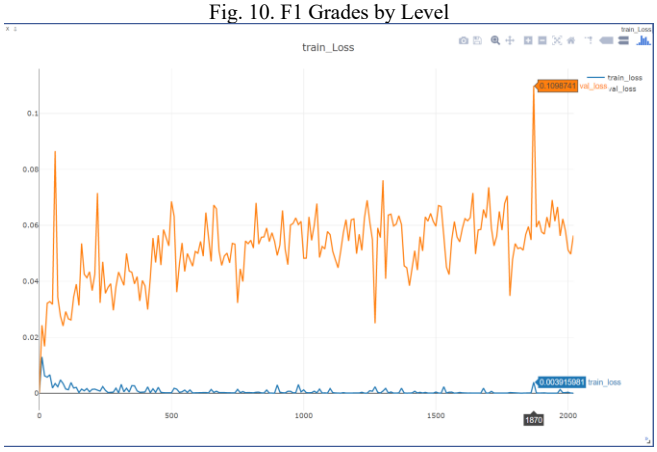


Fig. 11. Training and validation set loss curves during training

## REFERENCES

[1] 丁炎明.伤口护理学[M].人民卫生出版社,2017.
[2] 杨扬,陈德凤,李蓓,等.人工智能用于慢性伤口护理的研究进展[J].护理学杂志, 2024(005):039.
[3] Golinko MS, Clark S, Rennert R, Flattau A, Boulton AJ, Brem H. Wound emergencies: the importance of assessment, documentation, and early treatment using a wound electronic medical record [J] . Ostomy Wound Manage, 2009, 55(5): 54-61.

[4] Graves N, Phillips CJ, Harding K. A narrative review of the epidemiology and economics of chronic wounds. Br J Dermatol. 2022 Aug;187(2):141-148.
[5] van Koppen CJ, Hartmann RW. Advances in the treatment of chronic wounds: a patent review. Expert Opin Ther Pat. 2015;25(8):931-7.
[6] Mirhaj M, Labbaf S, Tavakoli M, Seifalian AM. Emerging treatment strategies in wound care. Int Wound J. 2022 Nov;19(7):1934-1954.
[7] Sun X, Ni P, Wu M, Huang Y, Ye J, Xie T. A Clinicoepidemiological Profile of Chronic Wounds in Wound Healing Department in Shanghai. Int J Low Extrem Wounds. 2017 Mar;16(1):36-44.
[8] 付小兵.进一步推进具有中国特色的创面防控创新体系建设[J].中华创伤杂志, 2017, 33(4):4.
[9] 孟浩.889例体表慢性难愈合创面住院患者临床流行病学研究[D].中国人民解放军医学院,2022.
[10] 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. 2020 May;105:103546.
[11] Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. N Engl J Med. 2023 Mar 30;388(13):1201-1208.
[12] Li D, Mathews C. Automated measurement of pressure injury through image processing. J Clin Nurs. 2017 Nov;26(21-22):3564-3575.
[13] 杨伯寅. 基于深度集成学习的慢性伤口分类研究[D].东华大学, 2023.
[14] Kolesnik M, Fexa A. Multi-dimensional Color Histograms for Segmentation of Wounds in Images [M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 1014–1022.
[15] Ahmad Fauzi M F, Khansa I, Catignani K, et al. Computerized segmentation and measurement of chronic wound Images [J]. Computers in Biology and Medicine, 2015, 60:
[16] Lee H, Lee B-U, Park J, et al. Segmentation of wounds using gradient vector Flow [C]. 2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 20

*This work was supported by the 2024 Wannan Medical College Student Innovation and Entrepreneurship Competition Project "Research on Image Segmentation and Multi-Classification of Chronic Wounds Based on Machine Learning" (Project No.: 202410368071), Research on Deep Learning-Based Automatic Segmentation Methods for Retinal OCT Fluid Images and System Implementation (Project No.: 202410368002), and Implementation of a Deep Learning-Based Intelligent Detection and Growth Prediction System for Pulmonary Nodules (S202510368019).*