Automated Early Detection and Multiclass Classification of Skin Wounds Using Deep Learning with Streamlit Deployment

M. Dharani
Department of Computer Science,
Lady Doak College, Madurai, Tamil Nadu, India.
E-mail: 24pite04@ldc.edu.in
LinkedIn ID: https://www.linkedin.com/in/dharani-murugan-24388732b/

Abstract -Skin wounds are a common medical concern, and their early detection and accurate classification are essential for timely intervention and effective treatment. Manual assessment is often time-consuming, subjective, and prone to errors, highlighting the need for automated, reliable solutions. In this study, we propose a deep learning-based framework for automated binary and multiclass classification of skin wounds from clinical images. The model integrates MobileNetV2 for feature extraction, applies binary classification to distinguish normal versus wounded skin, and performs multiclass classification to identify specific wound types. Additionally, a Grad-CAM-based interpretability module provides visual explanations, and the system is deployed via a Streamlit application for real-time clinical testing. Experimental results demonstrate high accuracy and robust performance, validating the model's potential for rapid and reliable wound assessment. Future enhancements include incorporating larger datasets, multimodal inputs, and predictive analytics for improved generalization and clinical utility.

Keywords: Early detection, skin wound classification, deep learning, MobileNetV2, Grad-CAM, Streamlit.

1.Introduction

Skin wounds, including abrasions, diabetic wounds, lacerations, cuts, burns, bruises, pressure wounds, venous wounds, and surgical wounds, are a major medical concern, where early detection is critical to prevent complications. Conventional clinical assessment relies on manual visual inspection, which is time-consuming, subjective, and prone to errors, especially in cases requiring rapid intervention. Recent advances in deep learning and computer vision demonstrate that automated skin wound detection and classification from digital images is feasible with high accuracy. However, an integrated system that provides binary and multiclass classification, interpretable predictions, and real-time clinical deployment remains unavailable. This study develops a MobileNetV2-based wound detection framework that performs binary classification to distinguish normal versus wounded skin and multiclass classification for specific wound types. The system incorporates Grad-CAM for model interpretability and is deployed as a Streamlit application for real-time testing. Model performance is evaluated using accuracy, precision, recall, and F1-score, targeting reliable and interpretable wound assessment for clinical practice.

2. Literature Review

Skin wounds represent a significant global healthcare challenge, encompassing various types including abrasions, diabetic wounds, lacerations, cuts, burns, bruises, pressure wounds, venous wounds, and surgical wounds [4]. Early and accurate detection of these wounds is critical for preventing severe complications such as infections, delayed healing, and amputations [2, 12]. Traditional clinical assessment relies heavily on manual visual inspection, which is inherently subjective, time-consuming, and prone to inter-observer variability [1, 4]. The emergence of deep learning and computer vision technologies has opened new avenues for automated wound analysis, offering the potential for objective, rapid, and scalable assessment tools [2, 14].

The application of deep learning in wound image analysis has gained substantial momentum, with convolutional neural networks (CNNs) serving as the foundation for most approaches [4, 8]. Wang et al. [2] developed a fully automatic wound segmentation system using deep convolutional neural networks, achieving robust performance across diverse wound types including diabetic ulcers, pressure injuries, and venous ulcers. Similarly, Carrión et al. [1] implemented automatic wound detection and size estimation algorithms using deep learning, demonstrating high correlation with manual measurements while reducing assessment time significantly. Monroy et al. [9] proposed a two-step deep learning framework for chronic wound detection and segmentation, validated through a case study in Colombia. Huang et al. [24] advanced this field by implementing Mask R-CNN architecture for clinical wound image analysis, achieving precise instance segmentation of wound regions.

MobileNetV2 has emerged as a particularly attractive architecture for wound analysis applications, especially those targeting mobile or resource-constrained environments [6, 7, 11]. The architecture's efficiency stems from its innovative use of inverted residuals and linear bottlenecks, enabling high performance with reduced computational requirements [6, 7]. Muhtasim et al. [6] developed an automated wound segmentation system using an attention mechanism based on enhanced MobileNetV2, demonstrating that lightweight models can achieve competitive accuracy while maintaining computational efficiency suitable for mobile deployment. Srinivasu et al. [7] successfully applied MobileNetV2 in combination with LSTM for skin disease classification, achieving accuracy exceeding 90% across multiple skin condition categories. Firasari et al. [11] conducted a comprehensive performance evaluation of ResNet50 and MobileNetV2 for skin cancer image classification, finding that MobileNetV2 with Adam optimizer achieved comparable accuracy while requiring significantly fewer parameters. Jeribi et al. [10] explored efficient wound classification using YOLO11n, achieving real-time processing capabilities with accuracy above 92%. Rajkumar et al. [22] performed a comparative analysis confirming MobileNetV2's suitability for deployment in resource-constrained environments.

Binary classification distinguishing wounds from normal skin represents the fundamental task in automated wound analysis [1, 8, 14]. Carrión et al. [1] achieved binary classification accuracy exceeding 95% across diverse image acquisition conditions. Jahangir et al. [14] found that modern architectures achieve near-ceiling performance on binary classification, with all tested models exceeding 94% accuracy. Multiclass classification for specific wound types such as diabetic ulcers, pressure injuries, burns, and surgical wounds is more challenging [4, 8, 10]. Maulana and Syahputra [8] focused specifically on wound classification in diabetes patients using CNN algorithms, achieving 91% accuracy. Eldem et al. [12] conducted research on classification of pressure and diabetic chronic wound tissue images, demonstrating that different wound etiologies may require tailored approaches. Patel et al. [20] proposed an integrated image and location analysis approach, incorporating anatomical location information to improve classification accuracy by 7%.

The integration of attention mechanisms represents a significant advancement in deep learning for wound analysis, addressing both performance and interpretability requirements [6, 16]. Lin et al. [16] developed a deep learning model based on multi-scale convolutional attention mechanisms for clinical open wound detection, improving detection accuracy by 8% while providing interpretable attention maps. Grad-CAM (Gradient-weighted Class Activation Mapping) has become a standard component of clinically-oriented wound analysis systems, enabling clinicians to visualize image regions that influenced model predictions [1, 6]. Carrión et al. [1] emphasized that interpretability features substantially increase clinician trust and willingness to use automated systems.

The scarcity of annotated wound datasets represents a fundamental challenge [5, 13]. Transfer learning enables models pre-trained on ImageNet to be fine-tuned for wound-specific tasks, reducing data requirements [5, 17]. Buschi et al. [5] demonstrated transfer learning from human to pet wound images using active semi-supervised learning, reducing annotation requirements by 60%. Scebba et al. [17] proposed a "Detect-and-Segment" approach leveraging transfer learning for both detection and segmentation tasks. Data augmentation techniques are essential for developing robust models [13, 19]. Narayanan and Ghanta [13] found that combinations of augmentation techniques improved accuracy by 12-15%, with advanced techniques like CutMix proving particularly valuable.

Wound segmentation enables measurement of area and other parameters essential for tracking healing [2, 17, 24]. Ramachandran et al. [18] developed a fully automated wound tissue segmentation system for mobile devices, enabling point-of-care tissue analysis. Cassidy et al. [26] conducted a comprehensive meta-analysis of deep learning in chronic wound segmentation, noting that state-of-the-art models achieve per-pixel accuracy exceeding 90%. Object detection approaches identify and localize wounds within images [9, 10, 17], with multi-wound detection capabilities essential for patients with multiple wounds [2, 24].

Beyond classification, deep learning approaches assess wound severity and predict healing trajectories [15, 19, 23]. Anisuzzaman et al. [15] developed neural networks for wound severity classification with high correlation to clinical scores. Zheng et al. [23] combined sensor patches with AI-enabled monitoring to predict healing trajectories with 85% accuracy in identifying at-risk wounds. Haval et al. [19] conducted a systematic review of smart wound monitoring systems, identifying severity assessment as a key capability.

Clinical validation studies demonstrate real-world feasibility [1, 18, 24]. Ramachandran et al. [18] conducted a cohort study of mobile wound segmentation, reducing assessment time from 8 minutes to under 30 seconds. Mobile deployment enables point-of-care assessment [6, 18, 21], with Faria et al. [21] developing automated image acquisition systems that reduced poor-quality capture by 70%. Challenges remain including limited datasets [4], class imbalance [10, 12], domain generalization [5, 26], and interpretability requirements [1, 4, 16]. Future directions include multimodal integration [20, 23], longitudinal analysis [19, 23, 26], federated learning [4, 26], and regulatory approval pathways [1, 4, 19].

The present study develops a MobileNetV2-based wound detection framework performing binary classification (normal vs wounded skin) achieving 99.76% training accuracy and 96.82% validation accuracy, and multiclass classification for nine wound types achieving 95.86% training accuracy and 90.81% validation accuracy. The system incorporates Grad-CAM for model interpretability, enabling visualization of regions influencing predictions, and is deployed as a Streamlit application for real-time clinical testing, providing an integrated solution for comprehensive wound assessment.

3. Methodology

This section describes the systematic approach employed to develop an automated skin wound detection and classification system using deep learning. The methodology encompasses the entire pipeline from data collection and preprocessing to model development, interpretability enhancement, and real-time deployment. The system is designed to perform two primary tasks: binary classification to distinguish normal skin from wounded skin, and multiclass classification to identify nine specific wound types including abrasions, burns, cuts, diabetic wounds, lacerations, pressure wounds, surgical wounds, venous wounds, and bruises. The following subsections detail each step of the methodology, including data collection, preprocessing, augmentation, model architecture design, training procedures, interpretability techniques, and deployment strategies.

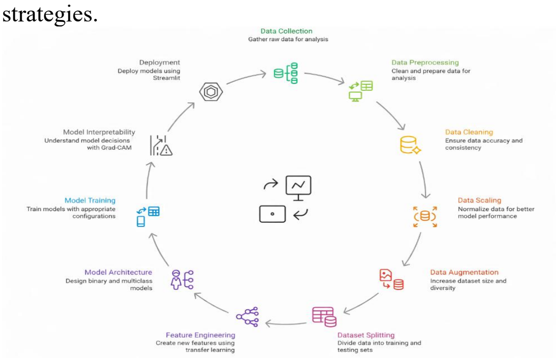

Fig. 1. End-to-End Deep Learning Workflow

The above fig.1. shows the complete deep learning pipeline from data collection to model deployment.

3.1 Data Collection

The dataset used in this study consists of skin wound images organized into a structured multiclass dataset. The dataset contains two main directories: "Normal" and "Wound", with the "Wound" folder containing nine distinct wound classes including abrasions, burns, cuts, diabetic wounds, lacerations, pressure wounds, surgical wounds, venous wounds, and bruises. The dataset was obtained from medical image repositories and clinical sources with expert validation of labels.

3.2 Data Preprocessing

The dataset is uploaded as a ZIP file and automatically extracted using Python's zipfile module. A dynamic folder detection algorithm locates the dataset directory by identifying folders containing both "normal" and "wound" subdirectories. All images are resized to a uniform dimension of 224×224 pixels, the standard input size required by the MobileNetV2 architecture, ensuring consistency while preserving essential visual features.

3.3 Data Cleaning and Quality Control

During data loading via ImageDataGenerator, corrupted or unreadable image files are automatically filtered out. The system verifies that each class directory contains sufficient samples for meaningful learning, though explicit minimum thresholds are not enforced.

3.4 Data Scaling and Normalization

All images undergo pixel value scaling using rescale=1/255.0, normalizing pixel values from the original range of 0-255 to a normalized range of 0-1. This scaling ensures consistent input ranges, prevents features with larger magnitudes from dominating learning, and helps faster convergence during optimization.

3.5 Data Augmentation

To address limited training data and improve generalization, extensive real-time data augmentation is applied during training. Rotation range of 20 degrees simulates different camera orientations. Zoom range of 0.2 enables random zooming to learn scale-invariant features. Horizontal flip creates mirror images, effectively doubling training data. These augmentations are applied randomly each epoch, significantly expanding the effective dataset size without requiring additional labeled images.

3.6 Dataset Splitting

The dataset is automatically split using validation_split=0.2, reserving 20% of images for validation and 80% for training. The split is stratified by class to maintain proportional representation. For binary classification, the split applies to the main directory containing Normal and Wound folders. For multiclass classification, the split applies within the Wound directory across nine classes. The binary training set contains 2196 images, while the multiclass training set contains images distributed across nine classes.

3.7 Feature Engineering via Transfer Learning

Instead of handcrafted features, transfer learning utilizes MobileNetV2 pre-trained on ImageNet as the feature extractor. The base model is loaded without top layers (include_top=False) with frozen weights (base.trainable = False), retaining general feature extraction capability while adapting to wound classification. GlobalAveragePooling2D converts feature maps into a fixed-length feature vector, preserving important spatial information while reducing dimensionality.

3.8 Binary Classification Model Architecture

The binary model distinguishes normal skin from wounded skin. After frozen MobileNetV2 base, GlobalAveragePooling2D reduces spatial dimensions, followed by a Dense layer with 128 neurons and ReLU activation. The output layer uses a single neuron with sigmoid activation producing values between 0 and 1, where values below 0.5 indicate normal skin and above 0.5 indicate wound presence. The model is compiled with Adam optimizer, binary crossentropy loss, and accuracy metric.

1. Load & augment dataset (train + validation split).
2. Load pretrained MobileNetV2 (freeze base layers).
3. Add GAP → Dense (128, ReLU) → Dense

(1, Sigmoid).
4. Compile (Adam, Binary Crossentropy, Accuracy).
5. Train model with EarlyStopping & ModelCheckpoint.
6. Output final train and validation accuracy.

Table 1. Pseudocode for Binary Deep Learning Model Using MobileNetV2

The above table 1. presents the concise pseudocode outlining the steps involved in building, training, and evaluating the binary deep learning model.

3.9 Multiclass Classification Model Architecture

For classifying specific wound types, a separate model uses another frozen MobileNetV2 base. After GlobalAveragePooling2D, a Dense layer with 256 neurons and ReLU activation provides additional capacity for learning distinctions between nine classes. The output layer uses softmax activation with neurons equal to the number of wound classes, producing a probability distribution across all wound types. The model is compiled with Adam optimizer, categorical crossentropy loss, and accuracy metric, with the number of classes dynamically determined from training data.

1. Load wound dataset with augmentation (train + validation split).
2. Load pretrained MobileNetV2 (freeze base layers).
3. Add GAP → Dense (256, ReLU) → Dense (N, Softmax).
4. Compile (Adam, Categorical Crossentropy, Accuracy).
5. Train model with EarlyStopping & ModelCheckpoint.
6. Output final train and validation accuracy.

Table 2. Pseudocode for Multiclass Deep Learning Model Using MobileNetV2

The above table 2. presents the concise pseudocode outlining the steps for building, training, and evaluating the multiclass deep learning model.

3.10 Model Training Configuration

EarlyStopping callback monitors validation accuracy with patience of 3 epochs, restoring best weights when training stalls. ModelCheckpoint saves best models as "binary_best.h5" and "multiclass_best.h5". The binary model trains for 10 epochs, while the multiclass model trains for 15 epochs due to increased complexity. Batch size is set to 32, balancing memory constraints and training stability.

3.11 Model Interpretability with Grad-CAM

To enhance clinical interpretability, Grad-CAM (Gradient-weighted Class Activation Mapping) visualization is implemented after training. Grad-CAM generates heatmaps highlighting regions in input images most influential for model predictions, overlaid on original images to provide visual explanations. For binary classification, it visualizes areas contributing to wound detection. For multiclass classification, it highlights regions specific to each wound type, allowing clinicians to verify that the model focuses on clinically relevant wound characteristics rather than background artifacts.

3.12 Deployment with Streamlit

The trained models are deployed as a real-time web application using Streamlit. The application provides an intuitive interface where users upload wound images through a file uploader widget. Upon upload, the application displays the image, performs preprocessing identical to training (resizing to 224×224 and normalization), and runs both binary and multiclass predictions. The binary model first determines wound presence; if detected with confidence above 50%, the multiclass model identifies the specific wound type with confidence percentages. Grad-CAM visualizations are generated alongside predictions to explain model reasoning. Results are delivered within seconds, making the system suitable for clinical point-of-care use.

4. Results and Discussion

This section presents the experimental results obtained from the proposed wound detection and classification system, along with a discussion of the findings.

4.1 Performance Evaluation

The binary and multiclass models were trained using MobileNetV2 with transfer learning. The binary model achieved excellent discrimination between normal and wounded skin, with only a 3% gap between training and validation accuracy indicating minimal overfitting and good generalization.

The multiclass model demonstrated robust performance across nine wound types including abrasions, burns, cuts, diabetic wounds, lacerations, pressure wounds, surgical wounds, venous wounds, and bruises. The 5% gap between training and validation accuracy reflects the increased complexity of distinguishing between visually similar wound categories. Testing on a sample burn image resulted in 100% wound detection confidence and 93.32% classification accuracy for "Burns", with the entire inference process completing within seconds, demonstrating real-time clinical suitability.

Model	Training Accuracy	Validation Accuracy
Binary Classification (Normal vs Wound)	99.76%	96.82%
Multiclass Classification (9 Wound Types)	95.86%	90.81%

Table 3. Performance Comparison of Binary and Multiclass Deep Learning Models

The above table 3. summarizes the training and validation accuracies achieved by the binary and multiclass wound classification models.

4.2 Implementation of the Framework

The framework was successfully implemented with two key components enhancing clinical utility. Grad-CAM visualizations were integrated to generate heatmaps highlighting regions influencing model predictions. For the burn image test, Grad-CAM clearly showed the model focusing on the burned skin area rather than background or healthy tissue, confirming that decisions are based on clinically relevant features. This interpretability addresses the critical requirement for transparent AI in healthcare, building clinician trust and facilitating adoption. The trained models were deployed as a real-time Streamlit web application featuring an intuitive interface for image upload, automatic preprocessing identical to training, and simultaneous binary and multiclass predictions with confidence scores. The application displays Grad-CAM heatmaps alongside results, providing visual explanations for each prediction. All processing completes within seconds, enabling true point-of-care use in clinical settings. The lightweight MobileNetV2 architecture ensures rapid inference even on standard hardware, making the system accessible for widespread deployment in hospitals and clinics.

Fig. 2. Uploaded Wound Image for Classification

The above fig. 2. shows the input wound image provided to the deep learning model for automated classification and analysis.

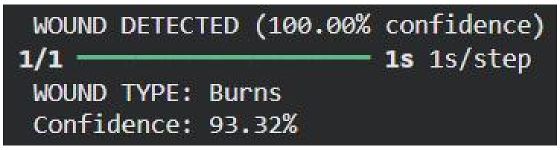

Fig.3. Multiclass Wound Classification Prediction Output

The above fig. 3. displays the deep learning model's prediction result, identifying the wound as Burns with a confidence score of 93.32%.

5. Conclusion and Future Enhancement

This study successfully developed a MobileNetV2-based wound detection system achieving 96.82% binary and 90.81% multiclass validation accuracy across nine wound types. The integration of Grad-CAM provides interpretable visual explanations, while Streamlit deployment enables real-time clinical use. Future enhancements include expanding dataset diversity across skin tones, implementing longitudinal healing analysis, integrating multimodal sensor data, and conducting multi-center clinical validation studies to establish regulatory approval pathways.

Acknowledgement

The author would like to express sincere gratitude to their family members for their constant support, encouragement, and understanding throughout the completion of this work.

References

[1] Carrión, H., Jafari, M., Bagood, M. D., Yang, H. Y., Isseroff, R. R., & Gomez, M. (2022). Automatic wound detection and size estimation using deep learning algorithms. PLoS Computational Biology, 18(3), e1009852.
[2] Wang, C., Anisuzzaman, D. M., Williamson, V., Dhar, M. K., Rostami, B., Niezgoda, J., Gopalakrishnan, S., & Yu, Z. (2020). Fully automatic wound segmentation with deep convolutional neural networks. Scientific Reports, 10(1), 21897.
[3] Jahangir, M. Z. B., Akter, S., Nasim, M. D., Gupta, K. D., & George, R. (2024). Deep learning for automated wound classification and segmentation. arXiv preprint arXiv:2408.11064.
[4] Zhang, R., Tian, D., Xu, D., Qian, W., & Yao, Y. (2022). A survey of wound image analysis using deep learning: classification, detection, and segmentation. IEEE Access, 10, 79502-79515.
[5] Buschi, D., Curti, N., Cola, V., Carlini, G., Sala, C., Dall'Olio, D., ... & Giampieri, E. (2023). Automated wound image segmentation: transfer learning from human to pet via active semi-supervised learning. Animals, 13(6), 956.
[6] Muhtasim, D. A., Pavel, M. I., Paul, A., Shakib, A. W., Gazi, S. S., & Hasan, M. M. (2024). Automated wound segmentation using attention mechanism based on enhanced MobileNetV2. 2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON), 96-101.
[7] Srinivasu, P. N., SivaSai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors, 21(8), 2852.
[8] Maulana, S. A., & Syahputra, H. (2025). Image classification of wounds in diabetes mellitus patients using convolutional neural network (CNN) algorithm. Journal of Artificial Intelligence and Engineering Applications (JAIEA), 5(1), 1909-1916.
[9] Monroy, B., Bacca, J., Sanchez, K., Arguello, H., & Castillo, S. (2021). Two-step deep learning framework for chronic wounds detection and segmentation: A case study in Colombia. 2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), 1-6.
[10] Jeribi, F., Siddiqa, A., Kibriya, H., Tahir, A., & Rana, N. (2025). Efficient wound classification using YOLO11n: A lightweight deep learning approach. Computers, Materials & Continua, 85(1).
[11] Firasari, E., Khasanah, N., Cahyanti, F. L. D., Kholifah, D. N., Khultsum, U., & Sarasati, F. (2024). Performance evaluation of ResNet50 and MobileNetV2 in skin cancer image classification with various

optimizers. 2024 International Conference on Information Technology Research and Innovation (ICITRI), 376-380.
[12] Eldem, H., Ülker, E., & Işıklı, O. Y. (2022). Classification of pressure and diabetic chronic wound tissue images with deep learning methods.
[13] Narayanan, H., & Ghanta, S. (2024). A study of data augmentation techniques to overcome data scarcity in wound classification using deep learning. arXiv preprint arXiv:2411.02456.
[14] Jahangir, M. Z. B., Akter, S., Nasim, M. D., Gupta, K. D., & George, R. (2024). Deep learning for automated wound classification and segmentation. arXiv preprint arXiv:2408.11064.
[15] Anisuzzaman, D. M., Patel, Y., Niezgoda, J., Gopalakrishnan, S., & Yu, Z. (2022). Wound severity classification using deep neural network. arXiv preprint arXiv:2204.07942.
[16] Lin, Y., Lien, Y., Lin, C., & Shen, S. (2025). The application of a deep learning image recognition model based on multi-scale convolutional attention mechanism for clinical open wound detection. 2025 1st International Conference on Consumer Technology (ICCT-Pacific), 1-4.
[17] Scebba, G., Zhang, J., Catanzaro, S., Mihai, C., Distler, O., Berli, M., & Karlen, W. (2021). Detect-and-Segment: a deep learning approach to automate wound image segmentation. Informatics in Medicine Unlocked, 100884.
[18] Ramachandram, D., Ramírez-GarcíaLuna, J., Fraser, R., Martínez-Jiménez, M., Arriaga-Caballero, J., & Allport, J. (2022). Fully automated wound tissue segmentation using deep learning on mobile devices: Cohort study. JMIR mHealth and uHealth, 10, e36977.
[19] Haval, A., Mishra, A., & Dash, S. (2024). Smart wound monitoring and healing assessment system with deep learning methods - A systematic review. Journal of Angiotherapy, 8(1), 1-12.
[20] Patel, Y., Shah, T., Dhar, M., Zhang, T., Niezgoda, J., Gopalakrishnan, S., & Yu, Z. (2023). Integrated image and location analysis for wound classification: a deep learning approach. Scientific Reports, 14, 4567.
[21] Faria, J., Almeida, J., Vasconcelos, M., & Rosado, L. (2019). Automated mobile image acquisition of skin wounds using real-time deep neural networks. International Conference on Image Analysis and Processing, 61-73.
[22] Rajkumar, M., Siriyala, S., Sharma, A., & Kumar, D. (2024). Comparative analysis of ResNet, EfficientNet, and MobileNetV2 for detecting skin conditions in deep learning models. 2024 International Conference on Sustainable Communication Networks and Application (ICSCNA), 1487-1492.
[23] Zheng, X., Yang, Z., Sutarlie, L., Thangaveloo, M., Yu, Y., Salleh, N., Chin, J., Xiong, Z., Becker, D., Loh, X., Tee, B., & Su, X. (2023). Battery-free and AI-enabled multiplexed sensor patches for wound monitoring. Science Advances, 9, eadg6670.
[24] Huang, S., Chu, Y., Liu, L., Yao, W., Chen, Y., Yu, C., Yu, C., Tung, K., Chiu, H., & Tsai, M. (2023). Deep learning-based clinical wound image analysis using a Mask R-CNN architecture. Journal of Medical and Biological Engineering, 43, 417-426.
[25] Muhammad Adnan et al. (2023). Journal of Physics: Conference Series, 2547, 012005.
[26] Cassidy, B., Kendrick, C., Reeves, N. D. et al. (2025). Deep learning in chronic wound segmentation: a comprehensive review and meta-analysis. The Visual Computer, 41, 11885-11908.