

Optimization of plantar foot thermogram for diabetic foot ulceration early detection: An image enhancement approach

Muhammad Nuril Huda, Aina Musdholifah*, Aufaclav Zatu Kusuma Frisky
Department of Computer Science and Electronics, Universitas Gadjah Mada, Sleman 55281, Indonesia

ABSTRACT

Diabetes mellitus (DM) is a critical health condition caused by insulin production failure, leading to elevated blood glucose levels. DM often results in severe complications such as heart disease, stroke, and diabetic foot ulcers (DFU), which pose risks of infection and potential amputation. This study developed a machine learning model for early detection of diabetic foot ulcers, using thermogram images and the thermo dataset containing detailed foot temperature data. The multi-classifier model integrates CNNs for processing thermogram images and an ANN for tabular data analysis. Various image enhancement techniques were applied, including solarize, CLAHE, posterize, and gamma adjustment, to improve the visibility of key temperature distribution patterns. The results demonstrate that solarize consistently emerged as the most effective image enhancement method, significantly improving model performance across all evaluation metrics. Models enhanced with solarize achieved an impressive accuracy of 97.06%, alongside a perfect AUC score of 1,000. Additionally, the application of image enhancement techniques proved instrumental in reducing training and inference times, indicating computational efficiency. The integration of temperature data with enhanced thermogram images further boosted predictive accuracy while maintaining critical thermal information. This study underscores the transformative potential of image enhancement techniques, particularly solarize, in advancing the accuracy and efficiency of early detection models for diabetic foot ulcers. These findings contribute meaningfully to the development of medical imaging technologies, offering a robust framework for improving disease diagnosis and management.

ARTICLE INFO

Article history:

Received Jan 13, 2025

Revised Feb 14, 2025

Accepted Feb 21, 2025

Keywords:

Deep Learning
Diabetic Foot Ulceration
Image Enhancement
Multi Classifier
Thermogram

This is an open access article under the [CC BY](#) license.



* Corresponding Author

E-mail address: aina_m@ugm.ac.id

1. INTRODUCTION

Diabetes Mellitus (DM) is a chronic disease characterized by the pancreas's inability to produce sufficient insulin, leading to elevated blood glucose levels. While the primary etiology remains uncertain, genetic predisposition and lifestyle factors significantly contribute to its onset [1]. If left unmanaged, diabetes can result in severe complications, including cardiovascular disease, stroke, renal failure, diabetic foot ulcers (DFUs), and retinopathy [2]. Common symptoms include polyuria, polydipsia, and polyphagia [3].

Among these complications, DFUs are particularly prevalent and serious. DFUs manifest as open sores on the plantar surface of the feet in diabetic patients and carry a high risk of infection. Contributing factors include diabetic neuropathy, arthropathy, and peripheral vascular disease. DFUs can progress to tissue infection, gangrene, and may ultimately necessitate amputation [4].

Thermography has emerged as a non-invasive technique for early detection of DFUs by analyzing temperature distributions on the foot's surface [5–7]. This contactless imaging modality

measures real-time surface temperatures, offering insights into blood flow and skin integrity without physical interaction [7]. However, the inherently low quality of thermographic images mandates the use of image enhancement techniques to improve their diagnostic utility.

Image enhancement is crucial in medical imaging to adjust brightness, contrast, and other parameters, thereby improving image quality [8]. A multitude of enhancement algorithms have been developed and effectively applied in medical imaging, remote sensing, and computer vision [9–11]. Applying these techniques to thermograms can reveal critical features otherwise difficult to detect, enhancing diagnostic accuracy.

In this study, we developed a machine learning model for early detection of DFUs using enhanced thermographic foot images. We employ image enhancement techniques such as solarization, posterization, Contrast Limited Adaptive Histogram Equalization (CLAHE), and gamma adjustment to improve image clarity. While these methods accentuate key visual patterns, they may obscure essential temperature details crucial for diagnosis. To address this, we incorporate tabular data containing detailed temperature readings from the thermograms and additional patient information to enrich the model's performance and preserve vital diagnostic information.

Our proposed approach aims not only to enhance the accuracy of early DFU detection but also to contribute to the advancement of medical imaging technologies in disease diagnosis. By integrating image enhancement techniques with machine learning and supplementary data, we seek to improve medical image quality and diagnostic outcomes.

2. THEORETICAL REVIEW

2.1. Research on Thermography

Thermography has emerged as a promising non-invasive technique for early detection of diabetic foot ulceration by analyzing plantar temperature distributions. Vardasca et al. [12] classified neuro-ischemic and ischemic diabetic foot ulcers (DFUs) using data from 39 patients with active DFUs. By employing Support Vector Machines (SVM), they achieved an accuracy of 87.5% without applying any image enhancement techniques, underscoring the potential of thermograms in detecting severe diabetic complications.

Adam et al. [13] investigated thermograms from 33 healthy and 33 diabetics individuals, using Discrete Wavelet Transform (DWT) and Higher Order Spectra (HOS) for feature extraction. Their SVM model yielded an accuracy of 89%, sensitivity of 82%, and specificity of 97%. Similarly, Balasenthilkumaran et al. [14] processed thermograms from 122 diabetic and 45 non-diabetic subjects using segmentation, noise reduction, and feature extraction methods such as Gray-Level Co-occurrence Matrix (GLCM). They achieved a classification accuracy of 93.3% with a Multilayer Perceptron (MLP) classifier. Jayapal et al. [15] analyzed 314 diabetic and 160 non-diabetic cases, employing Chi-square feature selection and achieving 93% accuracy with an SVM classifier, further validating the efficacy of thermographic analysis in diabetes classification.

Mounika and Thirunavukkarasu [16] attained 90% accuracy using a Naive Bayes classifier without preprocessing or feature extraction, relying on variables like alpha values and confidence intervals for feature selection. Vítor and Teixeira [17] achieved 85% accuracy through pixel clustering and thresholding techniques on thermograms from 122 diabetic and 45 non-diabetic cases, highlighting the role of preprocessing in enhancing classification performance. Khandakar et al. [6] applied Adaptive Histogram Equalization (AHE) and gamma correction, extracting features such as Estimated Temperature Difference (ETD). Using multiple classifiers, they reported that an MLP classifier was achieved 91% accuracy, demonstrating the benefits of combining image enhancement and features selection for improved model performance.

Despite these advances, the inherently low quality of thermographic images often hinders accurate information extraction. This limitation indicates a need for advanced data processing techniques to enhance image quality and facilitate more effective analysis.

2.2. Research on Medical Data Using Image Enhancement

Image enhancement techniques are vital for improving the clarity and diagnostics precision of medical images. Gamara et al. [18] applied Contrast Limited Adaptive Histogram Equalization

(CLAHE) and Wiener filtering to chest X-rays, increasing classification accuracy to 78%. Deepak and Bharanidharan [19] utilized Parabolic Balance Contrast Enhancement (PBCE) for osteosarcoma classification, achieving an impressive 98.5% accuracy with an ensemble model. This highlights PBCE's capability to enhance contrast and reveal intricate details.

Tasci et al. [20] employed CLAHE for tuberculosis detection on chest X-ray datasets, achieving accuracies of 97.5% and 97.6% with Inception V3 and Xception models, respectively. Anand et al. [9] improved chest X-ray classification accuracy from 54% to 66% by using techniques like CLAHE and histogram equalization, underscoring the value of enhancing image details. Kuruba and Gopalan [10] applied contrast adjustment for retinal vessel detection, achieving up to 99% accuracy on the CHASE_DB1 dataset using IterNet, illustrating the role of image enhancement in simplifying complex medical image patterns.

However, the application of image enhancement techniques to DFU thermograms remains limited, and even fewer studies have integrated enhanced images with tabular data, such as detailed temperature readings and patient information. Existing research on thermogram-based DFU detection often reports lower accuracy levels, indicating substantial room for improvement. This gap presents an opportunity to enhance diagnostic performance by applying image enhancement techniques that have proven successful in other medical imaging domains. By integrating these techniques with thermogram images and combining them with comprehensive tabular data, critical features that are otherwise difficult to detect can be identified, thereby improving the overall accuracy of diagnostic models.

2.3. Image Enhancement

The selection of image enhancement techniques—Contrast Limited Adaptive Histogram Equalization (CLAHE), Posterization, Solarization, and Gamma Adjustment—was strategically based on their unique abilities to enhance thermogram images, making critical features related to foot temperature distribution more discernible.

CLAHE and Gamma Adjustment represent computational approaches focused on optimizing image contrast and brightness to clarify key features. CLAHE enhances local contrast and highlights fine details without over-amplifying noise, making it particularly useful in medical image processing where subtle variations are diagnostically significant. Gamma Adjustment modifies the brightness levels across the image, enhancing areas of low or high intensity to make subtle temperature variations more distinct.

In contrast, Posterization and Solarization offer more stylized transformations aligned with artistic approaches. Posterization reduces the number of tonal levels in the image, which can decrease noise and simplify temperature distribution patterns, aiding in the identification of significant anomalies through an abstract representation. Solarization partially inverts pixel intensities, creating a dramatic effect that emphasizes temperature anomalies and highlights significant differences in thermal patterns.

By employing these diverse enhancement techniques, we aim to provide multiple perspectives on foot temperature features. This variety allows the machine learning model to extract richer information from the image, potentially capturing features that might be overlooked using standard enhancement methods. Ultimately, integrating these techniques is expected to improve the model's accuracy in detecting diabetic conditions by enhancing the visibility of key diagnostic features in thermogram images.

2.4. Posterize

Posterization is an image enhancement technique that reduces the number of colors intensity levels in an image, creating a visual effect reminiscent of a poster with simplified and segmented color regions. In the context of 8-bit RGB images—where each color channel (red, green, and blue) ranges from 0 to 255—posterization is applied individually to each channel to simplify the overall color palette and accentuate important visual features.

For images with three color channels (RGB), posterization must be applied individually to each color channel. The general formula for applying posterization to an RGB image is as follows:

$$Interval_i = \frac{I_{max_i}}{levels-1} \quad (1)$$

where I_{max} represents the maximum value in the color channel i (255 for an 8-bit RGB image), and levels is the desired number of color intensity levels.

In the context of an 8-bit RGB image, where each color channel (red, green, and blue) can take on a maximum value of 255, the process of posterization begins by first determining the desired number of color intensity levels, referred to as "levels." These levels dictate how many distinct intensity values each channel will be reduced to, thereby simplifying the color range of the image.

Once the desired number of intensity levels is set, the next step involves recalculating the pixel values for each color channel. This is done by taking the original pixel value of a given color channel and dividing it by the interval corresponding to the number of levels. The resulting value is then rounded to the nearest intensity level, reducing the number of distinct colors in the image.

After obtaining the new intensity level for each color channel, the pixel value is recalculated by multiplying the new intensity level by the same interval that was used in the previous step. This operation is repeated for each pixel in image, effectively reduces the color complexity of entire image.

$$Posterized_i = round\left(\frac{i}{Interval_i}\right) \times Interval_i \quad (2)$$

where i represents the red, green, and blue color channels respectively, and $interval_i$ refers to (1).

2.5. Solarize

Solarization is an image enhancement technique that inverts the pixel intensities of an image beyond a specific threshold value. In an RGB image, solarization can be applied to each color channel (red, green, and blue) individually. This process transforms the appearance of the image by reversing the colors in the brighter areas while leaving the darker areas mostly unaffected. The solarization operation can be mathematically defined as follows:

$$I'(x, y) = \begin{cases} I(x, y) & \text{if } I(x, y) \leq T \\ 255 - I(x, y) & \text{if } I(x, y) > T \end{cases} \quad (3)$$

where $I(x, y)$ represents the original pixel intensity at location (x, y) , $I'(x, y)$ is the solarized pixel intensity, T is the solarization threshold, and the intensity values are assumed to be in the range $[0, 255]$. For pixel values greater than T , the inversion operation enhances contrast by flipping high-intensity values to their complementary low values.

2.6. Contrast Limited Adaptive Histogram Equalization (CLAHE) for RGB Image

Contrast Limited Adaptive Histogram Equalization (CLAHE) is an image enhancement technique designed to improve image contrast while minimizing noise amplification. Unlike traditional histogram equalization, which globally adjusts contrast across the entire image, CLAHE operates on small, localized regions called tiles. This localized enhancement brings out subtle details in areas with varying contrast without over-amplifying noise.

When applying CLAHE to RGB images, processing each color channel (red, green, and blue) independently can lead to color distortions because contrast enhancements may alter each channel non-uniformly. To avoid such artifacts, a common practice is to convert the RGB image to the LAB color space. In LAB space, the L channel represents lightness (intensity), while the A and B channels carry color information (chrominance). By applying CLAHE exclusively to the L channel, we enhance brightness and contrast without affecting the image's color balance.

After the image is converted to LAB format, CLAHE is applied to the L channel. CLAHE operates by enhancing the contrast in local regions of the image, dividing it into small tiles, and equalizing the histogram within each tile. For each tile in the L channel, the CLAHE algorithm involves the following steps:

1. Divide the image into tiles, the L channel is divided into non-overlapping tiles of size $m \times n$.
2. Clip the histogram, the histogram for each tile is computed, and bins that exceed a predefined threshold T are clipped. This step prevents over-enhancement of noise or other small details.
3. Redistribute clipped values, the excess values from the clipped bins are redistributed across the histogram to balance the intensity levels and avoid extreme contrast in small areas.
4. Apply Histogram Equalization, the pixel values within each tile are redistributed based on the cumulative distribution function (CDF) of the clipped histogram. The formula for the new intensity $I'(x, y)$ for each pixel is:

$$I'(x, y) = \frac{CDF(I(x, y)) - \min(CDF)}{\max(CDF) - \min(CDF)} \times (L_{max} - L_{min}) + L_{min} \quad (4)$$

where $I(x, y)$ is the original pixel intensity in the L channel at position (x, y) , $CDF(I(x, y))$ is the cumulative distribution function of the pixel's intensity in the tile, $\max(CDF)$ and $\min(CDF)$ are the minimum and maximum values of the CDF in the tile, L_{max} and L_{min} the minimum and maximum intensity values in the L channel.

Once CLAHE has been applied to the L channel, the modified LAB image is converted back to RGB format. During this conversion, the A and B channels, which store color information, remain unchanged. The transformation from LAB back to RGB can be written as:

$$R, G, B = f^{-1}(L', A, B) \quad (5)$$

where f^{-1} represents the inverse transformation from LAB to RGB, ensuring that the enhanced lightness L' is reflected in the final RGB image.

2.7. Gamma Adjustment

Gamma Adjustment or also known as Gamma Correction is a process used to optimize the tonal scale of an image by adjusting its brightness values. Mathematically, gamma correction can be expressed using the following power-law equation:

$$I'(x, y) = C \times I(x, y)^\gamma \quad (6)$$

where $I(x, y)$ is the original pixel intensity at position (x, y) (in the range $[0, 1]$ for normalized pixel values or $[0, 255]$ for 8-bit images), $I'(x, y)$ is the gamma-corrected pixel intensity, C is a constant used for normalization, typically set to 1 when dealing with normalized pixel values (e.g., when the pixel values are in the $[0, 1]$ range), γ is the gamma value (typically between 0.8 and 2.2 for most displays). For RGB images, gamma correction can be applied to each color channel (R, G, B) individually using the same formula.

3. MATERIALS AND METHOD

The research workflow shown in Figure 1, the thermogram images of the soles of the feet obtained from the IEEE Dataport dataset, published by Hernandez-Contreras et al. [7].

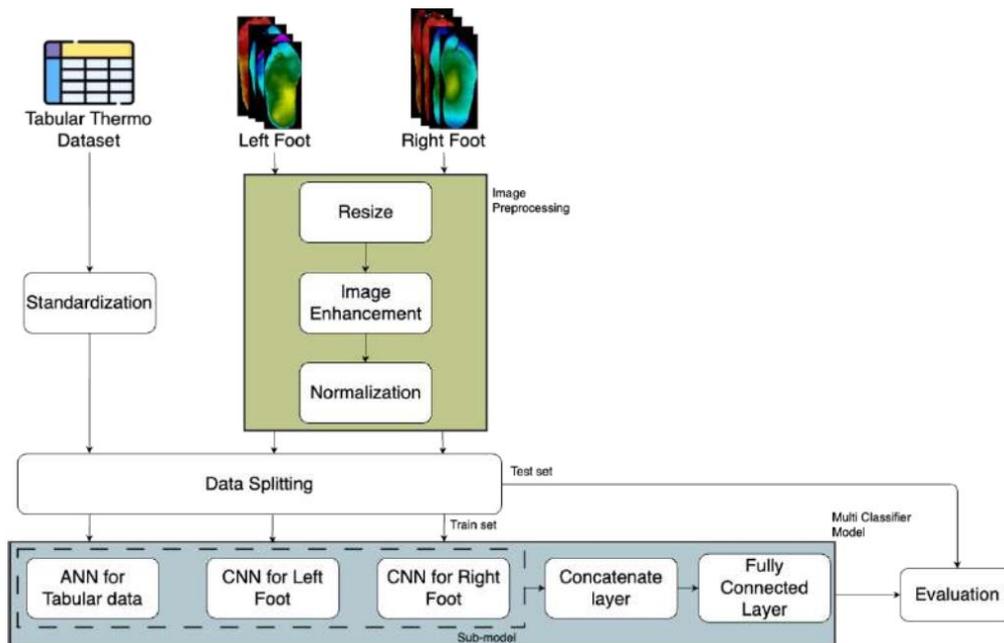


Figure 1. Research workflow.

The dataset includes thermogram images and the corresponding Thermo Dataset, which contains detailed temperature measurements and patient information. The dataset consists of 122 diabetic patients and 45 non-diabetic patients. The thermogram images visualize the temperature distribution on the surface of the feet, providing valuable insights into patterns associated with diabetes-related complications.

The tabular data in the Thermo Dataset as shown in Table 1 undergoes normalization using the Standard Scaler defined as follows:

$$z = \frac{x - \mu}{\sigma} \quad (7)$$

where, Z is the standardized value, x is the original data point, μ is the mean (average) of the dataset, and σ is the standard deviation of the dataset. This process standardizes the values, ensuring a mean of zero and a standard deviation of one.

Table 1. Thermo dataset features.

| Features | Description |
|---------------|---|
| Subject | Represents the unique identity of each subject in the dataset (DM001, DM002, CG001, etc.); this feature is used as a label. |
| Gender | The gender of each subject. |
| General_Right | The general temperature of the right sole. |
| LCA_Right | The temperature in the area of the Lateral Calcaneal Artery (outer heel) of the right foot. |
| LPA_Right | The temperature in the area of the Lateral Plantar Artery (outer sole) of the right foot. |
| MCA_Right | The temperature in the area of the Medial Calcaneal Artery (inner heel) of the right foot. |
| MPA_Right | The temperature in the area of the Medial Plantar Artery (inner sole) of the right foot. |
| TCL_Right | Thermal Change Index, the temperature change in the right foot over time. |
| General_Left | The general temperature of the left sole. |

Each thermogram image was resized to 64×149 pixels, matching the average dimensions of the dataset. We applied various image enhancement techniques: Posterize with 8, 64, and 512 colors; Solarize with thresholds of 64, 128, and 192; CLAHE with adjusted clip limits and tile grid sizes; and Gamma Adjustment with gamma values of -1.5, 1.25, 1.5, 1.75, 2, and 5. To standardize the input data and facilitate model convergence during training, pixel intensity values were normalized to the [0, 1] range by dividing by 255.

Following preprocessing, the thermogram images and accompanying tabular data were randomly split into training and testing sets with an 80:20 ratio, ensuring even distribution for optimal model training and evaluation.

We developed a custom multi-input classifier specifically designed for diabetic foot ulcer detection. The architecture components:

1. Convolutional Neural Networks (CNNs): Two CNN sub-models process thermogram images of the left and right feet independently. Each CNN extracts spatial and temperature-related features from the respective thermogram images.
2. Multi-Layer Perceptron (MLP): An MLP processes the normalized tabular data from the Thermo Dataset, extracting information from temperature measurements and other patient-related data. The MLP consists of fully connected layers with activation functions such as ReLU to model complex relationships in the tabular data.
3. Concatenate Layer: The extracted features from the CNNs and the MLP are concatenated, followed by fully connected layers to generate the final prediction.

The detailed architectures of the MLP, CNNs, and Fully Connector layers are provided in Tables 2, 3, and 4, respectively. To mitigate overfitting—where a model performs exceptionally well on training data but fails to generalize to unseen data—we incorporated L2 regularization and dropout in the fully connected layers.

Table 2. MLP architecture for tabular data.

| Layers | Neuron | Activation |
|---------|--------|------------|
| Dense | 64 | relu |
| Dense | 128 | relu |
| Dense | 256 | relu |
| Dropout | 0.5 | - |

Table 3. CNNs architecture for plantar foot thermogram data.

| Layers | Left foot | | | Right foot | | |
|--------------------|--------------|------------------|------------|--------------|------------------|------------|
| | Total filter | Kernel/pool size | Activation | Total filter | Kernel/pool size | Activation |
| Conv2D | 64 | (3, 3) | relu | 128 | (3, 3) | relu |
| MaxPooling2D | - | (2, 2) | - | - | (2, 2) | - |
| BatchNormalization | - | - | - | - | - | - |
| Flatten | - | - | - | - | - | - |

In the fully connected layers depicted, L2 regularization and dropout are both employed to prevent overfitting, which is a common challenge in training machine learning models, particularly when the model learns to perform exceptionally well on training data but fails to generalize to new, unknown data.

By integrating thermogram images and tabular data within a unified framework, our custom model leverages the complementary nature of visual and numerical information to enhance the early detection of diabetic foot ulceration.

Table 4. Fully connected layer architecture.

| Layers | Neuron/dropout size | Regularizer | Activation |
|-------------|---------------------|-------------|------------|
| Concatenate | - | - | - |
| Dense | 32 | L2 | relu |
| Dropout | 0.2 | - | - |
| Dense | 64 | L2 | relu |
| Dropout | 0.2 | - | - |
| Dense | 128 | L2 | relu |
| Dropout | 0.2 | - | - |
| Dense | 1 | - | sigmoid |

We compiled the model using the Adam optimizer with a learning rate of 1×10^{-4} . Binary cross-entropy loss was employed, aligning with the binary classification objective of diabetes detection. Model performance was monitored using the accuracy metric throughout the training process.

Training was conducted with a batch size of 32 over 200 epochs for each image enhancing method, ensuring consistency and enabling a fair comparison of their respective impacts on model performance. Independent training for each enhancement technique provided a detailed analysis of its specific effect on the model.

The evaluation model incorporated standard performance metrics—including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC)—along with an assessment of inference time. These metrics offered a comprehensive view of each model's effectiveness and the influence of different image enhancement techniques on performance, particularly in distinguishing diabetic patients from non-diabetic controls.

By considering both performance metrics and inference time, we aimed to balance accuracy and efficiency, which is crucial for real-world applications. This approach ensures that the proposed models are not only theoretically robust but also practically applicable within the constraints of time-sensitive medical diagnostics.

4. RESULTS AND DISCUSSIONS

In this study, we evaluated four image enhancement techniques—CLAHE, Gamma Adjustment, Posterization, and Solarization—applied to thermogram images of diabetic and non-diabetic patients' soles. Each method was tested independently to enhance image clarity and improve the model's accuracy in early diabetic foot ulcer detection.

4.1. Image Enhancement Results

The application of different enhancement techniques led to notable variations in the thermogram images, influencing the model's ability to extract meaningful features. Each method uniquely alters the images' contrast, color distribution, and overall clarity.

4.1.1. Posterization Results

Figure 2 illustrates the effects of Posterization with three color levels: 8, 64, and 512 colors. This technique reduces the number of colors in the thermograms, simplifying temperature gradients and accentuating distinct thermal regions.

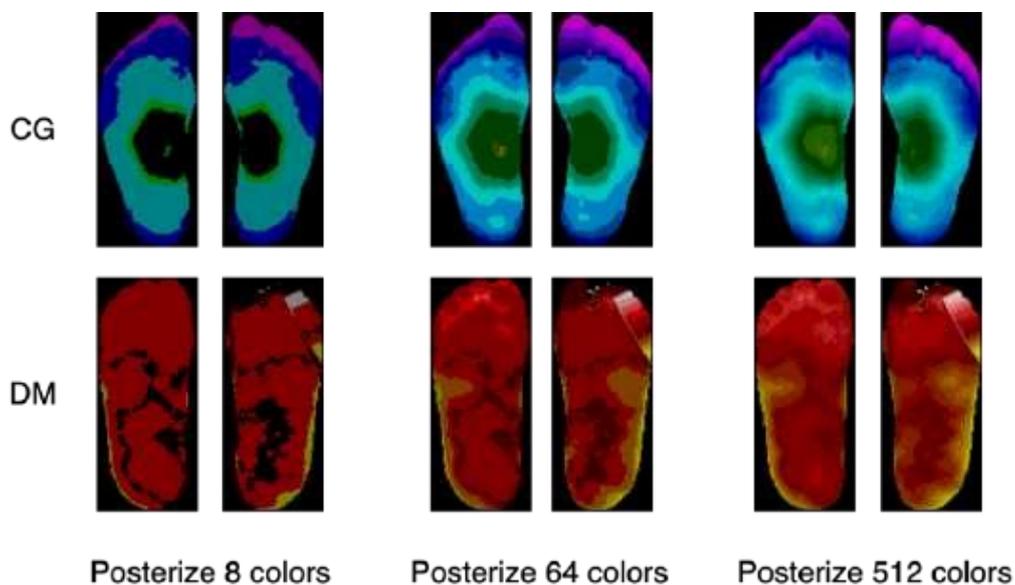


Figure 2. Posterize result.

Posterize 8 Colors significantly simplifies the thermograms, highlighting large thermal regions while eliminating finer details. Although it aids in identifying broad patterns, it may fail to capture subtle temperature differences critical for nuanced diagnoses, potentially diminishing diagnostic utility in complex cases.

Posterize 64 Colors balance simplicity and detail, this configuration effectively detects intermediate thermal patterns. It maintains sufficient resolution to preserve key diagnostic features while reducing noise, providing a more balanced input for the model.

Posterize 512 nearly as detailed as the original images. They capture subtle thermal variations and preserve the most fine details. However, the increased complexity may introduce noise and divert the model's focus from larger thermal patterns.

Posterization offers flexibility in preprocessing by adjusting the level of detail to align with specific modeling needs. Higher simplicity benefits general pattern detection, while greater detail preserves diagnostic subtleties.

4.1.2. Solarization Results

Figure 3 presents the results of applying Solarization with thresholds of 64, 128, and 192. This technique inverts pixel intensities beyond a specific threshold, enhancing temperature contrasts.

Overall, Solarization provides a powerful means of enhancing temperature contrasts in thermograms, with each threshold level catering to different diagnostic priorities. While lower thresholds (e.g., 64) capture subtle variations, they risk introducing noise, whereas higher thresholds (e.g., 192) focus on broader patterns at the expense of finer details.

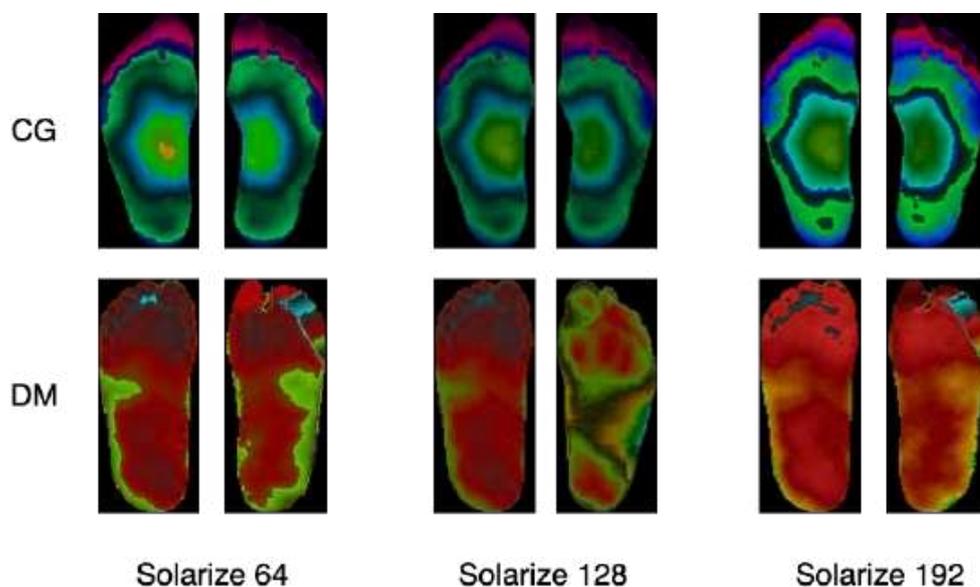


Figure 3. Solarize results.

4.1.3. CLAHE Results

Figure 4 shows the results of CLAHE with various clip limits and tile grid sizes. Visually, these configurations appear similar in terms of enhanced contrast and structural preservation.

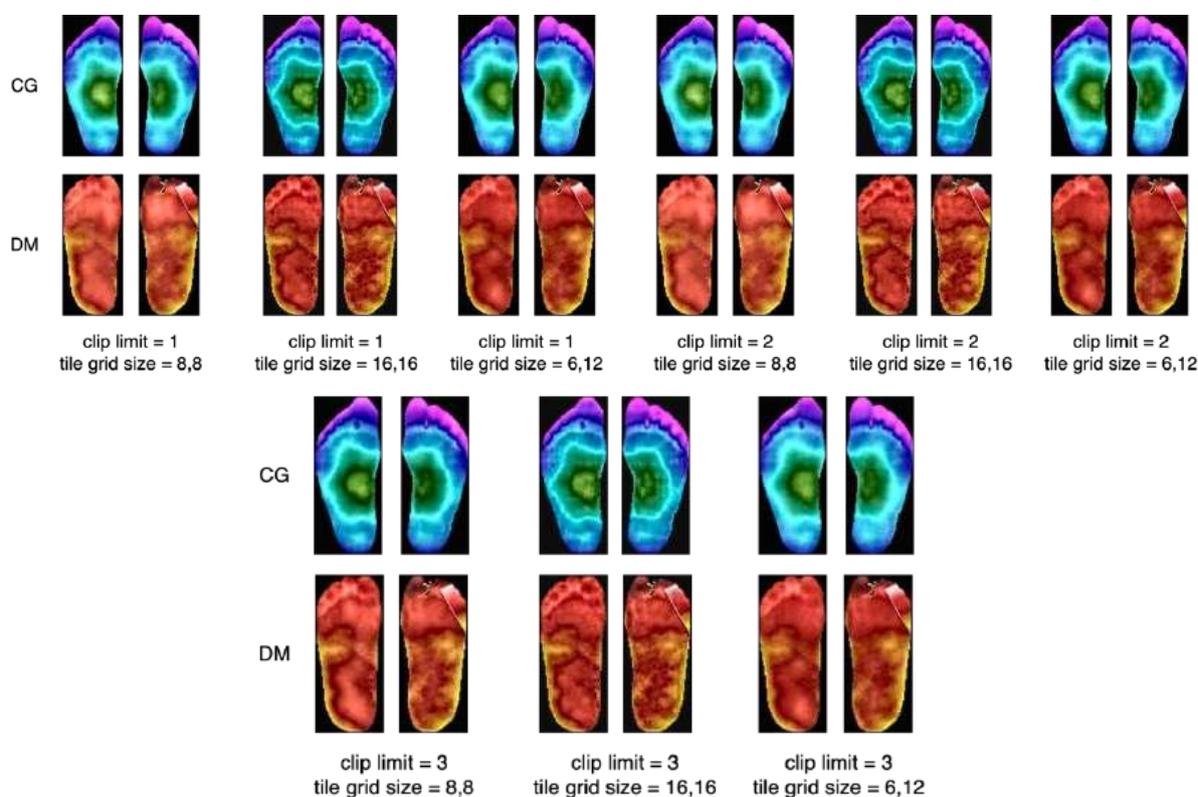


Figure 4. CLAHE results.

To quantitatively assess and identify optimal parameters, we conducted Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) analyses, summarized in Table 5.

PSNR is a widely used metric that quantifies the reconstruction quality of an image by comparing the original image to its enhanced version, with higher values indicating better image quality. The formula of PSNR is:

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (8)$$

where MAX is the maximum possible pixel value of the image (e.g., 255 for 8-bit images), and MSE is the Mean Squared Error between the original and the compressed image.

SSIM, on the other hand, focuses on assessing the perceptual quality by comparing the structural information of the images, where a higher SSIM value indicates greater similarity between the original and the enhanced image. The formula of SSIM is:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (9)$$

where μ_x and μ_y are the mean pixel values of images x and y , σ_x^2 and σ_y^2 are the variances of x and y , σ_{xy} is the covariance between x and y , C_1 and C_2 are constants to stabilize the division.

The analysis revealed that several configurations produced identical PSNR and SSIM values. For example, a clip limit of 2 with tile grid sizes of 8×8 , 16×16 , and 6×12 yielded the same metrics as a clip limit of 1 with corresponding grid sizes. This uniformity suggests that training models on all these variations would be redundant.

Table 5. PSNR and SSIM results.

| Image enhancement | PSNR (dB) | SSIM |
|--|-----------|--------|
| CLAHE (clip = 1, grid = 8×8) | 17.69 | 0.7310 |
| CLAHE (clip = 1, grid = 16×16) | 19.83 | 0.5989 |
| CLAHE (clip = 1, grid = 6×12) | 19.38 | 0.7380 |
| CLAHE (clip = 2, grid = 8×8) | 17.69 | 0.7310 |
| CLAHE (clip = 2, grid = 16×16) | 19.83 | 0.5989 |
| CLAHE (clip = 2, grid = 6×12) | 19.38 | 0.7380 |
| CLAHE (clip = 3, grid = 8×8) | 19.10 | 0.6792 |
| CLAHE (clip = 3, grid = 16×16) | 19.83 | 0.5989 |
| CLAHE (clip = 3, grid = 6×12) | 19.38 | 0.7380 |

To optimize computations, we narrowed down the configurations. We selected a clip limit of 2 with a grid size of 8×8 as one optimal setup due to its balance of performance and efficiency. Additionally, we chose a clip limit of 3 with grid sizes of 8×8 , 16×16 , and 6×12 , as they offered distinct PSNR and SSIM values compared to other clip limits.

4.1.4. Gamma Adjusted Results

Figure 5 illustrates the application of Gamma Adjustment with gamma values of -1.5 , 1.25 , 1.5 , 1.75 , 2 , and 5 . This technique manipulates pixel intensity, compressing or expanding the dynamic range to emphasize different aspects of thermal distribution.

Gamma -1.5 generates images with extreme contrast and significant noise, as shown in Figure 6. Thermal gradients appear inverted, and smooth temperature transitions are replaced with chaotic patterns. Despite reduced interpretability, the model performed unexpectedly well with gamma -1.5 , possibly due to enhanced visibility of distinct temperature anomalies critical for classification. This suggests that extreme transformations may reveal subtle patterns impossible in standard images.

Gamma 1.25 to 1.75 produces smoother transitions and emphasizes either lighter or darker regions, depending on the specific value. They increase broader temperature patterns, aiding in detecting general thermal anomalies. Gamma 1.5 , in particular, balances enhancement and interpretability, preserving critical features while maintaining a natural look appearance.

Gamma 2 exhibit noticeably higher brightness levels, with warm regions becoming more pronounced and transitions appearing smoother. This adjustment increases the visibility of areas related to diabetic foot symptoms but may potentially obscure finer details in cooler regions, reducing the model's ability to detect subtle variations.

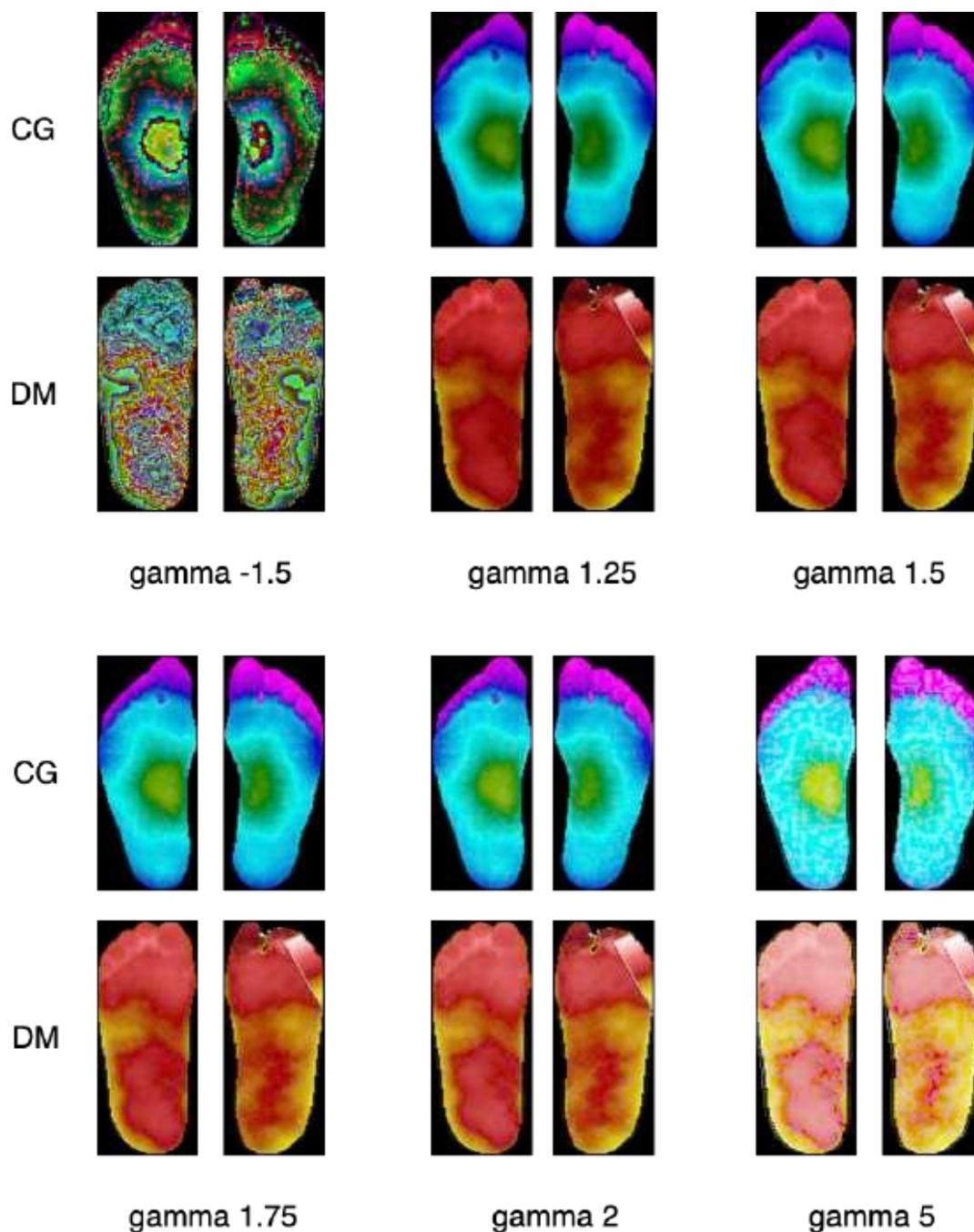


Figure 5. Gamma adjustment results.

At Gamma 5, images become exceedingly bright, and distinctions between thermal gradients diminish significantly. Lower-temperature regions are washed out, and the necessary contrast for identifying subtle anomalies is reduced. This may hinder the model's generalization due to the loss of critical image details.

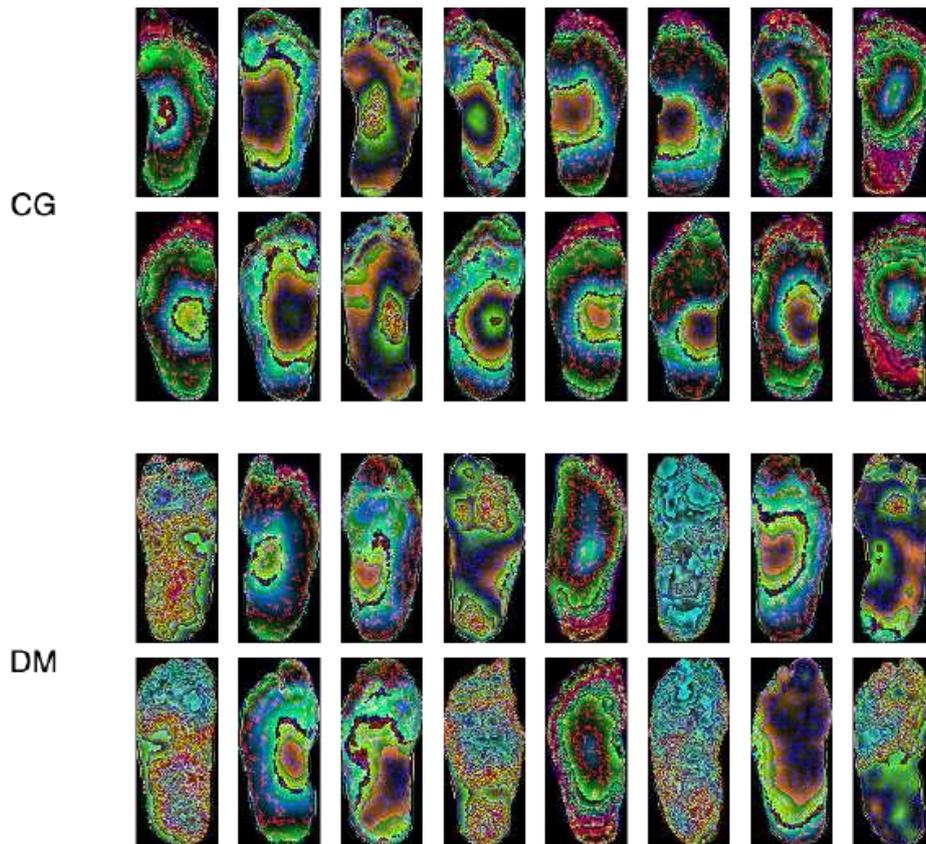


Figure 6. Sample images on gamma -1.5.

4.2. Modeling Results

The application of various image enhancement techniques had distinct impacts on model performance, each presenting unique strengths and challenges as summarized in Table 6. The "Diff" column in the table represents the difference between the model's overall accuracy and its fold mean accuracy, providing insights into the consistency and generalization of each method. A smaller "Diff" value indicates better alignment between the overall accuracy and cross-validation results, suggesting robust generalization across different folds.

Table 6. Results of modeling for all image enhancement.

| Image enhancement | Evaluation metrics (%) | | | | | Time (s) | | Fold mean accuracy (%) | Diff |
|---------------------|------------------------|-----------|--------|----------|-------|----------|-----------|------------------------|-------|
| | Accuracy | Precision | Recall | F1-score | AUC | Training | Inference | | |
| Original | 88.24 | 100 | 86.21 | 92.59 | 97.24 | 76.64 | 0.2 | 83.23 | 5.01 |
| Posterize 8 | 94.12 | 93.55 | 100 | 96.67 | 100 | 61.65 | 0.14 | 80.80 | 13.32 |
| Posterize 64 | 91.18 | 93.33 | 96.55 | 94.92 | 97.93 | 66.56 | 0.15 | 83.78 | 7.4 |
| Posterize 512 | 88.24 | 100 | 86.21 | 92.59 | 97.24 | 78.38 | 0.18 | 85.03 | 3.21 |
| Solarize 64 | 88.24 | 100 | 86.21 | 92.59 | 97.93 | 66.12 | 0.15 | 86.26 | 1.98 |
| Solarize 128 | 97.06 | 100 | 96.55 | 98.25 | 100 | 68.59 | 0.17 | 89.27 | 7.79 |
| Solarize 192 | 97.06 | 96.67 | 100 | 98.31 | 99.31 | 64.42 | 0.15 | 85.1 | 11.96 |
| CLAHE (2, (8, 8)) | 97.06 | 100 | 96.55 | 98.25 | 100 | 65.74 | 0.15 | 85.58 | 11.48 |
| CLAHE (3, (8, 8)) | 91.18 | 100 | 89.66 | 94.55 | 97.24 | 66.57 | 0.15 | 85.04 | 6.14 |
| CLAHE (3, (16, 16)) | 94.12 | 100 | 93.1 | 96.43 | 98.62 | 67.40 | 0.15 | 84.4 | 9.72 |
| CLAHE (3, (6, 12)) | 100 | 100 | 100 | 100 | 100 | 66.08 | 0.18 | 83.78 | 16.22 |
| Gamma -1.5 | 94.12 | 96.55 | 96.55 | 96.55 | 99.31 | 66.55 | 0.15 | 88.65 | 5.47 |
| Gamma 1.5 | 91.18 | 96.43 | 93.1 | 94.74 | 97.93 | 70.24 | 0.2 | 83.28 | 7.9 |
| Gamma 1.25 | 91.18 | 96.43 | 93.1 | 94.74 | 97.93 | 66.56 | 0.15 | 83.85 | 7.33 |
| Gamma 1.75 | 94.12 | 96.55 | 96.55 | 96.55 | 97.93 | 66.47 | 0.15 | 82.09 | 12.03 |
| Gamma 2 | 88.24 | 96.3 | 89.66 | 92.86 | 97.93 | 65.83 | 0.14 | 86.22 | 2.02 |
| Gamma 5 | 94.12 | 96.55 | 96.55 | 96.55 | 99.31 | 66.94 | 0.14 | 87.49 | 6.63 |

4.2.1. Original Images

Using the original, unenhanced thermogram images, the model achieved a stable performance with an accuracy of 88.24% and a precision of 100%. However, the fold mean accuracy was slightly lower at 83.23%, indicating moderate generalization. The unenhanced image quality retained all visual noise, which likely contributed to higher variance in model performance across folds.

4.2.2. Posterization

Applying the Posterize technique with 8 colors simplified the color palette and enhanced recall to 100% by considering distinct temperature regions. Nevertheless, the significant reduction in color detail hindered generalization, as evidenced by a fold mean accuracy of 80.80% and a high "Diff" value of 13.32%. Increasing the color levels to 64 struck a balance between simplicity and detail, resulting in a higher accuracy of 91.18% and improved consistency (fold mean accuracy of 83.78%). Using 512 colors produced results similar to the original images, with an accuracy of 88.24% and a fold mean of 85.03%, indicating that increasing colors beyond a certain threshold adds minimum value while maintaining consistent generalization.

4.2.3. Solarization

The Solarize technique at a threshold of 64 introduced high contrast but retained subtle details, achieving consistent performance with a fold mean accuracy of 86.26% and the lowest "Diff" value of 1.98%. Solarize at threshold 128 yielded the best overall results, with accuracy, precision, and recall all at 97.06%, and the highest fold mean accuracy of 89.27%. This threshold optimally balances contrast enhancement and detail visibility. Conversely, Solarize at threshold 192 accentuated extreme contrasts, boosting the F1-score to 98.31% but increasing fold variation (fold mean accuracy of 85.10% and "Diff" of 11.96%), indicating sensitivity to extreme values and reduced generalization.

4.2.4. CLAHE

Applying CLAHE with a clip limit of 2 and a tile grid size of 8×8 demonstrated consistency across metrics, achieving 97.06% accuracy and perfect scores in precision, recall, and AUC, with a stable fold mean accuracy of 85.58%. Increasing the clip limit to 3 with the same grid size resulted in slightly lower performance (91.18% accuracy) but maintained reliability. The configuration with a clip limit of 3 and a grid size of 6×12 achieved perfect scores across all metrics (100%) but struggled to generalize across folds, indicated by the highest "Diff" value of 16.22%. Using a grid size of 16×16 with a clip limit of 3 balanced enhanced detail and generalization, achieving a fold mean accuracy of 84.40%.

4.2.5. Gamma Adjustment

Gamma Adjustment at -1.5 delivered strong results with an accuracy of 94.12% and a fold mean accuracy of 88.65%, effectively highlighting low-temperature regions despite the extreme visual distortion. This suggests that emphasizing certain thermal anomalies can enhance classification performance. A gamma value of 1.5 provided a balanced approach, achieving consistent results with 91.18% accuracy and a fold mean accuracy of 83.28%. For higher gamma values (2 and 5), images become overly bright, impacting performance. Gamma 2 showed moderate generalization (fold mean accuracy of 86.22%), while gamma 5 maintained strong recall and AUC but exhibited higher "Diff" values (6.63%), indicating less consistent generalization.

4.2.6. Overall Findings

The implementation of image enhancement techniques significantly improved model performance, as evidenced by the enhanced evaluation metrics. Enhanced images contributed to higher accuracy, precision, recall, and F1-score, and also demonstrated faster training and inference times compared to using the original, unenhanced data. These results highlight the dual benefits of image enhancement in boosting both the effectiveness of the model and computational efficiency.

Among the enhancement methods evaluated, Solarize at threshold 128 emerged as the most effective, balancing high performance with strong generalization. CLAHE and Gamma Adjustment also demonstrated considerable potential, offering unique advantages depending on specific task requirements. This analysis underscores the importance of balancing enhancement intensity with fold

consistency to achieve optimal results. Selecting appropriate enhancement parameters is crucial, as it directly influences the model's ability to generalize and accurately classify diabetic conditions. Future work may explore combining these enhancement techniques to capitalize on their complementary strengths, potentially leading to further improvements in diagnostic accuracy.

Figure 7 illustrates the training history plots for models using the Solarize enhancement technique at thresholds of 64, 128, and 192 over 200 epochs.

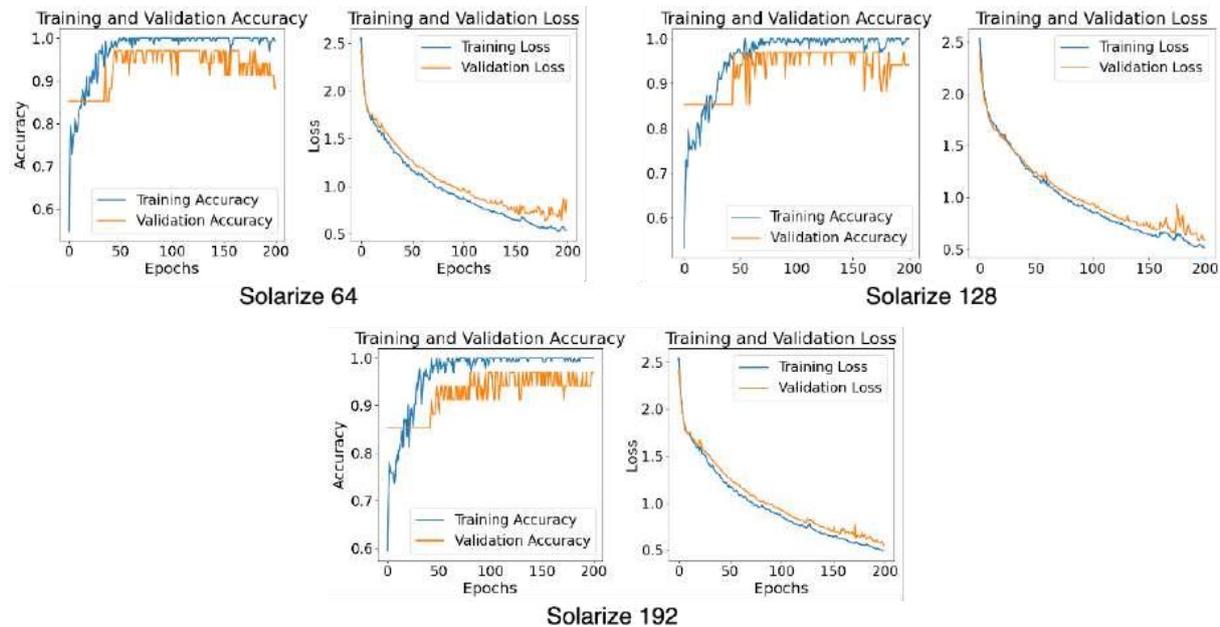


Figure 7. History plot of solarize.

The training history plots show that the Solarize threshold of 192 offers the most balanced performance, combining high accuracy, stable loss metrics, and minimal fluctuations. This indicates that threshold 192 effectively enhances thermogram images by optimally inverting pixel intensities to highlight critical temperature contrasts without adding excessive noise or obscuring important details. Models trained with this threshold generalize well to validate data, maintaining consistent performance across epochs.

In contrast, the lower threshold of 64, while quickly achieving high accuracy, exhibits greater volatility in both accuracy and loss—possibly due to overemphasis on minor temperature variations and increased noise sensitivity. The threshold of 128 provides stable training loss metrics but shows fluctuations in validation loss.

4.3. Modeling on Bias Sets

An additional evaluation was conducted using a bias-prone subset of data. This biased dataset included thermograms lacking distinct CG or DM characteristics, thermograms that visually exhibited CG traits but specific to DM cases, and thermograms with asymmetrical patterns between the right and left feet. Selection of these biased samples was based on the author's judgment, guided by medical literature [7], [21-25]. The subset consisted of 33 biased samples used as test data, while the remaining 134 samples were used for training. The model was trained using Solarize, identified earlier as the most effective image enhancement technique. This setup allowed us to analyze the model's ability to handle complex and ambiguous thermographic patterns while leveraging the performance improvements provided by Solarize.

Table 7 is the results from the bias dataset offering critical insights into the model's performance under challenging conditions. Using Solarize 64, the model achieved the highest accuracy at 69.7% with an F1-score of 79.17%. Although recall was relatively low at 65.52%, precision reached an ideal 100%. This indicates the model was highly selective and confident in its positive predictions but struggled to generalize across the entire test set. The imbalance suggests that

the biased dataset, with thermograms showing ambiguous or conflicting features, presented patterns that differed significantly from the training data, limiting the model's adaptability.

Table 7. Results of modeling for all image enhancement.

| Image enhancement | Evaluation metrics (%) | | | | | Time (s) | |
|-------------------|------------------------|-----------|--------|----------|-------|----------|-----------|
| | Accuracy | Precision | Recall | F1-score | AUC | Training | Inference |
| Solarize 64 | 69.7 | 100 | 65.52 | 79.17 | 67.24 | 65.04 | 0.004 |
| Solarize 128 | 48.48 | 100 | 41.38 | 58.54 | 52.59 | 65.89 | 0.004 |
| Solarize 192 | 45.45 | 92.31 | 41.38 | 57.14 | 70.69 | 65.72 | 0.005 |

In contrast, Solarize 128 and 192 showed even lower accuracies of 48.48% and 45.45%, respectively. The sharp drop in F1-score and recall for Solarize 128, along with its AUC of only 52.59%, highlighting the model's inability to effectively distinguish between CG and DM in scenarios with extreme or conflicting patterns. Solarize 192, despite a slightly higher AUC of 70.69%, also suffered from poor generalization, reflected in its low recall of 41.38%. These findings emphasize the trade-off with Solarize thresholds: while Solarize 64 effectively captures broader temperature contrasts, higher thresholds like 128 and 192 may introduce distortions that obscure critical features in highly biased data.

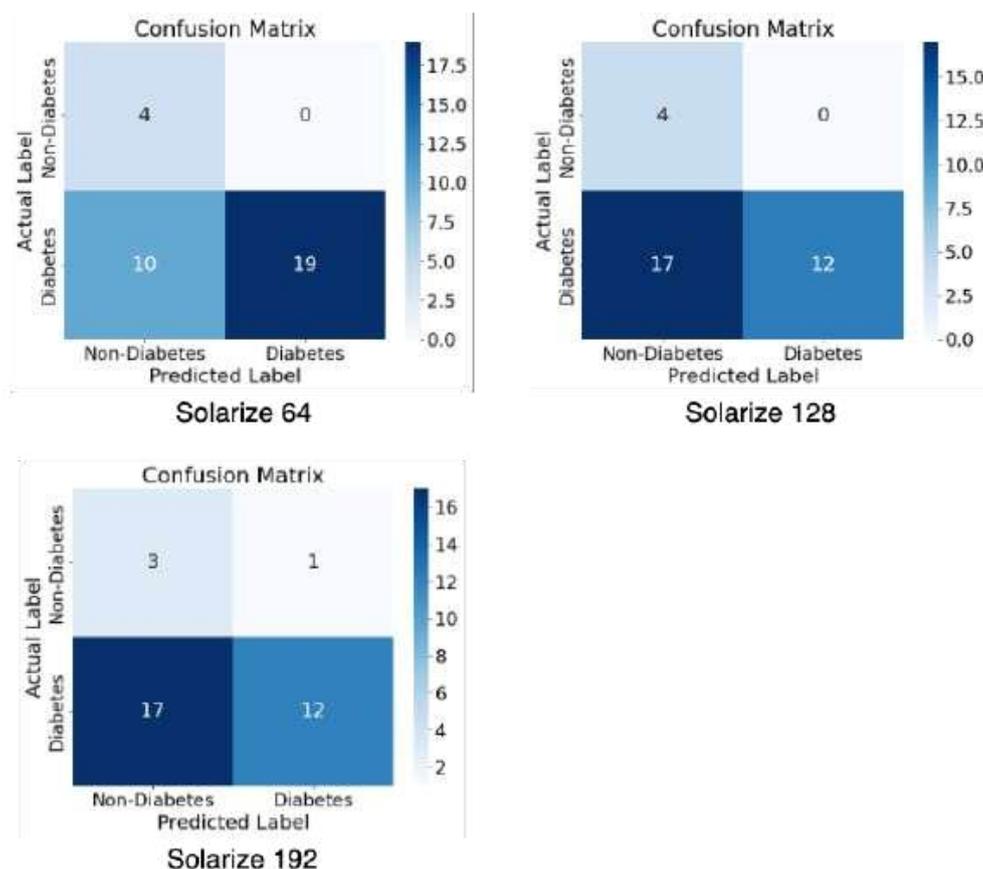


Figure 8. Confusion Matrices for bias sets.

An intriguing observation is the significantly lower inference time across all Solarize thresholds for the bias dataset, averaging around 0.004 – 0.005 seconds. This sharp reduction compared to the standard dataset suggests the model processes the bias data more quickly due to reduced complexity or variability in the patterns encountered. However, this efficiency comes at the cost of performance, as the lower inference time correlates with declines in recall and overall

accuracy. This trade-off raises concerns about the model's robustness and highlights the need for further refinement, especially in handling edge cases that deviate from the training distribution.

Despite the low accuracy, the model exhibits exceptionally high precision across all Solarize thresholds, particularly with Solarize 64 and 128, where precision reaches 100%. This occurs because the model makes virtually no false positive predictions for the positive class (DM). As shown in Figure 9, the false positive count is zero for both Solarize 64 and 128, indicating that every positive prediction aligns with the ground truth labels.

Despite the low accuracy across all Solarize thresholds, the model exhibits exceptionally high precision, particularly with Solarize 64 and Solarize 128, where precision reaches a perfect 100%. This phenomenon arises because the model makes virtually no false positive predictions for the positive class (DM). As reflected Fig. 9, the false positive (FP) count is zero for both Solarize 64 and Solarize 128. This indicates that every positive prediction made by the model is indeed correct according to the ground truth labels.

However, the low recall scores reveal a significant limitation in the model's capabilities to capture all positive cases. For example, under Solarize 64, of the 29 instances that truly belong to the DM class, the model only identifies 19, leaving 10 cases undetected. This demonstrates a strong bias toward precision over sensitivity, which can be particularly problematic in medical diagnostics where missing true positives (false negatives) can have severe consequences.

5. CONCLUSION

This research evaluated various image enhancement techniques to improve early diabetic foot ulcer detection using thermogram images. We used a comprehensive dataset from IEEE Dataport by Hernandez-Contreras et al. [7], which included thermograms from 122 diabetic and 45 non-diabetic patients, along with detailed foot temperature data. Rigorous preprocessing ensured data consistency, including normalization using the Standard Scaler for tabular data and pixel normalization for images.

We regularly applied four image enhancement methods: CLAHE, Posterize, Solarize, and Gamma Adjustment. These techniques aim to enhance temperature distribution patterns critical for early ulcer detection. The model, which integrated submodules for both image and tabular data, was trained using an 80:20 split for training and testing. Performance was assessed using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

Among the techniques, Solarize at a threshold of 128 was most effective. It achieved 97.06% accuracy at 200 epochs, with a mean fold accuracy of 89.27% and a manageable error of 7.79%, showing strong generalization. Solarize also resisted overfitting, outperforming other methods like CLAHE, which, despite high precision in some setups, showed overfitting and inconsistent folds. Posterize and Gamma Adjustment showed potential but were less consistent. Notably, Gamma -1.5 offered surprising generalization despite its extreme visual transformations.

We also tested the model on a highly biased data subset with ambiguous thermograms, selected based on medical literature. While accuracy dropped (e.g., 69.7% for Solarize 64), the model achieved near-perfect precision, avoiding false positives even in challenging cases. This underscores the need for optimization when handling complex, biased datasets.

Future work could combine Solarize with other enhancements such as CLAHE or Gamma Adjustment to further improve feature visibility and model performance. Adopting advanced architectures such as Kolmogorov-Arnold Networks or Vision Transformers might capture subtle temperature variations more effectively. Additionally, expanding the dataset with patient-specific metrics—such as blood glucose levels, age, BMI, and diabetes duration—could increase accuracy and clinical relevance. These steps aim to develop a more robust, precise, and clinically useful diagnostic system for detecting diabetic foot ulcers.

REFERENCES

- [1] Mahajan, S., Sarangi, P. K., Sahoo, A. K., & Rohra, M. (2023). Diabetes mellitus prediction using supervised machine learning techniques. *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, 587–592.

- [2] Ouyang, S., Zhang, X., Li, H., Tang, X., Ning, X., Li, R., Huang, F., Liu, B. and Fang, Y., & Liang, Y. (2023). Cataract, glaucoma, and diabetic retinopathy are independent risk factors affecting falls in the older adult with eye diseases. *Geriatric Nursing*, **53**, 170–174.
- [3] Bettencourt-Silva, R., Aguiar, B., Sá-Araújo, V., Barreira, R., Guedes, V., Ribeiro, M. J. M., Carvalho, D., Östlundh, L., & Paulo, M. S. (2019). Diabetes-related symptoms, acute complications and management of diabetes mellitus of patients who are receiving palliative care: a protocol for a systematic review. *BMJ open*, *9*(6), e028604.
- [4] Zhao, N., Yu, L., Fu, X., Dai, W., Han, H., Bai, J., Xu, J., Hu, J., & Zhou, Q. (2024). Application of a Diabetic Foot Smart APP in the measurement of diabetic foot ulcers. *International Journal of Orthopaedic and Trauma Nursing*, **54**, 101095.
- [5] Araújo, A. L. D., Negreiros, F. D. D. S., Florêncio, R. S., Oliveira, S. K. P. D., Silva, A. R. V. D., & Moreira, T. M. M. (2022). Effect of thermometry on the prevention of diabetic foot ulcers. *Revista Latino-Americana de Enfermagem*, **30**, e3567.
- [6] Khandakar, A., Chowdhury, M. E., Reaz, M. B. I., Ali, S. H. M., Abbas, T. O., Alam, T., Ayari, M. A., Mahbub, Z. B., Habib, R., Rahman, T., Tahir, A. M., Bakar, A. A. A., & Malik, R. A. (2022). Thermal change index-based diabetic foot thermogram image classification using machine learning techniques. *Sensors*, **22**(5), 1793.
- [7] Hernandez-Contreras, D. A., Peregrina-Barreto, H., de Jesus Rangel-Magdaleno, J., & Renero-Carrillo, F. J. (2019). Plantar thermogram database for the study of diabetic foot complications. *IEEE Access*, *7*, 161296–161307.
- [8] Yang, S., Tian, Y., Zheng, M., Du, Y., Chen, H., Song, F., Gao, X., & Li, L. (2021). A review of image enhancement technology research. *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, 715–720.
- [9] Anand, S. & Roshan, R. K. (2023). Chest X ray image enhancement using deep contrast diffusion learning. *Optik*, **279**, 170751.
- [10] Kuruba, C. & Gopalan, N. P. (2023). Robust blood vessel detection with image enhancement using relative intensity order transformation and deep learning. *Biomedical Signal Processing and Control*, **86**, 105195.
- [11] Pei, X., hong Zhao, Y., Chen, L., Guo, Q., Duan, Z., Pan, Y., & Hou, H. (2023). Robustness of machine learning to color, size change, normalization, and image enhancement on micrograph datasets with large sample differences. *Materials & Design*, **232**, 112086.
- [12] Vardasca, R. (2019). Diabetic foot monitoring using dynamic thermography and AI classifiers. *ISLA - Atas de Conferência Nacional*, 027.
- [13] Adam, M., Ng, E. Y., Oh, S. L., Heng, M. L., Hagiwara, Y., Tan, J. H., Tong, J. W., & Acharya, U. R. (2018). Automated characterization of diabetic foot using nonlinear features extracted from thermograms. *Infrared Physics and Technology*, **89**, 325–337.
- [14] Balasenthilkumaran, N. V., Ram S, B., Gorti, S., Rajagopal, S., & Soangra, R. (2022). Design and comparison of machine learning-based computer-aided diagnostic techniques to aid diagnosis of diabetes and detection of ulcer-prone regions in the feet using thermograms. *Research on Biomedical Engineering*, **38**(3), 781–795.
- [15] Jayapal, S., Murugesan, N. B., & Mohan, S. (2022). Detection of Diabetic Foot Using Statistical Features. *Diabetic Foot-Recent Advances*.
- [16] Mounika, N. & Thirunavukkarasu, U. (2023). Classification of healthy and diabetic mellitus individuals by extracted textural features from left plantar thermograms and classifying using SVM and NB classifiers. *AIP Conference Proceedings*, **2822**(1).
- [17] Filipe, V., Teixeira, P., & Teixeira, A. (2020). A clustering approach for prediction of diabetic foot using thermal images. *International Conference on Computational Science and Its Applications*, 620–631).
- [18] Gamara, R. P. C., Loresco, P. J. M., & Bandala, A. A. (2022). Medical chest x-ray image enhancement based on CLAHE and wiener filter for deep learning data preprocessing. *2022 IEEE 14th international conference on humanoid, nanotechnology, information technology, communication and control, environment, and management (HNICEM)*, 1–6.
- [19] Deepak, K. V. & Bharanidharan, R. (2023). Osteosarcoma detection in histopathology images using ensemble machine learning techniques. *Biomedical Signal Processing and Control*, **86**, 105281.

- [20] Tasci, E., Uluturk, C., & Ugur, A. (2021). A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection. *Neural Computing and Applications*, **33**(22), 15541–15555.
- [21] Mori, T., Nagase, T., Takehara, K., Oe, M., Ohashi, Y., Amemiya, A., Noguchi, H., Ueki, K., Kadowaki, T., & Sanada, H. (2013). Morphological pattern classification system for plantar thermography of patients with diabetes. *Journal of Diabetes Science and Technology*, **7**(5), 1102–1112.
- [22] Page, M. J., Sterne, J. A., Higgins, J. P., & Egger, M. (2021). Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. *Research Synthesis Methods*, **12**(2), 248–259.
- [23] Hiebl, M. R. (2023). Sample selection in systematic literature reviews of management research. *Organizational Research Methods*, **26**(2), 229–261.
- [24] Althubaiti, A. (2023). Sample size determination: A practical guide for health researchers. *Journal of General and Family Medicine*, **24**(2), 72–78.
- [25] Afonso, J., Ramirez-Campillo, R., Clemente, F. M., Büttner, F. C., & Andrade, R. (2024). The perils of misinterpreting and misusing “publication bias” in meta-analyses: an education review on funnel plot-based methods. *Sports Medicine*, **54**(2), 257–269.