

Received May 20, 2022, accepted May 30, 2022, date of publication June 3, 2022, date of current version June 9, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3180036

Detection of Diabetes Mellitus With Deep Learning and Data Augmentation Techniques on Foot Thermography

ANDRÉS ANAYA-ISAZA^{1,2} AND MATHA ZEQUERA-DIAZ¹, (Member, IEEE)

¹Electronics Department, BASPI-FootLab, Pontificia Universidad Javeriana, Bogota 410001, Colombia

²INDIGO Research, Bogota 410010, Colombia

Corresponding author: Andrés Anaya-Isaza (anayaandresj@javeriana.edu.co)

This work was supported in part by the Research Division from BASPI FOOTLAB (<https://ingenieria.javeriana.edu.co/investigacion/laboratorio-footlab/>), and in part by INDIGO Technologies (<https://indigo.tech/>).

ABSTRACT Diabetes mellitus (DM) is a metabolic disorder characterized by increased blood glucose. The pathology can manifest itself with different conditions, including neuropathy, the main consequence of diabetic disease. Statistics show worrying figures worldwide, diagnosed an estimated 1.6 million people with DM by 2025. In this sense, alternative and automated methods are necessary to detect DM, allowing it to take the pertinent measures in its treatment and avoid critical complications, such as the diabetic foot. On the other hand, foot thermography is a promising tool that allows visualization of thermal patterns, patterns that are altered as a consequence of shear and friction associated with lower limb neuropathy. Based on these considerations, we explored different strategies to detect patients with DM from foot thermography in this research. Initially, the study focused on finding a classification index like Thermal Change Index (TCI). Subsequently, we used the deep convolutional neural networks paradigm, implementing 12 different data augmentation methods, of which four are conventional, and 8 are newly proposed methods. The results showed that the proposed and the conventional methods increased the network's performance, where a 100% detection was achieved by weighting the DM probability percentages for both images of the feet. Finally, it was also possible to demonstrate the importance of transfer learning, which does not depend on the type of database, but on the data corpus with which the transfer was trained.

INDEX TERMS Artificial intelligence, biomedical imaging, computational and artificial intelligence, diabetes, machine learning, medical diagnostic imaging.

I. INTRODUCTION

Diabetes Mellitus (DM) is a metabolic disorder characterized by increased blood glucose, known as hyperglycemia [1]. The pathogenesis of DM is due to immune disorders or insulin resistance, generated by genetic predisposition or environmental factors [2], [3]. In general, risk factors associated with DM are first-degree relatives with the pathology or a high Body Mass Index (BMI). In fact, high BMI has increased the number of cases worldwide due to poor eating habits and excessive sugar consumption [4], [5]. The impact is seen in alarming statistics worldwide. For example, World Health Organization (WHO) reported DM as the leading cause of death in 2019, with approximately 1.5 million cases [6].

The associate editor coordinating the review of this manuscript and approving it for publication was M. Shamim Kaiser¹.

Projections indicate that this number could increase to approximately 1.6 million by 2025 [7]. Similarly, the International Diabetes Federation (IDF) reported 537 million adults (20-79 years) with diabetes in 2021, i.e., one in ten adults are living with diabetes, where more than half are undiagnosed. Estimates project a global increase to 643 million by 2030 and 784 million by 2045. In addition, the IDF estimated approximately 6.7 million deaths by 2021 [8].

DM manifests as a group of metabolic disorders that can range from heart disease, stroke, renal failure, blindness, ischemia, and peripheral vascular disease, among the most common conditions [9]. Peripheral vascular disorders generate the well-known diabetic foot, being this also the main neuropathic alteration [10]. The disorder generates infections, ulcerations, or destruction of deep tissue in the lower limb, limiting the patient's mobility [11]. In this order of ideas, the

pathology requires early and continuous medical attention to prevent complications such as lower limb amputation [12].

On the other hand, body temperature change is a widely used strategy to identify abnormalities in the medical field [13]. The research addresses different applications in the medical field, and diabetes is no exception to this trend. For example, different studies are currently based on foot thermography and even on the detection of type 2 DM from facial or tongue thermography, as explored by Thirunavukkarasu *et al.* in 2020 [14], [15]. Thermography is presented as a novel method for the exploration of different conditions [16], [17]; even such images have presented a significant growth in recent decades due to their noninvasive nature, being used even in face detection [18]. Additionally, thermographic images can be integrated with different Artificial Intelligence (AI) strategies due to their digital character. Advances have not been long in coming, and there are already several research studies aimed at diagnosing DM or the detection of conditions that are a consequence of it, such as neuropathic complications or diabetic foot [19]–[24]. The growing interest has even motivated the development of low-cost proprietary systems to register such thermographic images, as has been done by Bayareh [25] and Sruthi *et al.* [26].

At the same time, the topic has been approached from different perspectives, covering topics ranging from image processing to state-of-the-art deep learning networks. For example, Christy Evangeline *et al.* suggest a method of asymmetry analysis between angiosomes for the exclusive characterization of subjects with diabetes [27]. Filipe *et al.* propose a new coefficient like the Thermal Change Index (TCI). The method consists of parceling the foot into different regions based on the clustering algorithm applied to the intensity levels of the image (temperature). The average temperature is estimated from the regions, and they calculate the coefficient called Classification Temperature Threshold (CTT) [28]. Prabhu *et al.* developed an automatic system for segmentation of the plantar foot region in thermographic images, as it is an inherent need for feature extraction in such images. The development was done using the C means algorithm [29]. On the other hand, Carlos Padierna *et al.* extract 12 features from the top of the foot to train a Support Vector Machine (SVM) [30]. Similarly, Adam *et al.* also rely on a support vector machine for classification. The model uses Gray Level Co-occurrence Matrix (GLCM), Hu's invariant moment, Local Binary Pattern (LBP), Laws Texture Energy (LTE), and entropy, extracted from the decomposed form of the Discrete Wavelet Transform (DWT) and Higher-Order Spectrum (HOS). The development achieves an accuracy of over 89%. However, the authors point out that the work was based on a reduced database; moreover, the system is semi-automated, as it requires feature extraction [31]. In the same year, Adam *et al.* present new research on diabetic foot classification. The authors report higher values than in their previous research.

Similarly, the research is based on feature extraction but through decomposition into coefficients using Double

Density-Double Tree-complex Wavelet Transform (DD-DT-CWT). The features were used to train different machine learning algorithms, where they managed to achieve the best results through four LSDA features from bilateral foot thermograms with k nearest neighbors (kNN) classifier [32]. In turn, Josephine Selle *et al.* follow the same classification path through SVM. The implementation is based on 15 features, of which 14 are generated from the GLCM, being the features in mention: autocorrelation, contrast, correlation, dissimilarity, energy, entropy, homogeneity, max probability, sum of squares, sum average, sum entropy, sum variance, difference entropy, and difference variance. In this particular case, the model performance reaches an accuracy of 96.42%; however, the authors highlight that the number of the database is limited, the features are not optimized, and the classification is not performed on higher-order machine learning algorithms [33]. Reddy *et al.* implement four machine learning algorithms to diagnose foot ulcers in diabetic patients in a more comprehensive approach. The research is based on 22 recorded characteristics of each patient. The 22 attributes range from gender, age, body mass index, and qualitative indices such as smoking habit [34]. Nag *et al.* perform the classification with five previous steps, starting from acquisition, data preprocessing, analysis, segmentation of regions of interest, and feature extraction. The latter were obtained with first-order statistical features, textures based on GLCM, and features based on Wavelets. Finally, they perform the classification using three machine learning methods: SVM, kNN, and Decision Tree (DT), reaching an accuracy of 97.78% through the latter method [35].

Contrary to the previous developments, Thirunavukkarasu *et al.* perform a study base on standing thermographs' acquisition. In the process, they measure the temperature of the toe, metatarsal 1, 3, 5, instep, and heel, achieving an accuracy of over 81% through SVM [36]. For their part, Vardasca *et al.* use dynamic thermography to classify the diabetic foot. The proposed study was based on three machine learning algorithms using 12 temperature landmarks for subject discrimination. The development achieved 81% accuracy by implementing the kNN algorithm [37]. Finally, Cruz-Vega *et al.* use an artificial neural network and a support vector machine from the deep learning paradigm. The development is performed by obtaining several features after preprocessing. Additionally, they include the AlexNet deep learning network with the raw data, achieving the best classification performance in cross-validation [38]. A year later, Cruz-Vega *et al.* introduced a new convolutional neural network like AlexNet, which they call DFTNet. Additionally, the authors explore the multilayer perceptron and the support vector machine. The first model is trained with the raw data, while the third and second are trained with features extracted from the regions of interest [39], achieving an accuracy close to 95% [39]. Similarly, Khandakar *et al.* propose diabetic foot detection from two perspectives. The first one is from detection with different deep learning networks, and the second one uses conventional machine learning methods

based on the subjects' characteristics, reaching an accuracy close to 96% [40].

Table 1 summarizes the related works described above, where the best metrics obtained by the different authors with the proposed methods are indicated.

The previous review was based on the most recent and relevant articles in the study of DM diagnosis in thermographic images. The systematic review shows significant shortcomings in this area of interest. While it is evident that artificial intelligence has reached DM diagnosis, approaches to deep learning are scarce. On the other hand, despite the small corpus of data, approaches toward data augmentation or transfer learning are also not addressed in depth. Moreover, from previous reviews, one can confirm critical limitations such as small data volume, data variability, and low data quality [42].

At the same time, convolutional neural networks have had exponential growth in different research areas, and every day new architectures, models, and variations emerge that make them more and more efficient [43], [44].

Likewise, DL strongly impacts the medical area, addressing issues ranging from diagnosis to the prediction of possible pathologies [45]–[47]. Research is focused on networks with greater generalization capabilities, increasing the performance of models while trying to outperform human performance [48], [49]. The impact is seen in novel architectures such as ResNet [50], Inception [51], DenseNet [52] and transformers [53] among the most relevant developments. In addition, performance enhancement has been addressed with practical solutions such as regularization [54], batch normalization [55], transfer learning [56], and data augmentation. While the strategies are promising, data augmentation has been more widely embraced because it largely corrects the overfitting problem suffered by DL networks with little training data [57]. Traditionally, data augmentation consists of generating new images starting from reference images that retain their attributes, i.e., the new images must have the same labels as the original images [58]. The methods can address linear transformations, color space changes, and even intuitive transformations such as flipping images [59]. In other more inventive and current approaches, Szlobodnyik *et al.* [60] put forward a new method based on guided deep interpolation with an auxiliary linear self-expressive layer. In turn, Raj *et al.* [61] suggest a nonlinear image blending technique, which challenges the classical notion of data augmentation. In this sense, exploring new data augmentation methods is an inherent need for the constant demand of large data corpora for training DL networks.

On the other hand, most DL applications are focused on conventional images, i.e., images of people, objects, or animals [62]. The trend was expected since high-quality digital cameras are now available, even at low cost and accessible to most users. Consequently, there are now genuinely novel developments that have not been exploited in medical imaging or are not known to be effective cases, such as foot thermography. For example, it has recently been shown that the

ResNet network outperforms human performance in classification tasks and is even part of the top 5 networks [63]. Moreover, this network introduces residual connections to stabilize parameter training, allowing for deepening the model and, in turn, tackling increasingly complex tasks [50].

In this sense, we propose a framework for exploring different approaches to the detection of subjects with diabetes mellitus (DM). Initially, we propose the study of a new coefficient of characterization of subjects based on characteristics or descriptors such as age, general temperature, temperature by angiosomes, and rate of thermal change. Subsequently, due to the limited corpus of data, we propose the implementation of one of the most novel neural networks implemented with 12 data augmentation methods, of which 8 are newly proposed methods. Additionally, as a comparative analysis, we study the behavior of the network under three different training conditions. First, we train the network from scratch. Secondly, we use transfer learning from the conventional ImageNet image database, and, finally, we perform the same process but from a thermographic database, hoping that the thermal patterns of the data will allow us to achieve a higher accuracy of the model.

- Therefore, this work shows the following relevant contributions.
- A coefficient for the characterization of subjects with DM was sought.
- One of the state-of-the-art convolutional neural networks was implemented for accurate detection of DM subjects.
- Eight new data augmentation methods were introduced to overcome the limitation of a low amount of data.
- A comparative analysis of the training conditions of the convolutional models was performed.

II. MATERIALS AND METHODS

A. DATASET

The investigation was performed using the public database “Plantar Thermogram Database for the Study of Diabetic Foot Complications” [64]. The database has 122 patients diagnosed with DM and 45 subjects in the Control Group (CG). Each subject has thermograms of both feet stored in .csv files in a temperature array. Temperatures range from 15.98 to 37.28 degrees Celsius, except the bottom, which was stored with zero after a manual segmentation process. The images have the partitions of the four angiosomes; Medial Plantar Artery (MPA), Lateral Plantar Artery (LPA), Medial Calcaneal Artery (MCA), and Lateral Calcaneal Artery (LCA) (See Figure 1). Additionally, the set includes the description of gender, age, TCI, height, weight, and BMI. It should be noted that the last three descriptors are not for all subjects.

B. DATA PREPROCESSING

The .csv files were converted to single-channel images in the 'float32' format. The size was set to 199 × 88, padding with zeros the edges of all images that did not meet that size.

TABLE 1. Compilation of the results obtained by the different authors in research associated with the detection of the diabetic foot through artificial intelligence.

Author	Journal	Year	Best performing algorithm	Number of data	Evaluation metrics												
					Accuracy	Sensitivity	Specificity	F1 score	Precision	MCC	AUC	ZOL	TS/CSI	FOR	FDR	Dice	RMSD
Khandakar [41]	Others	2022	MLP	167	91.0	95.0	-	95.0	96.0	-	-	-	-	-	-	-	-
Khandakar [40]	Elsevier	2021	RF	167	96.7	96.7	94.6	96.7	96.7	-	-	-	-	-	-	-	-
Khandakar* [40]	Elsevier	2021	MobileNetv2	167	95.8	95.8	94.2	95.8	95.8	-	-	-	-	-	-	-	-
Josephine Selle [33]	IEEE	2021	SVM	50	96.4	97.5	94.4	-	-	-	-	-	-	-	-	-	-
Nag [35]	Others	2021	TD	167	97.8	97.5	97.5	-	-	-	-	-	-	-	-	-	-
Prabhu [29]	Others	2021	ACM	50	-	-	-	-	-	-	-	-	-	-	-	94.1	6.0
Carlos Padierna [30]	Elsevier	2020	SVM	43	92.6	91.8	93.6	-	-	93.3	-	-	-	-	-	-	-
Evangeline [27]	IEEE	2020	IP	39	-	-	-	-	-	-	-	-	-	-	-	-	-
Cruz-Vega* [39]	Others	2020	DFTNet	167	94.5	95.3	93.8	94.6	94.0	-	94.6	-	-	-	-	-	-
Thirunavukkarasu [36]	Others	2020	SVM	60	81.6	-	-	-	-	-	-	-	-	-	-	-	-
Filipe [28]	Others	2020	Clustering	167	85.3	85.3	60.0	-	78.4	-	-	-	-	-	-	-	-
Cruz-Vega [38]	IEEE	2019	AlexNet	-	99.9	-	-	-	-	-	-	-	-	-	-	-	-
Vardasca [37]	Others	2019	KNN	39	81.3	80.0	100.0	-	-	-	-	-	-	-	-	-	-
Adam [31]	Elsevier	2018	SVM	66	89.4	81.8	97.0	-	-	-	-	-	-	-	-	-	-
Adam [32]	Elsevier	2018	KNN	66	93.2	90.9	98.0	-	-	-	-	-	-	-	-	-	-
Reddy [34]	Others	2018	ELM	133	96.2	-	-	-	-	-	-	0.0	1.0	0.0	0.1	-	-

*Research with transfer learning, data augmentation or both

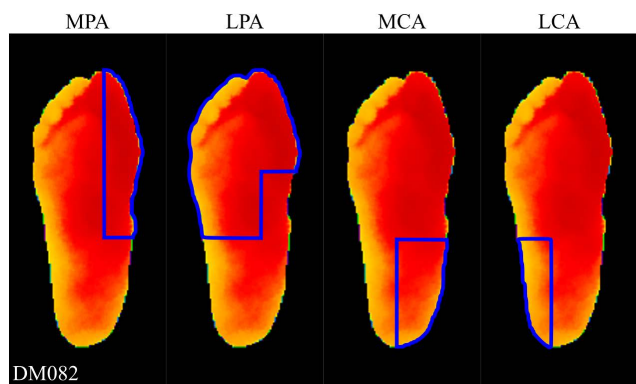


FIGURE 1. Diabetic foot with the plots corresponding to the 4 angiosomes.

Finally, the images were normalized with respect to the maximum and minimum of the data set, i.e., temperatures from 15.98 to 37.28 °C were taken as values from 0 to 1.

C. CONVOLUTIONAL NEURAL NETWORK

Artificial intelligence is one of the areas of computational sciences with many ramifications. AI ranges from the most straightforward systems composed of linear models to the most recent deep learning methods. One of the fundamental elements in AI is convolutional artificial neural networks or simply convolutional networks (See Appendix A). Networks are one of many bio-inspired systems based on the functioning of the central nervous system, more specifically, the brain.

Convolutional networks emulate the primary visual cortex to perform tasks such as detection, classification, or feature extraction to achieve a specific task [65].

This concept has led to the emergence of a large number of networks; in fact, there are currently several networks that differ from the first convolutional model developed by Lecun *et al.* in 1998 [66] or from the first convolutional DL model, known as AlexNet [65]. The various architectures and their depth allow them to perform increasingly complex tasks. However, the increase in depth has two drawbacks. First, the depth implies a more significant number of layers and, consequently, a larger number of training parameters. Consequently, a more significant number of training data is needed to fit many model parameters. In this sense, this is one of the reasons that motivated us to use data augmentation methods and include new strategies to address this problem. Secondly, convolutional layer augmentation limits the model’s training generated by gradient fading [66].

For this reason, in this research, we chose to use the ResNet50v2 network [50], which has an elegant solution to this drawback. The network is based on the concept of connection or residual mapping. In general, the connection creates parallel paths to the convolutional layer sequences, allowing smooth transmission of the gradient through the layers and preventing the gradient value from being zero. In addition, the connection forces the network to learn the residual mapping $f(x) - x$, being easier to train if the residual mapping is the identity function $f(x) = x$ [50], [67].

D. DATA AUGMENTATION

As mentioned above, neural networks require a large corpus of data to train all the model parameters. However, this is one of the main drawbacks in medical imaging-related research due to limited data sets. Fortunately, it has already been demonstrated that neural networks can be trained with synthetic images, allowing to reach the optimal values of the training parameters, and even avoiding overtraining the model [57], [58]. In this order of ideas, in the following, we present the four most used conventional methods in data augmentation, and we also propose eight new strategies based on image reduction methods.

1) FLIP

Among the most used conventional methods are geometric transformations [58]. The operations change the spatial position of pixels while maintaining the relationship between them, which can be subject to a mathematical model. One of the simplest methods is image flipping. The process consists of taking each pixel and assigning it a new inverted position concerning one of the two axes (in the case of two-dimensional images). The mathematical model of this process is governed by Equations (1) and (2) for horizontal and vertical flipping, respectively. The result is illustrated in Figure 2.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x_{max} \\ y \end{pmatrix} - \begin{pmatrix} x \\ 0 \end{pmatrix} \tag{1}$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y_{max} \end{pmatrix} - \begin{pmatrix} 0 \\ y \end{pmatrix} \tag{2}$$

where, x_{max} and y_{max} are the last positions reached by the image pixels on the respective axes.

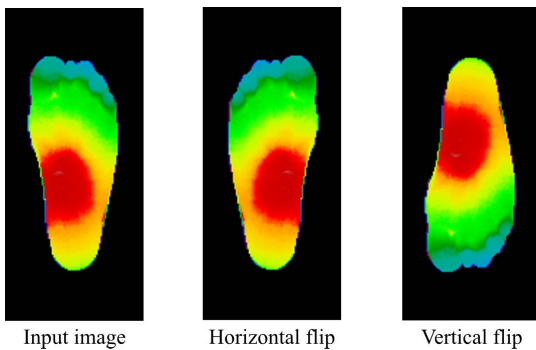


FIGURE 2. Data augmentation using geometric image flipping transformation.

2) COLOR SPACE CHANGE

Conventional images are stored as arrays of three channels with the exact dimensions. The channels represent each of the intensity levels that make up the RGB image, i.e., the intensities of red, green, and blue colors. However, the intensity ratios of these channels can be combined to generate new intensity levels, creating a new color space [68]. Since the

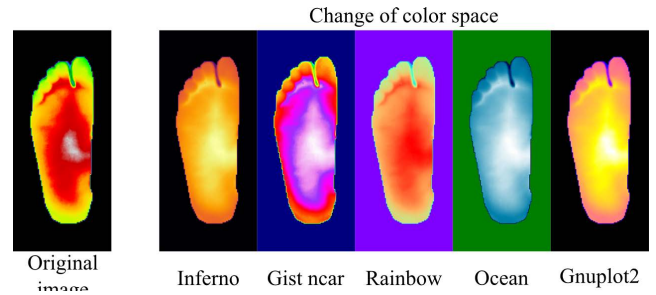


FIGURE 3. Examples of color space change in 5 different color maps.

channels can be combined in different ways, there are many color spaces, e.g., Figure 3 shows five examples.

For this particular case, we preferred to use our own color space, which was obtained with the equivalent transformations from RGB space to YCbCr, and from YCbCr to HSV space. However, the first transformation was applied by copying the gray channel to the remaining channels. That is, the transformation performed is the one expressed in equation (3) [69].

$$\begin{bmatrix} Y' \\ Cb' \\ Cr' \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} G \\ G \\ G \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \tag{3}$$

The transformation generated 3 channels from a single channel; however, we make the clarification that equation (3) is the equivalent transformation to YCbCr, but it is not such transformation because to obtain the Y, Cb and Cr channels, the operation must be applied to the RGB channels.

Subsequently, the channels obtained in the previous process were subjected to the equivalent RGB to HSV transformation. Mathematically, the process was performed with equations (4), (5) y (6), to generate the 3 definitive channels used in the training.

$$H = 60 * \begin{cases} \frac{Cb-Cr}{\Delta}, & \text{if } Y = \max(Y, Cb, Cr) \\ \frac{Cr-Y}{\Delta} + 2, & \text{if } Cb = \max(Y, Cb, Cr) \\ \frac{Y-Cr}{\Delta} + 4, & \text{if } Cr = \max(Y, Cb, Cr) \end{cases} \tag{4}$$

$$S = \frac{\Delta}{\max(Y, Cb, Cr)} \tag{5}$$

$$V = \max(Y, Cb, Cr) \tag{6}$$

where $\Delta = \max(Y, Cb, Cr) - \min(Y, Cb, Cr)$ [70].

3) LINEAR FILTERS

Another of the most used strategies is linear filters. Generally, filters are used to focus and blur the image, as illustrated in Figure 4. The method consists of sliding the filter through the whole image, obtaining new values for each filter position, generating the new image [71].

The new images were generated using a 3×3 Gaussian filter with a deviation of 90 pixels ($\sigma = 90$).

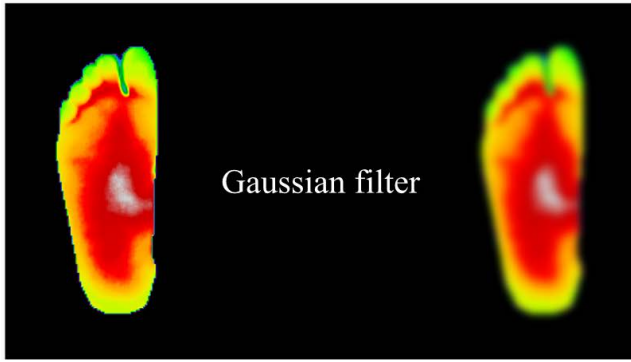


FIGURE 4. Example of linear transformation with a Gaussian filter.

4) PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a method used to represent a set of observations in an equivalent, represented in latent variables of lower dimensionality [72]. The method takes the observations and combines them linearly to create new components. These components are generated hierarchically, each retaining a percentage of the variability of the original data. The first component, z_1 , is created with the highest percentage of variability, while subsequent components take smaller percentages than the previous ones. In principle, the first principal component is expressed as shown in Equation (7) or (8) in the form of a dot product (For more details see appendix B).

$$z_1 = u_{11}x_1 + u_{12}x_2 + u_{13}x_3 + \dots + u_{1m}x_m \quad (7)$$

$$z_1 = U_1 \cdot X \quad (8)$$

5) KERNEL PCA

The kernel PCA is an extension of principal component analysis (PCA), developed to work with nonlinear data. Kernel PCA implements a function to project the data set into a higher dimensional feature space. That is, the original data set is projected into a linearly separable higher dimensional space and then reduced in dimensionality through the PCA method described in the previous section. In general, this method can be used with different filters, where the most common is the Gaussian filter, as shown in Equation (9).

$$x_3 = e^{-\left(\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right)} \quad (9)$$

6) INCREMENTAL PCA

Like the previous case, incremental PCA is another extension of conventional PCA but used for too large data sets. The method consists of a low-rank approximation for the data, independent of the number of input data samples.

7) FACTOR ANALYSIS

Factor analysis is one of the most popular multivariate analysis methods used to study the relationship between sets of

variables and project them into a space of lower dimensionality, known as factors. In particular, the mathematical model of factor analysis assumes that the original p variables are functions of m common factors (with $m < p$) plus a single factor ε . Mathematically, this can be expressed as shown in Equations (10) and (11).

$$\begin{aligned} x_1 &= u_{11}f_1 + u_{12}f_2 + u_{13}f_3 + \dots + u_{1m}f_m + \varepsilon_1 \\ x_2 &= u_{21}f_1 + u_{22}f_2 + u_{23}f_3 + \dots + u_{2m}f_m + \varepsilon_2 \\ &\vdots \\ x_p &= u_{p1}f_1 + u_{p2}f_2 + u_{p3}f_3 + \dots + u_{pm}f_m + \varepsilon_p \quad (10) \\ X &= UF + E \quad (11) \end{aligned}$$

where X is the vector of the original variables ($X \in R^p$), U is the factor matrix or matrix with factor loadings ($U \in R^{p \times m}$), F is the vector of common factors ($F \in R^m$), and E is the vector of unique factors ($E \in R^p$) [73], [74].

8) INDEPENDENT COMPONENT ANALYSIS

In statistics, independent component analysis is a widely used technique to find a linear representation of non-Gaussian data such that the components of the new representation are statistically independent. Assuming a data set represented with m variables ($X \in R^m$), one can estimate the independent components as a linear combination of the original variables, as expressed in Equation (12) or (13).

$$\begin{aligned} s_1 &= w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + \dots + w_{1m}x_m \\ s_2 &= w_{21}x_1 + w_{22}x_2 + w_{23}x_3 + \dots + w_{2m}x_m \\ &\vdots \\ s_m &= w_{m1}x_1 + w_{m2}x_2 + w_{m3}x_3 + \dots + w_{mm}x_m \quad (12) \\ S &= W \cdot X \quad (13) \end{aligned}$$

Starting from this assumption, the components or latent variables are determined by the matrix W , which is estimated by maximizing the non-Gaussianity of $W \cdot X$ [75], [76].

9) NON-NEGATIVE MATRIX FACTORIZATION

Non-negative matrix factorization, or also known as non-negative matrix approximation, is a dimension reduction technique that factors the matrix X observations into the approximation of two matrices. The method assumes that the three matrices do not have negative elements, allowing the resulting matrices to be more easily inspected. Mathematically, the observations matrix can be defined as an approximation of two matrices, as shown in Equation (14).

$$X \approx WH \quad (14)$$

The estimation of the two new matrices is calculated keeping as much information as possible and using the objective function expressed by Equation (15). The WH matrices are calculated in such a way that the objective function reaches

its smallest possible value [77], [78].

$$\begin{aligned}
 O_F &= \frac{1}{2} \|X - WH\|^2 + \alpha_w L_1 m \|W\|_1 + \dots \\
 &+ \alpha_h L_1 n \|H\|_1 + \dots \\
 &+ \frac{1}{2} \alpha_w (1 - L_1) m \|W\|_F^2 + \dots \\
 &+ \frac{1}{2} \alpha_h (1 - L_1) n \|W\|_F^2 \quad (15)
 \end{aligned}$$

here, α_w and α_h are the regularization coefficients associated with each matrix. L_1 is the regularization mixture parameter, m the number of variables or features in the original set, n the number of samples or observations, $\|\cdot\|_1$ is the l_1 norm and $\|\cdot\|_F$ is the Frobenius norm or also called as the Hibert-Schmidt norm, defined by Equation (16).

$$\|A\|_F = \left(\sum_{j=1}^m \sum_{k=1}^n |a_{j,k}|^2 \right)^{\frac{1}{2}} \quad (16)$$

here, $a_{j,k}$ is each of the elements of the $m \times n$ -dimensional matrix A [79].

10) DICTIONARY LEARNING

Dictionary learning is a way to find a better sparse mapping matrix by using training data. The technique is like non-negative matrix factorization since it is based on the minimization of an objective function. In particular, the method consists of finding a dictionary D (Matrix) and a representation of the data Y such that the objective function expressed in Equation (17) is minimized.

$$\min_{D \in \zeta, Y} \left\{ \frac{1}{2} \|X - DY\|_F^2 + \lambda \Psi(Y) \right\} \quad (17)$$

here $\|\cdot\|_F$ is the Frobenius norm, λ the coefficient associated with the regularization and Ψ is the sparsity inducing function, which can represent different regularization terms, such as l_0 and l_1 norms. ζ represents the dictionary constraint such that $\zeta \equiv \{D \in Y^{n \times m} : \|d_{-i}\|_2 < 1 \forall i = 1, \dots, m\}$. Where, $\|\cdot\|_2$ is the length of the vector and d_i the elements that compose the dictionary D [80], [81].

11) LATENT DIRICHLET ALLOCATION

Latent Dirichlet Allocation (LDA) is a generative statistical model used to represent a set of observations (M) by latent subjects. LDA is a three-level hierarchical model, with each level underlying a finite set of the respective lower level. Suppose that an observation m can be modeled as the multinomial mixture of k latent topics z . In turn, each topic can be represented as a mixture of the probability distribution of n variables that make up the set of all possible variables constituting the space of observations ($V \in R^N$).

The LDA algorithm randomly assigns, to each variable of an observation, one of the k topics. Subsequently, it computes the proportion of the variable n in each observation of the total set and represents it as a Dirichlet probability distribution

over the latent topics, where the model of the whole corpus (D) of data is given by Equation (18).

$$\begin{aligned}
 p(D|\alpha, \beta) &= \prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(m_{dn}|z_{dn}, \beta) \right) d\theta_d \quad (18)
 \end{aligned}$$

here, α is the Dirichlet distribution parameter, β is the parameterization matrix of observation m , θ is the expected proportion of topics in observation m , and z represents the assignment of a word in a topic k [82].

E. IMAGE GENERATION (PROPOSED METHODS)

The above reasoning can be applied to images, where each pixel is taken as the features of the vectors subject to reduction. In other words, the 199×88 images are flattened to a vector of **17, 512** features, which can be represented in a latent space of lower dimensionality through the different methods described. In this case, for all transformations, the feature vector was reduced to **267** latent variables (principal components, for the case of PCA variations).

The latent variables are constructed with the different mathematical models, which are reversible and are generally referred to as inverse transformations. In fact, this concept is widely used in machine learning for data whitening. In principle, the data are represented with lower dimensionality latent variables, which may be associated with noise. Therefore, removing such variables (or components) would eliminate the noise in the images when they are returned to the original space through inverse transformation [83]. In this sense, altering the latent variables implies partially modifying the images without losing their spatial characteristics.

Based on the above consideration, altering the latent variables with random noise would imply the partial modification of the image, preserving its main attributes and creating new images from a reference image. In summary, our method consists of projecting the images to a latent space of lower dimensionality (see Equation (19)). Each latent variable is multiplied point by point (Hadamard Product) with a random vector V of the same dimensions but made up with values from a threshold t to 1 (see Equation (20)). The threshold value is determined by the expression $t = 1 - noise$, where $noise$ is the proportion of noise added. Finally, the modified latent variables are used to return to the original space through the inverse transformation, as indicated in Equation (21).

$$Z = TX \quad (19)$$

$$X \in R^m, Z \in R^n \text{ with } n < m$$

$$Z' = Z \odot V \quad (20)$$

$$V = (v_1, v_2, v_3, \dots, v_n) \forall v_i \in [t, 1]$$

$$X' = T^{-1} \{Z'\} \quad (21)$$

here, X represents the flattened image, $T\{\}$ the transformation depending on the reduction model used. Z the image in latent space, i.e., the vector of latent variables or components.

\odot is the Hadamard product. Z' the component vector modified with the random noise of the vector V , $T^{-1}\{\}$ the inverse transformation and X' the new flattened image.

F. TRANSFER LEARNING

DL networks require a large corpus of data due to many training parameters. Fortunately, some methods help overcome this drawback, such as transfer learning. The process consists of taking a network and training it with an extensive database, allowing filters to be tuned to this reference database. The adjusted filters are then used in a secondary network, allowing the model to train from the predefined filters. In general, if the database is large enough, the network learns the task with a high degree of generalization. The attribute can be retained for an equivalent task with another dataset. The model is trained based on the learned weights, and the network would achieve better performance, even if the data is small [84], [85].

Based on this premise, it was proposed to train the ResNet50v2 network with learning transfer from two different databases. First, the transfer was performed from the ImageNet database (1.28 million images), a widely accepted and used database [86]. Second, the ‘‘Thermal Image dataset for object classification’’ database (6695 images) was used [87], consisting of thermal images of cats, cars, and humans. In principle, we expect that the weights learned will be able to preserve thermal patterns that will allow better performance with diabetic foot thermography databases. It should be noted that the first database already has the weights organized in the structure of various neural networks, as is the case of the ResNet50v2 network. In contrast, the second database does not have this attribute; therefore, for this database, it was necessary to train the model to store the weights and transfer them to the classification model of subjects with DM. The training process for the latter is presented in Appendix C.

G. PERFORMANCE EVALUATION METRICS

The review of related work shows several performance evaluation metrics of artificial intelligence algorithms. Therefore, this research focused on evaluating the five most common metrics, namely accuracy, sensitivity, specificity, F_1 score, and precision. The metrics are expressed as shown in Equations (22) through (26) [88]–[90].

$$\text{Accuracy} = \frac{TP + TN}{TN + TP + FP + FN} \quad (22)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (23)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (24)$$

$$F_{1\text{score}} = \frac{2TP}{2TP + FP + FN} \quad (25)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (26)$$

The five metrics are given in terms of True Positives (TP), True Negatives (TN), False Positives (FP), and False

Negatives (FN). In addition, for this specific case, the metrics represent the following observations:

- Accuracy: The ability of a network to correctly classify different classes, i.e., subjects with DM and healthy or control group CG) subjects.
- Sensitivity: The ability of a network to classify actual DMs.
- Specificity: The ability of a network to correctly classify images of CG subjects.
- F_1 score: The ability to correctly identify the different classes in proportion to the number of classes.
- Accuracy: The proportion of real DM subjects among all detected DMs.

H. STATISTICAL ANALYSIS

For statistical estimation between groups, the non-parametric Kruskal Wallis test was implemented. The statistic evaluates whether two or more groups belong to the same distribution, based on the median of these groups. In other words, the test is based on the null hypothesis with the assumption that all samples come from the same distribution. Consequently, a p -value lower than 0.05 significance level would indicate that the null hypothesis is false, revealing a statistically significant difference between the two groups tested [91].

Assuming k groups with n observations, the method defines the H statistic given by the mathematical expression of Equation (27).

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i(\bar{r}_i - \bar{r})^2 \quad (27)$$

$$\bar{r}_{ij} = \frac{\sum_{i=1}^{n_i} r_{ij}}{n_i} \quad (28)$$

where, n_i is the number of observations of the i -th group, N is the total number of observations of the two groups, r_{ij} is the rank of the i -th observation over the j -th observation among all observations, and k is the number of groups [92], [93].

I. SEGMENTATION OF CRITICAL REGIONS

As mentioned above, diabetic foot neuropathy results in loss of lower limb sensation and reduced proprioception. The condition manifests with recurrent walking trauma, atrophying and causing dislocation of the protective pads, compromising the integrity of these regions [94]. Likewise, neuropathy decreases sweating and affects skin perfusion, generating dryness and hyperkeratosis, compromising the tissue with cracks or fissures [95], [96]. The consequences can be seen in the appearance of lesions preceded by ulcerations. However, trauma, recurrent peak pressures, and shearing produce significant temperature changes in the compromised regions [97]. In this regard, a simple model was used to highlight critical regions prone to injury and ulceration. The model consisted of highlighting regions that exceed the average plus the standard deviation of the plantar temperature of each foot in subjects classified with MD. That is, critical regions were segmented following the mathematical model

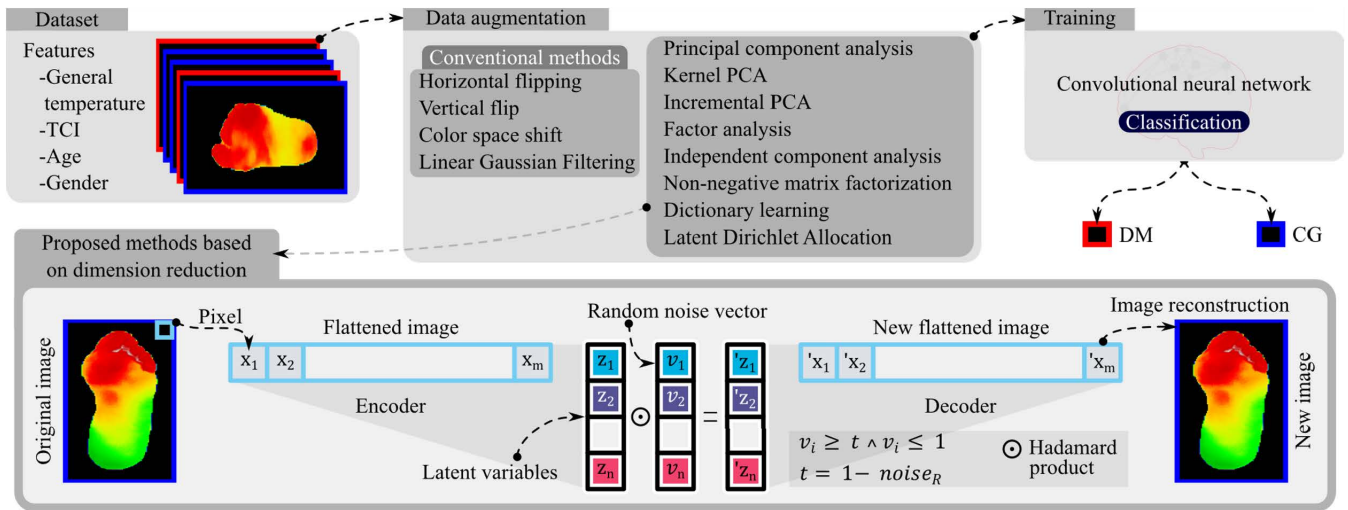


FIGURE 5. General scheme of the experimental design of the deep learning model.

of equation (29).

$$Segmentation = \begin{cases} Critical, & \text{if } T \geq \mu_f + \sigma_f \\ Normal, & \text{if } T < \mu_f + \sigma_f \end{cases} \quad (29)$$

Here, T is the temperature of each pixel of the plantar image, μ_f is the average temperature, and σ_f is the standard deviation of each foot.

J. EXPERIMENTAL DESIGN

Initially, the characteristics of the subjects were taken and evaluated through the Kruskal-Wallis test statistic. The evaluation was done between the control group (CG) and diabetes mellitus (DM) in the characteristics of general temperature (average), angiosome temperatures, TCI coefficient, age, and gender. Based on this validation, a new coefficient was constructed to discriminate between the different subjects. In addition, the temperature histogram was also explored, where the average curves and the error bands associated with the two groups and each foot were generated. Subsequently, the thermographic images were taken and processed as described in the methodology. The images were adjusted to 199×88 size, normalized, divided into training and test data, and evaluated through 5-fold cross-validation.

The training data were used as reference images to augment the images with the strategies of:

- Horizontal flipping.
- Vertical flip.
- Color space shift.
- Linear Gaussian filtering.
- Principal component analysis.
- Kernel PCA.
- Incremental PCA.
- Factor analysis.
- Independent component analysis.
- Non-negative matrix factorization.

- Dictionary Learning.
- Latent Dirichlet Allocation.

Once the augmented images were obtained, the ResNet50v2 convolutional neural network was trained under different training conditions (See Figure 5). Initially, it was trained from scratch, i.e., with the training parameters initialized with arbitrary values. Second, the network was trained by transfer learning with the weights generated by the ImageNet database. Third, the experiment was replicated with transfer learning but from the “Thermal Image dataset for object classification” database [87]. For the latter case, ResNet50v2 was trained, and the generated weights were used in the transfer. The results of this training are shown in the appendix section.

After training, the network was evaluated with the test set under accuracy, sensitivity, and specificity, F_1 score, and precision metrics. Additionally, the network was run six times per fold using the following hyperparameters:

- Number of epochs: 50
- Loss function: Cross-entropy
- Optimizer: Adadelta
- Weight initialization: Uniform Glorot or Transfer Learning, as appropriate.
- Bias initialization: Zeros or Transfer Learning, as appropriate.
- Batch size: 16.

The architecture was modeled under Python, mainly with the Keras and TensorFlow libraries. The execution was performed on the Colab platform configured with 25 GB of Ram and a Tesla T4 GPU. The implemented codes are left publicly available in the following GitHub repository (https://github.com/AndresAnayaIsaza/Data_augmentation_in_thermography-Diabetes_mellitus).

Next, the different data augmentation methods were evaluated using the non-parametric Kruskal-Wallis test statistic with the obtained accuracies. Finally, the best performing

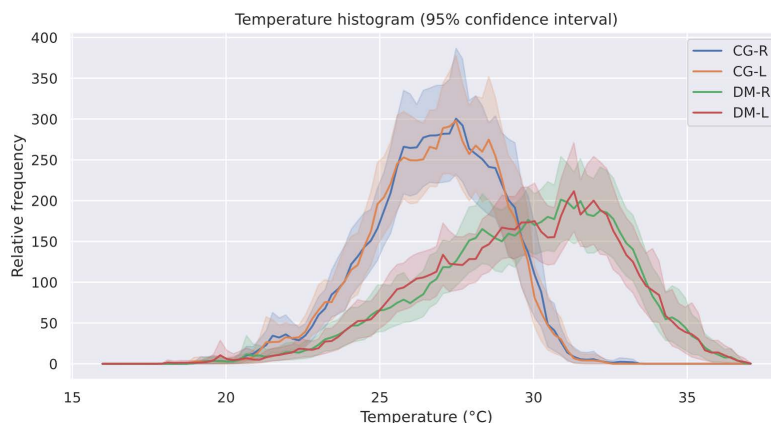


FIGURE 6. Average histogram with 95% error band for CG and DM subjects in both feet.

model was used to detect subjects with DM to which the critical temperature regions were segmented to establish power ulceration regions.

III. RESULTS

This section presents the results generated by the different parts described in the methodology. Initially, the problem was approached from the characteristics obtained from the subjects and the thermal images. Table 2 shows the p-value generated by the Kruskal-Wallis non-parametric test statistic. The statistic was evaluated for the eight characteristics between healthy or CG subjects and subjects with DM. In particular, the p-value was several orders below the significance level of 0.05 between CG and DM, i.e., the null hypothesis that the two groups come from the same distribution is rejected. Conversely, the p-value generated in comparing both feet is above the 0.05 level, indicating that the null hypothesis is retained. In other words, there is no statistically significant difference between the characteristics of both feet of both CG and DM subjects.

On the other hand, exploration of the histogram (see Figure 6) showed that the temperature distributions present differences between the two subjects. First, patients diagnosed with DM present a temperature shift above 30° Celsius, while healthy subjects present a temperature distribution centered between 25 to 30° Celsius. In other words, the prevalence of a higher proportion of the foot with temperatures above 30° would be an indication associated with DM.

Based on the above considerations, it is possible to assume that the characteristics studied serve to discriminate between the two groups. In this order of ideas, a new classification coefficient was constructed, governed by the mathematical expression of Equation (30).

$$Coefficient = \bar{T} + A^2 + TCI_R + TCI_L \quad (30)$$

here, \bar{T} is the average temperature of both feet, A is the age, TCI_R and TCI_L are the thermal change index of the right and left foot, respectively.

TABLE 2. p-value of the Kruskal-Wallis test statistic evaluated between the different characteristics of CG, DM, and between feet subjects.

Features	Between-subjects		Between-foot	
	CG vs DM	Left vs right of CG	Left vs right of DM	
General	1.32E-18	0.935	0.747	
LCA	1.01E-16	0.859	0.736	
LPA	1.17E-19	0.947	0.723	
MCA	1.31E-15	0.827	0.799	
MPA	1.40E-19	0.910	0.861	
TCI	3.76E-22	0.942	0.723	
Age	2.60E-70	-	-	
Gender	5.53E-06	-	-	

TABLE 3. Evaluation metrics using the hyperplane support vector machine.

Metrics	Coefficient	Right TCI	Left TCI
Classification threshold*	102.55	1.02	0.47
Accuracy (%)	94.61	77.25	75.45
Sensitivity (%)	97.54	89.34	96.72
Specificity (%)	86.67	44.44	17.78
F1 score (%)	96.36	85.16	85.20
Precision (%)	95.20	81.34	76.13

*Or hyperplane

Figure 7a, shows the score returned for the 167 study subjects. The coefficient showed a marked difference between CG and DM subjects concerning the TCI value (see Figure 7b). It was feasible to search for a classification threshold through a support vector machine in awareness. The strategy with the SVM was replicated in the TCI for the right and left feet. The results are shown in Table 3 with the conventional evaluation metrics.

Table 3 shows an increase of more than 17% with the new coefficient concerning the most accurate

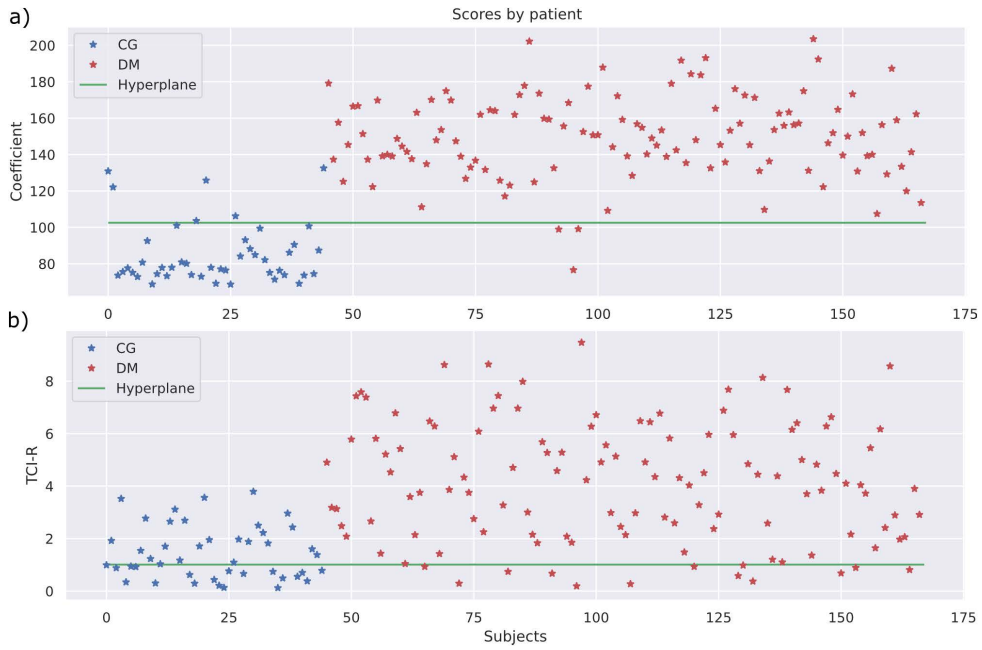


FIGURE 7. Scores with a) the new coefficient and b) the TCI value of the right foot. Each figure is shown with the hyperplane (green linear) generated by the support vector machine model.

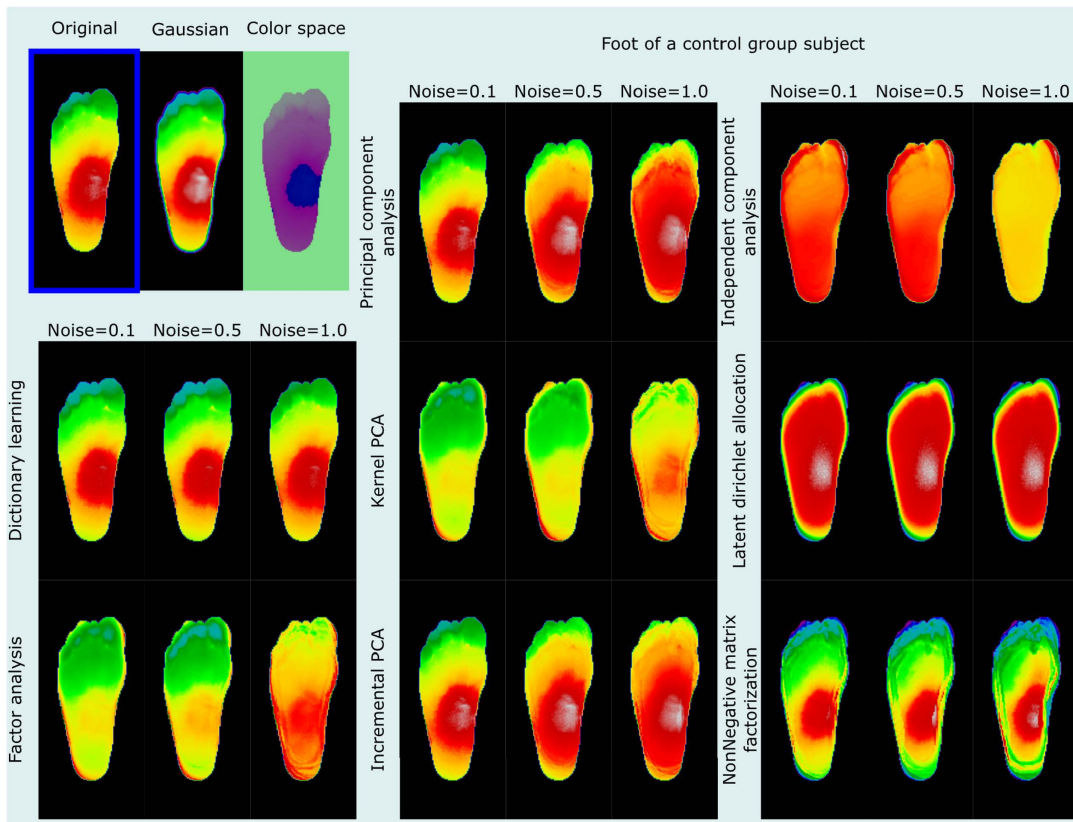


FIGURE 8. Images synthesized from the proposed methods with different percentages of noise. The results are shown by randomly taking a healthy subject or CG as a reference image.

TCI (right foot). In addition, the coefficient has a sensitivity of more than 97%, i.e., it has a high capacity to identify true DM cases.

As mentioned in the materials and methods section, in the second instance, the diagnosis of subjects with DM was approached through the novel deep learning

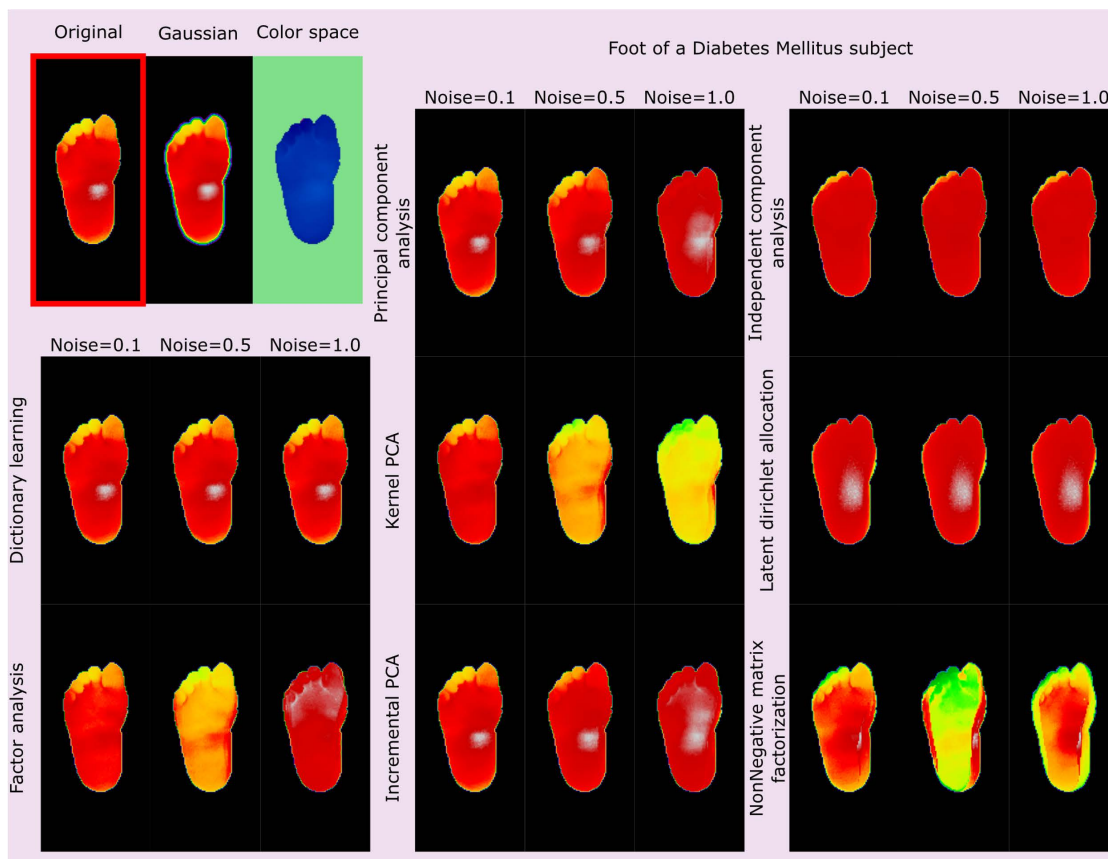


FIGURE 9. Images synthesized from the proposed methods with different percentages of noise. The results are shown by randomly taking a diabetic or DM subject as a reference image.

convolutional neural networks. In addition, this development was approached from data augmentation strategies to mitigate the effect generated by the limited data sets.

Figure 8 shows the images synthesized by the data augmentation methods for a reference image of a healthy subject. Additionally, the proposed methods project the data onto new latent variables to which a noise component is introduced. Consequently, the proposed methods are presented with different noise ratios. Similarly, Figure 9 shows synthesized images, but for the case of a subject with DM.

The results presented in Figure 8 and Figure 9 show that the feet retain their morphological structure with changes in temperature distributions. The patterns generated vary from one method to another. For example, principal component analysis with 0.1 noise preserves the reference image almost unchanged, while the LDA deforms the original foot pattern.

As mentioned above, the ResNet50v2 model was trained under different training conditions. Initially, the model was trained with the different data augmentation methods by initializing the training parameters with arbitrary values (under the Uniform Glorot distribution). The results of this first experiment are shown in Table 4. The values are expressed as average percentage values with the respective standard deviation. Subsequently, the same training was performed by

transferring the weights from the ImageNet database. The results of this are shown in Table 5. Finally, the process was replicated using the transfer learning from the database “Thermal Image dataset for object classification” [87], where the latter results are presented in Table 6.

The results in Table 4 show an increase of about 9% in accuracy between the highest and best scoring methods. That is, data augmentation based on incremental PCA presented a 9% increase in training without any data augmentation method (row None). At the same time, data augmentation with the conventional color space shift technique presented about a 5% increase in training without methods. In other words, the results demonstrate the high effectiveness of data augmentation methods, including the proposed new methods.

Similarly, Table 5 presents a similar behavior, i.e., the data augmentation methods increased the accuracy of the ResNet50v2 network. For example, the linear Gaussian filter-based strategy improved by 8.5% in training without data augmentation. Similarly, the conventional image flipping method increased the model’s accuracy by 5.6%, again guaranteeing the high effectiveness of training with data augmentation. On the other hand, it is worth noting that transfer learning produced a marked difference to the model trained from scratch. The results in Table 4 are close to 81%, while the results

TABLE 4. Training of the ResNet50v2 model from scratch as a function of different training conditions (data augmentation). Scores are given in average values with the respective standard deviation, calculated from the 30 different runs, i.e., the runs generated by the six repetitions of each fold (5 folds). Scores generated with the test data.

Data augmentation methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1_score (%)
Incremental PCA	81.72±4.63	77.27±14.97	77.27±14.97	78.01±13.20	76.67±12.24
Nonnegative matrix factorization	81.36±4.25	73.97±19.74	73.97±19.74	78.54±12.03	74.69±14.07
Dictionary learning	81.31±5.54	78.19±13.14	78.19±13.14	77.72±15.02	76.94±11.68
Principal component analysis	81.26±4.24	76.73±14.69	76.73±14.69	77.62±13.50	76.16±11.90
Horizontal flip	80.10±6.15	76.13±12.92	76.13±12.92	76.20±15.67	75.33±12.83
Vertical flip	79.65±6.36	75.85±13.13	75.85±13.13	75.76±15.85	74.90±12.77
Gaussian linear filter	79.49±3.61	67.18±29.06	67.18±29.06	78.85±12.20	68.56±19.86
Factor analysis	79.24±5.15	73.29±18.02	73.29±18.02	75.31±13.99	72.96±14.60
Latent dirichlet allocation	79.19±4.43	68.87±25.92	68.87±25.92	76.34±12.44	69.81±18.41
Kernel PCA	79.14±4.24	73.36±17.52	73.36±17.52	75.51±14.38	72.95±13.77
Independent component analysis	78.94±5.12	74.13±17.07	74.13±17.07	74.47±14.36	73.15±13.99
Change of color space	77.88±4.33	71.64±20.13	71.64±20.13	74.67±14.86	70.99±15.07
None	72.62±5.73	69.02±12.16	69.02±12.16	68.98±14.87	68.20±11.85

TABLE 5. Training of the ResNet50v2 model with learning transfer from ImageNet as a function of different training conditions (data augmentation). Scores are given in average values with the respective standard deviation, calculated from the 30 different runs, i.e., the runs generated by the six repetitions of each fold (5 folds). Scores generated with the test data.

Data augmentation methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1_score (%)
Gaussian linear filter	90.61±2.51	87.47±10.44	87.47±10.44	89.13±8.24	87.72±7.12
Independent component analysis	89.90±2.77	86.46±11.21	86.46±11.21	88.26±7.39	86.82±7.41
Nonnegative matrix factorization	89.80±4.86	85.61±13.54	85.61±13.54	88.59±8.37	86.50±9.56
Incremental PCA	89.70±3.79	86.39±12.09	86.39±12.09	88.12±9.52	86.52±8.70
Kernel PCA	89.55±3.38	86.60±10.49	86.60±10.49	87.54±9.66	86.50±8.06
Factor analysis	89.39±3.03	86.42±10.40	86.42±10.40	86.84±10.99	85.96±8.30
Change of color space	88.74±3.78	86.16±10.79	86.16±10.79	86.31±10.24	85.62±8.59
Dictionary learning	88.73±4.11	85.45±12.37	85.45±12.37	86.65±9.83	85.36±9.22
Latent dirichlet allocation	88.23±4.24	84.28±12.23	84.28±12.23	85.79±8.68	84.63±9.50
Horizontal flip	88.08±4.09	84.52±11.25	84.52±11.25	85.63±10.99	84.57±9.56
Principal component analysis	88.07±4.05	83.95±14.19	83.95±14.19	86.95±9.12	84.52±9.44
Vertical flip	87.73±3.64	83.76±13.28	83.76±13.28	85.84±9.98	83.95±9.42
None	82.08±4.26	78.07±12.96	78.07±12.96	80.21±9.96	78.43±9.56

with learning transfer (Table 5) are close to or above 90%, i.e., learning transfer significantly improves the network’s performance.

Finally, the results of the model trained with the data transfer from the thermographic database are shown in Table 6. Although the results present a significant increase concerning the training without transfer, it does not present any noticeable difference with respect to the transfer from the ImageNet database. In other words, transfer learning increases significantly; however, the database used in the transfer of the weights does not seem to be relevant, as long as the database has a large number of data.

Figure 10 shows the distribution of the accuracy obtained by the data augmentation methods and the three training conditions of the ResNet50v2 network. It should be noted that the graphs are shown with fractional values to percentage equivalents, i.e., with values from 0 to 1, indicating values

from 0 to 100%. In addition, the results of the remaining metrics are shown in Appendix D.

The distributions corroborate what was presented above. In the model trained from scratch, the box and whisker plots of the methods are centered near 80%, while the boxes of the methods with transfer learning are above 80 and centered at approximately 90%. In addition, it can also be observed that the distribution without the data augmentation method (None) is partially below the others in all three training conditions.

Table 7, Table 8 and Table 9 show the test statistic evaluated between the different data augmentation conditions for model training. The first one shows the values for training from scratch and the second and third ones show the values for training with transfer learning from ImageNet and the thermographic database. Although the p-values vary substantially between methods, it is evident that the comparison

TABLE 6. Training of the ResNet50v2 model with learning transfer from the thermographic database as a function of different training conditions (data augmentation). The scores are given in average values with the respective standard deviation, calculated from the 30 different runs, i.e., the runs generated by the six repetitions of each fold (5 folds). Scores generated with the test data.

Data augmentation methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1_score (%)
Factor analysis	90.05±3.32	87.15±10.52	87.15±10.52	88.17±8.36	87.18±7.50
Change of color space	89.75±3.34	87.49±10.22	87.49±10.22	87.62±9.49	86.96±7.51
Incremental PCA	89.60±3.02	86.18±10.72	86.18±10.72	87.64±8.67	86.43±7.99
Gaussian linear filter	89.55±3.01	85.06±13.07	85.06±13.07	88.09±7.87	85.95±8.90
Dictionary learning	89.49±3.03	86.45±11.16	86.45±11.16	87.65±8.84	86.42±7.76
Principal component analysis	89.34±2.48	85.74±11.85	85.74±11.85	87.70±9.01	86.02±7.90
Nonnegative matrix factorization	89.09±3.65	85.01±11.92	85.01±11.92	87.34±7.89	85.70±8.57
Vertical flip	88.94±3.23	85.02±12.44	85.02±12.44	87.20±8.66	85.46±8.66
Horizontal flip	88.79±2.54	84.88±11.47	84.88±11.47	86.99±9.33	85.31±8.09
Kernel PCA	88.64±2.78	85.14±11.05	85.14±11.05	86.36±9.68	85.23±8.38
Latent dirichlet allocation	88.64±3.39	84.70±11.82	84.70±11.82	86.48±8.62	85.12±8.84
Independent component analysis	88.59±2.81	84.53±12.48	84.53±12.48	87.05±9.54	84.99±8.48
None	82.39±3.96	78.65±11.86	78.65±11.86	80.32±8.76	79.07±9.09

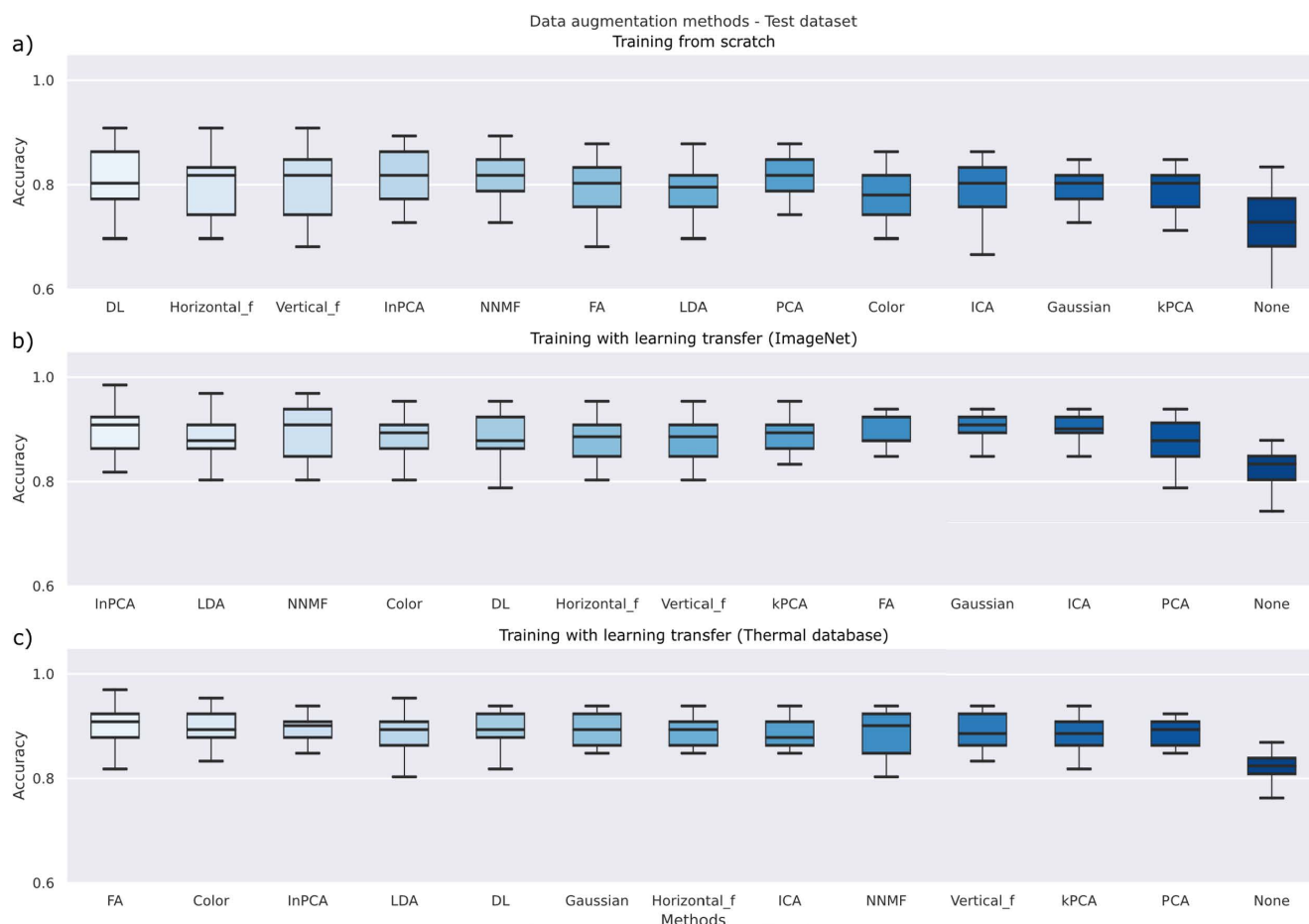


FIGURE 10. Distribution of the accuracies represented in boxes and whiskers for the data augmentation methods and under the conditions of a) training from scratch, b) transfer learning from ImageNet, and c) transfer learning from the thermographic database.

without data augmentation (None) shows p-values below the 0.05 significance level for all cases. That is, the null hypothesis that the scores come from samples with the same

distribution can be rejected. In other words, all data augmentation methods are statistically significant in neural network performance, including the proposed new methods.

TABLE 7. P-value generated by the Kruskal-Wallis test statistic evaluated between the data augmentation methods and the accuracy metric in the case of training from scratch.

		1	2	3	4	5	6	7	8	9	10	11	12	13	
Data augmentation methods	Dictionary learning	1	1.00	0.37	0.19	0.70	0.87	0.08	0.04	0.92	0.00	0.05	0.06	0.04	0.00
	Horizontal flip	2	0.37	1.00	0.74	0.15	0.28	0.42	0.28	0.32	0.03	0.23	0.35	0.26	0.00
	Vertical flip	3	0.19	0.74	1.00	0.08	0.17	0.64	0.47	0.20	0.08	0.35	0.60	0.47	0.00
	Incremental PCA	4	0.70	0.15	0.08	1.00	0.55	0.01	0.00	0.54	0.00	0.00	0.00	0.00	0.00
	Nonnegative matrix factorization	5	0.87	0.28	0.17	0.55	1.00	0.03	0.01	0.98	0.00	0.01	0.01	0.00	0.00
	Factor analysis	6	0.08	0.42	0.64	0.01	0.03	1.00	0.67	0.03	0.08	0.68	0.83	0.69	0.00
	Latent dirichlet allocation	7	0.04	0.28	0.47	0.00	0.01	0.67	1.00	0.00	0.11	0.76	0.71	0.92	0.00
	Principal component analysis	8	0.92	0.32	0.20	0.54	0.98	0.03	0.00	1.00	0.00	0.01	0.01	0.00	0.00
	Change of color space	9	0.00	0.03	0.08	0.00	0.00	0.08	0.11	0.00	1.00	0.10	0.05	0.13	0.00
	Independent component analysis	10	0.05	0.23	0.35	0.00	0.01	0.68	0.76	0.01	0.10	1.00	0.90	0.85	0.00
	Gaussian linear filter	11	0.06	0.35	0.60	0.00	0.01	0.83	0.71	0.01	0.05	0.90	1.00	0.79	0.00
	Kernel PCA	12	0.04	0.26	0.47	0.00	0.00	0.69	0.92	0.00	0.13	0.85	0.79	1.00	0.00
	None	13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

TABLE 8. p-value generated by the Kruskal-Wallis test statistic evaluated between the data augmentation methods and the accuracy metric in the case of transfer learning training from ImageNet.

		1	2	3	4	5	6	7	8	9	10	11	12	13	
Data augmentation methods	Incremental PCA	1	1.00	0.08	0.47	0.24	0.18	0.03	0.01	0.71	0.63	0.15	0.76	0.04	0.00
	Latent dirichlet allocation	2	0.08	1.00	0.03	0.59	0.56	0.65	0.40	0.13	0.20	0.00	0.01	0.80	0.00
	Nonnegative matrix factorization	3	0.47	0.03	1.00	0.08	0.10	0.02	0.00	0.42	0.40	0.93	0.39	0.01	0.00
	Change of color space	4	0.24	0.59	0.08	1.00	0.99	0.29	0.10	0.41	0.48	0.01	0.12	0.39	0.00
	Dictionary learning	5	0.18	0.56	0.10	0.99	1.00	0.35	0.18	0.22	0.42	0.00	0.04	0.31	0.00
	Horizontal flip	6	0.03	0.65	0.02	0.29	0.35	1.00	0.67	0.06	0.17	0.00	0.01	0.89	0.00
	Vertical flip	7	0.01	0.40	0.00	0.10	0.18	0.67	1.00	0.02	0.10	0.00	0.00	0.50	0.00
	Kernel PCA	8	0.71	0.13	0.42	0.41	0.22	0.06	0.02	1.00	0.61	0.03	0.36	0.05	0.00
	Factor analysis	9	0.63	0.20	0.40	0.48	0.42	0.17	0.10	0.61	1.00	0.01	0.15	0.11	0.00
	Gaussian linear filter	10	0.15	0.00	0.93	0.01	0.00	0.00	0.00	0.03	0.01	1.00	0.14	0.00	0.00
	Independent component analysis	11	0.76	0.01	0.39	0.12	0.04	0.01	0.00	0.36	0.15	0.14	1.00	0.01	0.00
	Principal component analysis	12	0.04	0.80	0.01	0.39	0.31	0.89	0.50	0.05	0.11	0.00	0.01	1.00	0.00
	None	13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

In particular, the results showed high performance of the ResNet50v2 network, reaching values close to 100% accuracy, training the model with transfer learning, and using the proposed data augmentation method: Incremental PCA.

However, it should be emphasized that the development was performed on each foot individually, i.e., the scores of the evaluation metrics were performed on each foot individually (test data). Therefore, to augment the model’s predictive model, we weighted the probability score of each foot to generate an overall score. Using the best-performing model, the implementation of this principle allowed us to reach 100% accuracy with both training and test data.

Finally, starting from this model, we implemented the detection of subjects with DM and highlighted the higher temperature regions, generating the critical temperature regions, regions prone to further ulceration due to temperature increase from friction and shear.

In this sense, Figure 11 and Figure 12 present the results of the detection model, where the probability of a subject suffering from diabetes mellitus is estimated [98]–[100].

The results indicate that the healthy subject with the highest probability of belonging to DM is 24.97%, i.e., in the worst case, the network has a certainty of 75.09 that the subject is healthy. Similarly, the subject with DM with the best probability of being DM is 69.19%, i.e., in the worst case, the network has a 69.19% certainty of classifying the subject with DM.

On the other hand, some subjects presented the maximum temperatures over the inner part of the foot or over the butterfly pattern. This would imply that the subject, although suffering from diabetes, is far from developing a diabetic foot ulcer. Although the classification is 100% accurate and the regions of maximum temperature could reveal the possibility of developing diabetic foot, longitudinal

TABLE 9. p-value generated by the Kruskal-Wallis test statistic evaluated between the data augmentation methods and the accuracy metric in the case of training with transfer learning from the thermographic database.

		1	2	3	4	5	6	7	8	9	10	11	12	13	
Data augmentation methods	Factor analysis	1	1.00	0.65	0.33	0.02	0.31	0.37	0.01	0.01	0.18	0.08	0.01	0.12	0.00
	Change of color space	2	0.65	1.00	0.81	0.08	0.69	0.66	0.09	0.02	0.53	0.16	0.03	0.37	0.00
	Incremental PCA	3	0.33	0.81	1.00	0.09	0.90	0.92	0.06	0.02	0.74	0.29	0.04	0.54	0.00
	Latent dirichlet allocation	4	0.02	0.08	0.09	1.00	0.14	0.15	0.92	0.69	0.30	0.64	0.89	0.24	0.00
	Dictionary learning	5	0.31	0.69	0.90	0.14	1.00	0.96	0.13	0.05	0.82	0.30	0.09	0.64	0.00
	Gaussian linear filter	6	0.37	0.66	0.92	0.15	0.96	1.00	0.11	0.09	0.57	0.35	0.12	0.62	0.00
	Horizontal flip	7	0.01	0.09	0.06	0.92	0.13	0.11	1.00	0.56	0.22	0.68	0.87	0.17	0.00
	Independent component analysis	8	0.01	0.02	0.02	0.69	0.05	0.09	0.56	1.00	0.27	0.43	0.73	0.10	0.00
	Nonnegative matrix factorization	9	0.18	0.53	0.74	0.30	0.82	0.57	0.22	0.27	1.00	0.66	0.30	0.97	0.00
	Vertical flip	10	0.08	0.16	0.29	0.64	0.30	0.35	0.68	0.43	0.66	1.00	0.48	0.62	0.00
	Kernel PCA	11	0.01	0.03	0.04	0.89	0.09	0.12	0.87	0.73	0.30	0.48	1.00	0.17	0.00
	Principal component analysis	12	0.12	0.37	0.54	0.24	0.64	0.62	0.17	0.10	0.97	0.62	0.17	1.00	0.00
	None	13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

studies are needed to determine if the proposed method has clinical validity.

IV. DISCUSSION

Diabetes Mellitus (DM) can manifest itself through different metabolic disorders. The disease has a high recurrence in people over 40 years of age, and the worldwide figures are alarming due to the high number of cases. In this sense, early detection of DM is a primordial necessity to avoid other affections. In this order of ideas, in this research, we proposed a framework to explore different approaches for detecting or diagnosing subjects with DM. First, we proposed to search for a new coefficient superior to the TCI to classify patients with DM. The proposed new coefficient proved to be more efficient than the TCI; the difference between the two is evident, as there was a nearly 17% increase in accuracy¹ (see Table 3). In addition, our new coefficient managed to outperform the most recent research-based on machine learning techniques and the TCI coefficient. While the latest research reported an accuracy of 90.1% [41], our coefficient reached 94.61%, a difference of more than 4%.

However, due to the lack of information with the data used, it is not possible to determine whether the value of the coefficient is proportional to the degree of severity of the pathology, i.e., it is not possible to determine whether the score corresponds to the real stratification of the pathology. In this sense, it is suggested to study whether subjects with different degrees of DM or diabetic foot correspond proportionally to the new estimated coefficient.

On the other hand, the study of the characteristics confirms what has been reported in the literature, where diabetic foot pathology is closely related to temperature changes due to friction and shearing due to lower limb

neuropathy [98]–[100]. Secondly, subjects with DM were detected from a deeper approach through convolutional neural networks, specifically through the ResNet50v2 network. Additionally, the implementation was approached from the data augmentation approach due to the lack of data in foot thermographs. Eight new proposed methods were integrated, which make a significant contribution to data augmentation. Although the synthesized images varied substantially among the data augmentation methods, this does not indicate that they are not efficient. For example, color space-shifting is one of the most accepted methods [58]; however, it differed quite a bit from the reference image (see Figure 8 and Figure 9). In other words, the data augmentation does not necessarily have to have an intuitive interpretation for the human, since, even if the image lacks visual interpretation, it could have real meaning for the artificial networks, a statement that has already been affirmed in previous research. Additionally, the values returned during training and the p-value of the Kruskal-Wallis test statistic ensure that the methods are statistically significant.

The results section shows that data augmentation has a significant difference concerning training without augmentation. Similarly, transfer learning also presented an improvement in network performance. However, the type of database does not show statistically significant differences. In other words, using transfer from ImageNet and the thermographic database is not relevant. These rules out the assumption that the network can learn thermal patterns that improve the performance of an application based on the same type of images. Although the ImageNet and the thermal database are based on two types of images, the transferred weights are taken as the initial values of the model, which are adjusted in training with the final database. Therefore, the fine-tuning would occlude the original transfer patterns, implying that the performance does not depend on the type of images

¹Accuracy validated through a support vector machine.

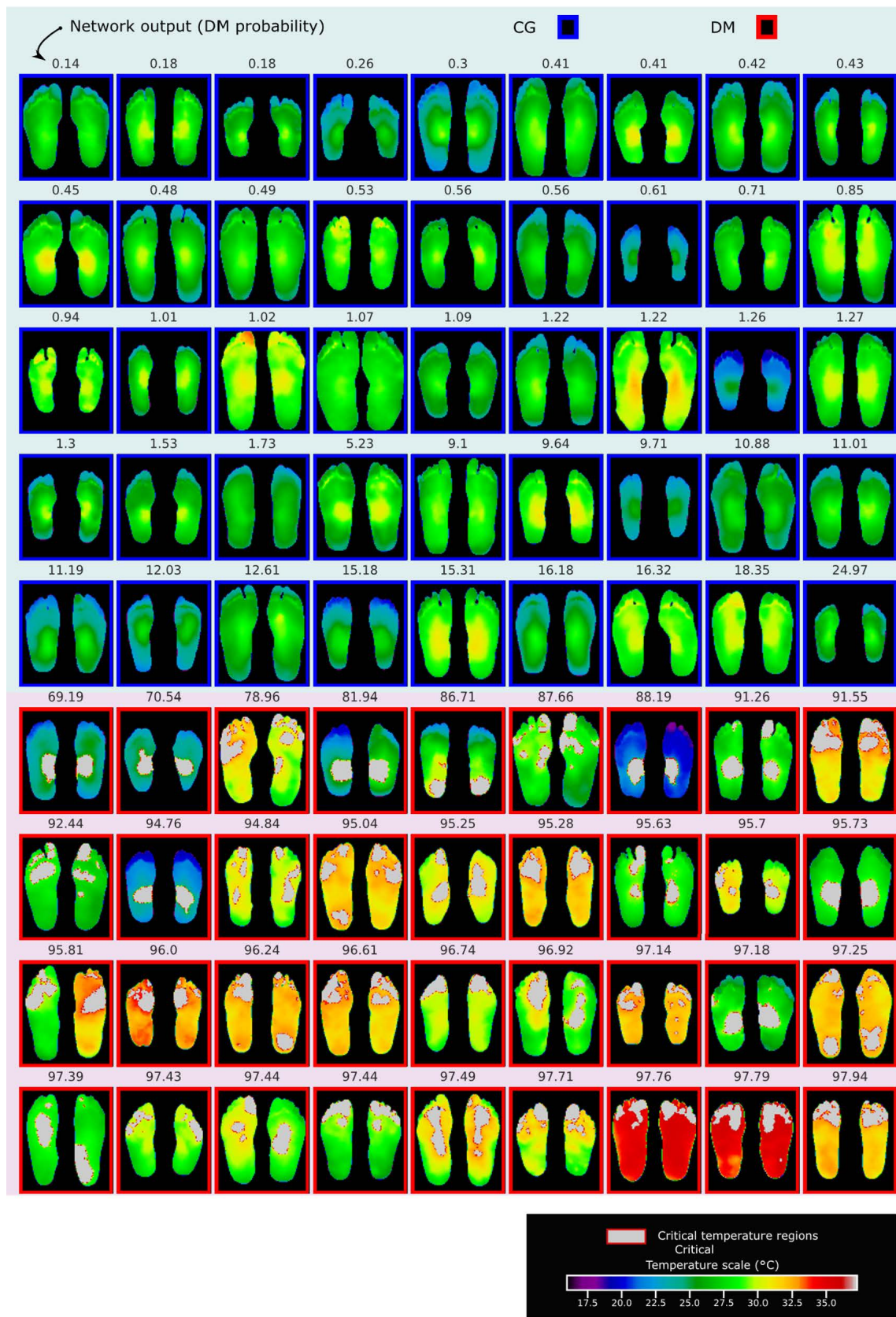


FIGURE 11. Output generated by the convolutional neural network. The scores are the weighted probability (between both feet) of being subjects with DM. Additionally, subjects diagnosed with DM have their critical temperature regions highlighted, i.e., regions prone to further ulceration.

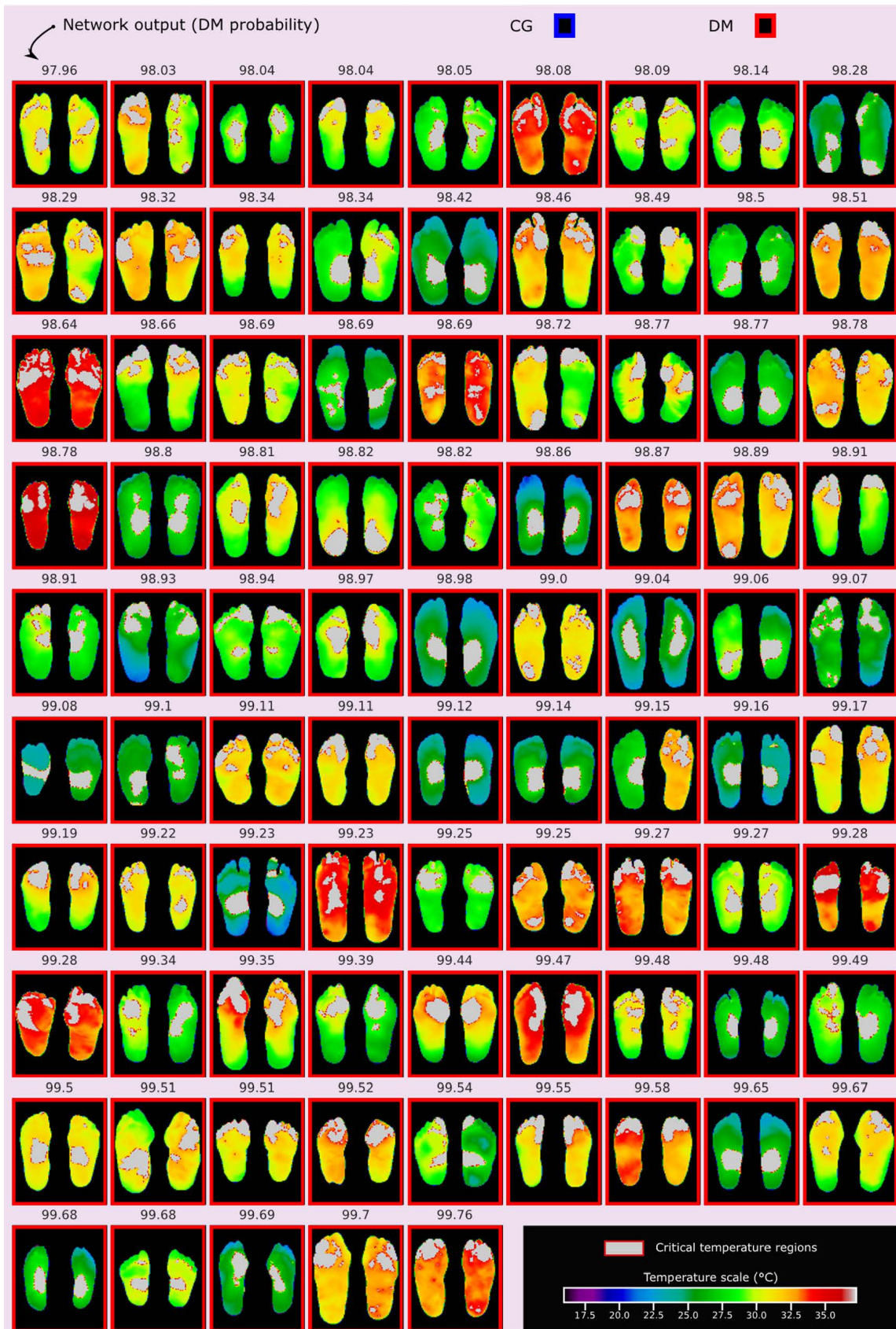


FIGURE 12. Output generated by the convolutional neural network. The scores are the weighted probability (between both feet) of being subjects with MD. Additionally, subjects diagnosed with DM have their critical temperature regions highlighted, i.e., regions prone to subsequent ulcerations.

in transfer learning. The results delivered by the network are highly satisfactory. 100% accuracy was achieved, i.e., true positives and true negatives were correctly classified, while no false positives or negatives were generated. That is, the network achieved 100% effectiveness in sensitivity and specificity, surpassing all previous investigations, where the maximum metric was achieved by Cruz-Vega *et al.* with an accuracy of 99.98% [38]. In this sense, we assume that our approach has two strong axes concerning the implementation performed by Cruz-Vega *et al.* First, our research was based on one of the state-of-the-art networks (ResNet50 Network), which has proven to be highly efficient in classification tasks. Second, the network training was based on proposed new data augmentation methods. Also, we assume that our framework is the most efficient because most of the recent research focused on conventional machine learning methods (See Table 1). For example, only three of the mentioned researches implemented deep learning [38]–[40]. However, our research is the first development that integrates a state-of-the-art network trained under new data augmentation methods to the best of our knowledge.

In addition, the critical temperature regions are a tool to determine regions with a high probability of ulceration. In fact, most of the processed images presented temperature maxima in the regions of highest pressure with a high probability of ulceration, i.e., the bottom of toes, pad of foot, and heel of foot [101].

The results are promising and are statistically significant concerning training without any data augmentation methods. Also, the methods were used to augment the existing data in the same proportion. For each reference image, a synthetic image was created, and with this data, the neural network was trained, allowing it to reach 100% accuracy. However, it should be clarified that the research has some limitations. First, the new coefficient generates a stratification of the subjects that correspond to their nature (Figure 7), but it was not verified whether these correspond to the degree of severity of the pathology. Furthermore, these values are not clinically corroborated; further studies are required to determine their validity. Secondly, the investigation could only be based on a single database, which we assume does not have sufficient heterogeneity, which is achieved with data from different centers.

Furthermore, this experimental framework leaves open some research questions that can be explored in future work. The methods allowed us to reach this excellent performance without considering that all methods were used individually. Therefore, it would be convenient to study if the combination of the methods would generate better performance, or posed as a question, could there be a combination of methods that improve the performance of convolutional networks? Finally, although ResNet50v2 showed excellent results, it has many training parameters; therefore, it would be interesting to validate if the proposed methods have the same impact on lower-order networks.

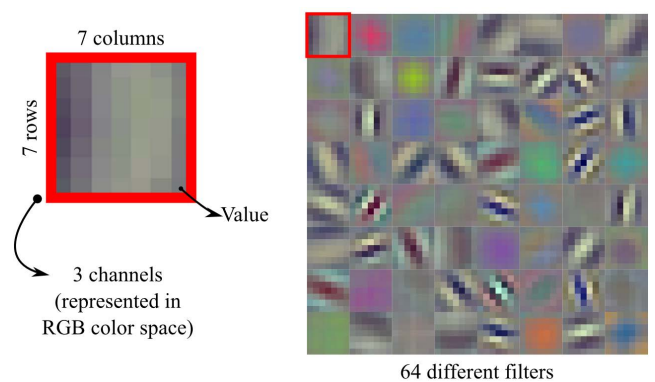


FIGURE 13. 64 different convolutional filters with 3 channels, 7 rows and 7 columns.

V. CONCLUSION

A framework for exploring different approaches for screening subjects with diabetes mellitus (DM) was proposed. Initially, we planned to study a new discrimination coefficient of the subjects based on the characteristics of average temperature, age, overall TCI of right and left feet. The new coefficient was able to exceed the accuracy of the TCI by up to 17%, which is a useful tool for the stratification of subjects with DM. Secondly, the ResNet50v2 network was explored to accurately classify subjects by integrating it with 12 data augmentation methods to make up for the low corpus of thermographic image data. The augmentation consisted of 4 conventional methods and eight proposed methods based on dimension reduction methods: principal component analysis, kernel PCA, incremental PCA, factor analysis, independent component analysis, non-negative matrix factorization, dictionary learning, and LDA. Initially, the images were taken to latent spaces of lower dimensionality using the aforementioned methods. The latent variables were altered with a random noise vector and returned to the original image space, generating new synthetic images with characteristics like the reference images. All methods were shown to be statistically significant with a significance level below 0.05, which is a relevant contribution to data enhancement.

Additionally, a comparative analysis was performed, we studied the behavior of the network under three training conditions. First, the network was trained from scratch. Secondly, we used transfer learning from the conventional ImageNet image database, and finally, we performed the same process, but from a thermographic database. The results showed that data augmentation and learning transfer are statistically significant strategies to improve the performance of convolutional neural networks. Furthermore, it was also shown that learning transfer has the same impact regardless of the nature of the images, as long as the data corpus is large enough to generate transferable learning patterns.

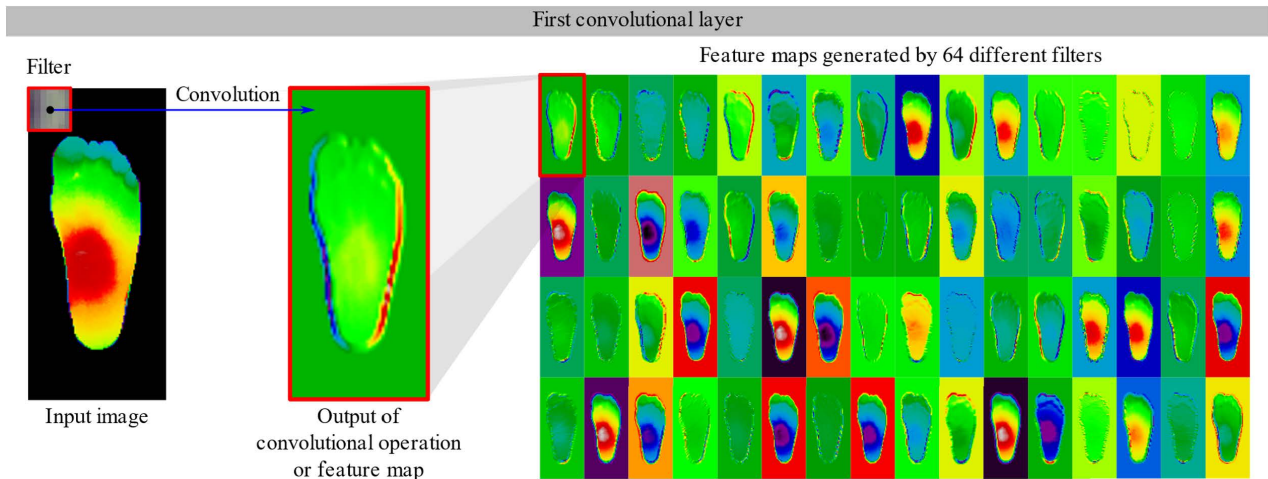


FIGURE 14. Activation maps of the 64 filters in Figure 13 used on the input image.

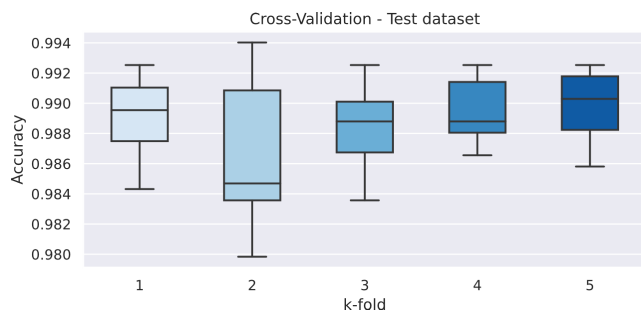


FIGURE 15. Distribution of accuracies of the different folds with the test data in the model trained with the thermographic database.

APPENDIX

The study presents some secondary results not presented in the main section. This section makes such useful results available to support those discussed throughout this paper.

A. CONVOLUTIONAL NEURAL NETWORK

In particular, as the name implies, convolutional networks are made up of a convolutional operator (or filter) that highlights some features of the image. The filter is a two-dimensional element belonging to the set of real numbers, which can have any number of combinations of values, as shown in Figure 13.

In general, networks are built with a different number of convolutional layers and with different filters per layer, which define the mathematical model of the network. Each filter is initially made up of random values that are progressively adjusted in the training process until the optimal parameters are reached. Once the model is fully trained, the network can take a new image, convolve it, and progressively generate feature maps as it moves through the different convolutional layers. For example, Figure 14 shows the first feature maps of a 64-filter model. The features of the first layer are then used by subsequent layers, which in turn create new and increasingly abstract feature maps.

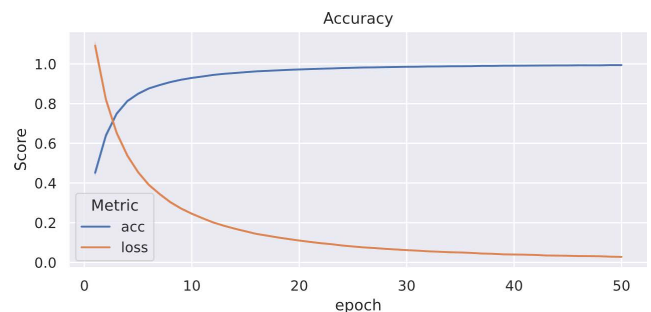


FIGURE 16. The plot of ResNet50v2 training as a function of epochs in accuracy (acc) and loss scores.

The features or feature maps predict a new result from what is learned from the training data. In essence, the mathematical model of the network is shaped by the expression of Equation (31).

$$A_j = \varphi \left(b_j + \sum_{i=1}^M X_i * K_{ij} \right) \tag{31}$$

Here, X_i is the input image in the i -th channel, M is the total number of channels or number of feature maps (as the case may be). K_{ij} is the i -th channel of the j -th convolutional filter. $(*)$ is the convolutional operator, b_j is the bias associated with the j -th filter, and φ is the activation function of the model. Therefore, A_j is the feature map generated by the j -th convolutional filter. It should be emphasized that Equation (31) is only the output of a single filter in a single layer of the network. The model is similar throughout all layers, but for anyone layer, the layer’s input is the previous layer’s output; therefore, this principle is represented as expressed in Equation (32).

$$A_j^{(l)} = \varphi^{(l)} \left(b_j^{(l)} + \sum_{i=1}^{M^{(l-1)}} A_i^{(l-1)} * K_{ij}^{(l)} \right) \tag{32}$$

TABLE 10. Average scores with standard deviation generated by the different folds in the training of the ResNet50v2 model with the thermographic database. Scores generated with the test data.

k_fold	Accuracy	Sensitivity	Specificity	Precision	F1_score
1	98.93±0.23	98.24±0.4	99.28±0.37	98.37±1.11	98.3±0.62
2	98.65±0.47	97.84±0.88	99.07±0.4	97.94±1.3	97.89±0.92
3	98.82±0.27	98.08±0.69	99.16±0.33	98.29±0.74	98.18±0.57
4	98.94±0.21	98.27±0.47	99.27±0.25	98.46±0.72	98.36±0.44
5	98.99±0.2	98.42±0.83	99.28±0.48	98.52±0.97	98.46±0.36

TABLE 11. Average scores with standard deviation generated by the different folds in the ResNet50v2 model trained from scratch. Scores generated with the test data.

k_fold	Accuracy	Sensitivity	Specificity	Precision	F1_score
0	81.59±3.24	74.52±16.88	74.52±16.88	77.07±11.17	75.13±13.51
1	77.22±4.40	77.33±9.48	77.33±9.48	70.55±22.64	71.59±13.46
2	81.90±3.65	75.39±17.00	75.39±17.00	78.03±9.56	76.12±13.22
3	83.30±4.02	72.66±25.97	72.66±25.97	84.22±7.11	74.99±17.15
4	72.87±3.76	67.65±17.68	67.65±17.68	70.11±9.69	68.05±13.18

TABLE 12. Average scores with standard deviation generated by the different folds in the ResNet50v2 model trained with transfer learning from the ImageNet database. Scores generated with the test data.

k_fold	Accuracy	Sensitivity	Specificity	Precision	F1_score
0	90.59±3.46	87.90±8.30	87.90±8.30	88.03±7.93	87.83±7.48
1	87.00±3.77	86.16±7.17	86.16±7.17	80.59±15.82	82.52±10.12
2	90.23±3.34	86.49±10.38	86.49±10.38	88.86±6.20	87.41±7.50
3	90.21±3.02	85.34±12.46	85.34±12.46	89.76±4.74	87.05±7.73
4	84.44±4.04	79.04±17.92	79.04±17.92	85.81±6.29	80.90±9.95

TABLE 13. Average scores with standard deviation generated by the different folds in the ResNet50v2 model trained with learning transfer from the thermographic database. Scores generated with the test data.

k_fold	Accuracy	Sensitivity	Specificity	Precision	F1_score
0	89.98±3.06	86.35±10.26	86.35±10.26	87.85±7.15	86.82±7.73
1	87.90±3.59	86.17±8.68	86.17±8.68	81.82±14.39	83.36±10.07
2	89.52±2.98	86.77±8.14	86.77±8.14	87.17±6.95	86.85±7.02
3	90.18±3.34	85.27±12.40	85.27±12.40	89.91±5.49	87.06±7.72
4	85.83±3.30	80.83±16.25	80.83±16.25	87.33±5.75	82.78±8.46

where, the superscript (l) indicates the l-th layer and (l - 1) the previous layer.

B. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a method used to represent a set of observations in an equivalent, represented in latent variables of lower dimensionality [72]. The method takes the observations and combines them linearly to create new components. These components are generated hierarchically, each retaining a percentage of the variability of the original data. The first component, z₁, is created with the highest percentage of variability, while subsequent components take smaller percentages than the previous ones. In principle, the first principal component is expressed as shown in

Equation (7) or (8) in the form of a dot product.

$$z_1 = u_{11}x_1 + u_{12}x_2 + u_{13}x_3 + \dots + u_{1m}x_m \quad (33)$$

$$z_1 = U_1 \cdot X \quad (34)$$

In other words, let X be an observation of m variables ($X \in R^m$); this can be represented from a smaller number of latent variables Z, as shown in Equation (35).

$$Z = WX \quad (35)$$

$$W = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1m} \\ u_{21} & u_{22} & \dots & u_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \dots & u_{nm} \end{pmatrix} \quad (36)$$

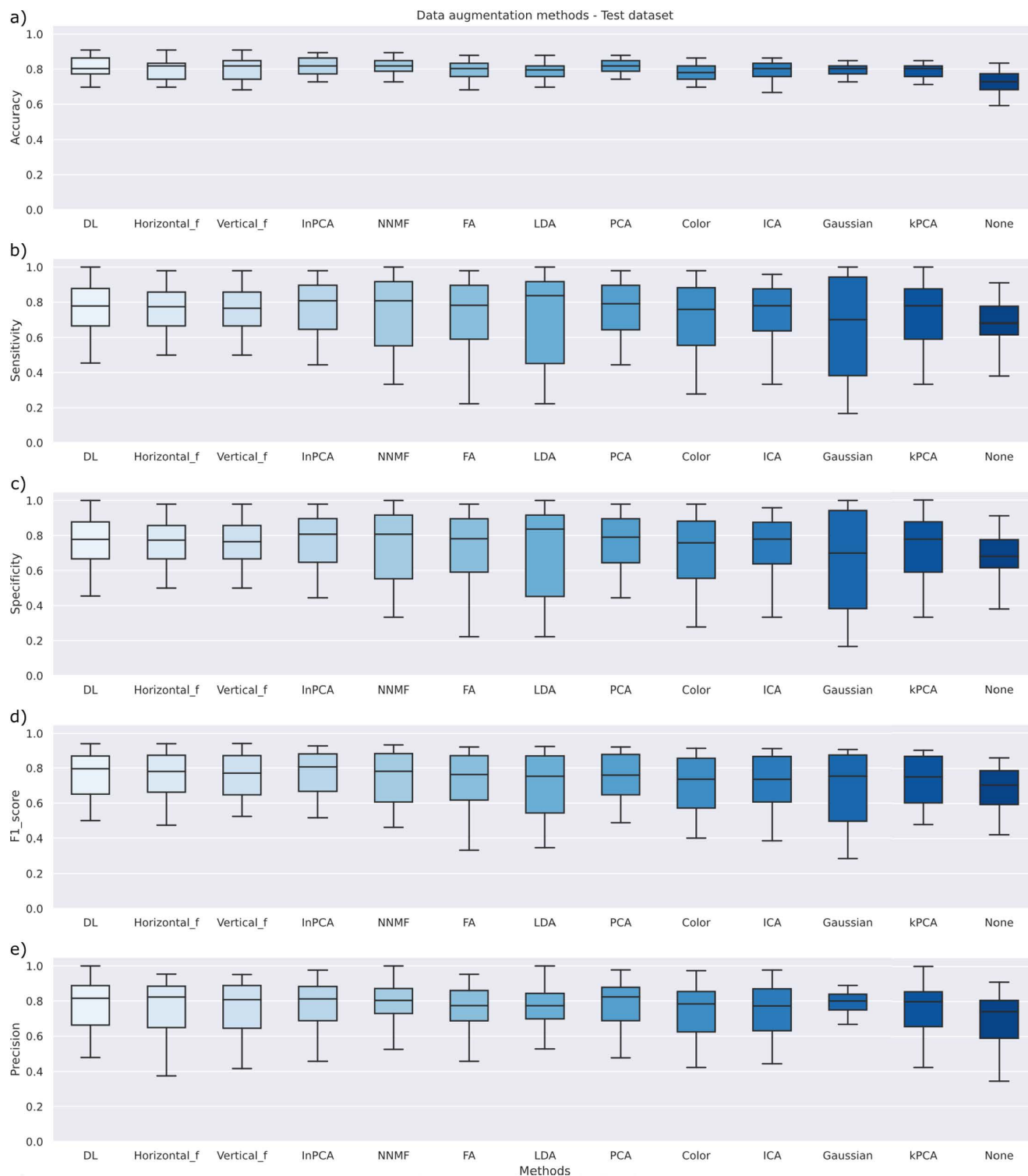


FIGURE 17. Distribution of the scores generated by the data augmentation methods in the ResNet50v2 model trained from scratch for the metrics of a) accuracy, b) sensitivity, c) specificity, d) sensitivity, and e) precision.

where, Z is the vector of n principal components, with $n < m$. In other words, Z^T is equal to $(z_1, z_2, z_3, \dots, z_n)$, with each principal component z_i being the linear combination of the

m original variables that make up X . W is the matrix of the coefficients of these linear combinations, which are calculated according to the following considerations: The first

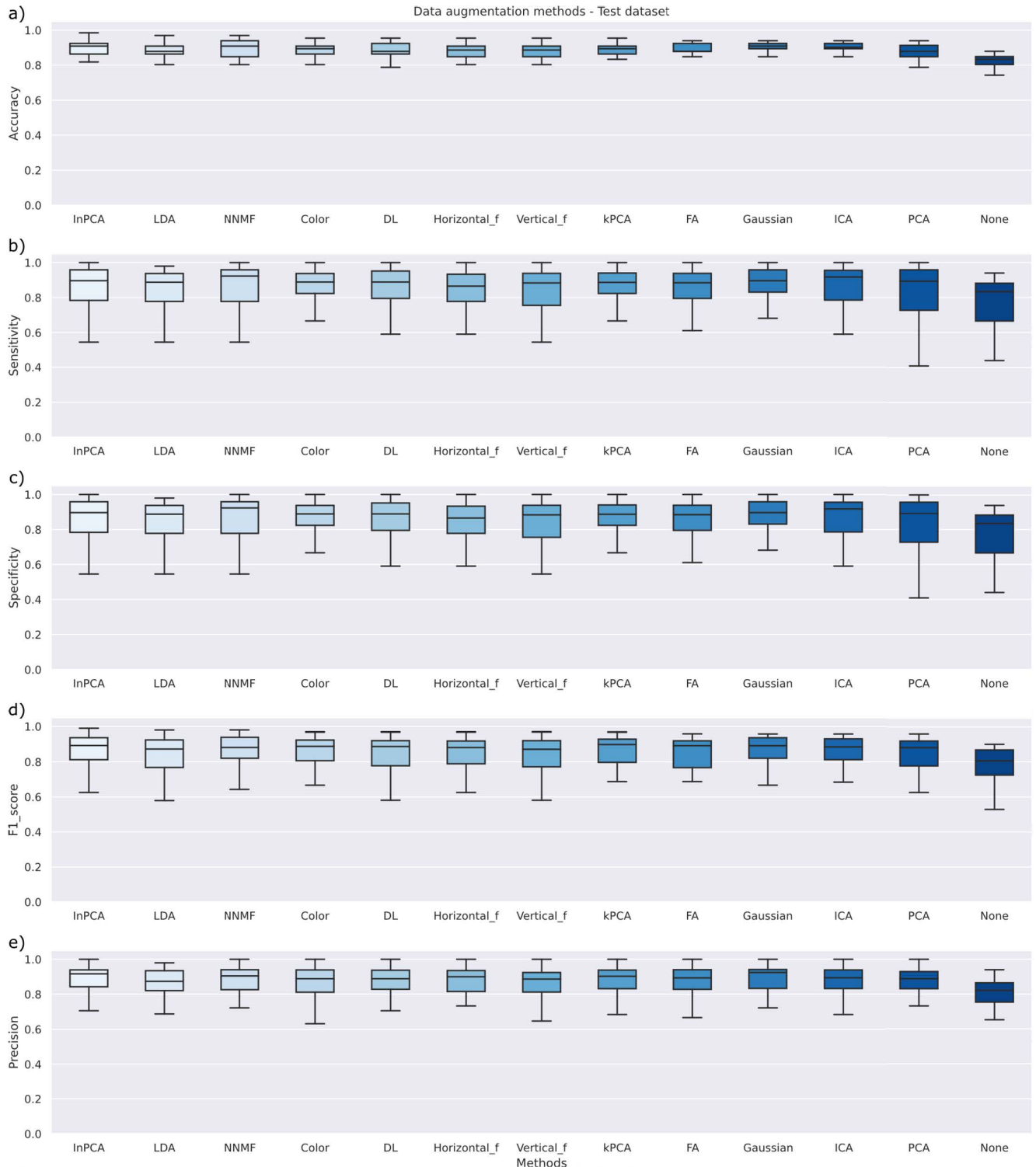


FIGURE 18. Distribution of scores generated by data augmentation methods on the ResNet50v2 model trained with transfer learning from the ImageNet database for the metrics of a) accuracy, b) sensitivity, c) specificity, d) sensitivity, and e) precision.

component z_1 must satisfy the maximum variance subject to the constraint of Equation (37).

$$U_1 U_1^t = \sum_{i=1}^m u_{1i} u_{1i} = 1 \quad (37)$$

The subsequent components are calculated under the same reasoning but considering that the new components must be orthogonal to the previous ones. Therefore, the i -th component must fulfill the restriction of Equation (38).

$$U_i U_j^t = 0 \quad \forall j < i \quad (38)$$

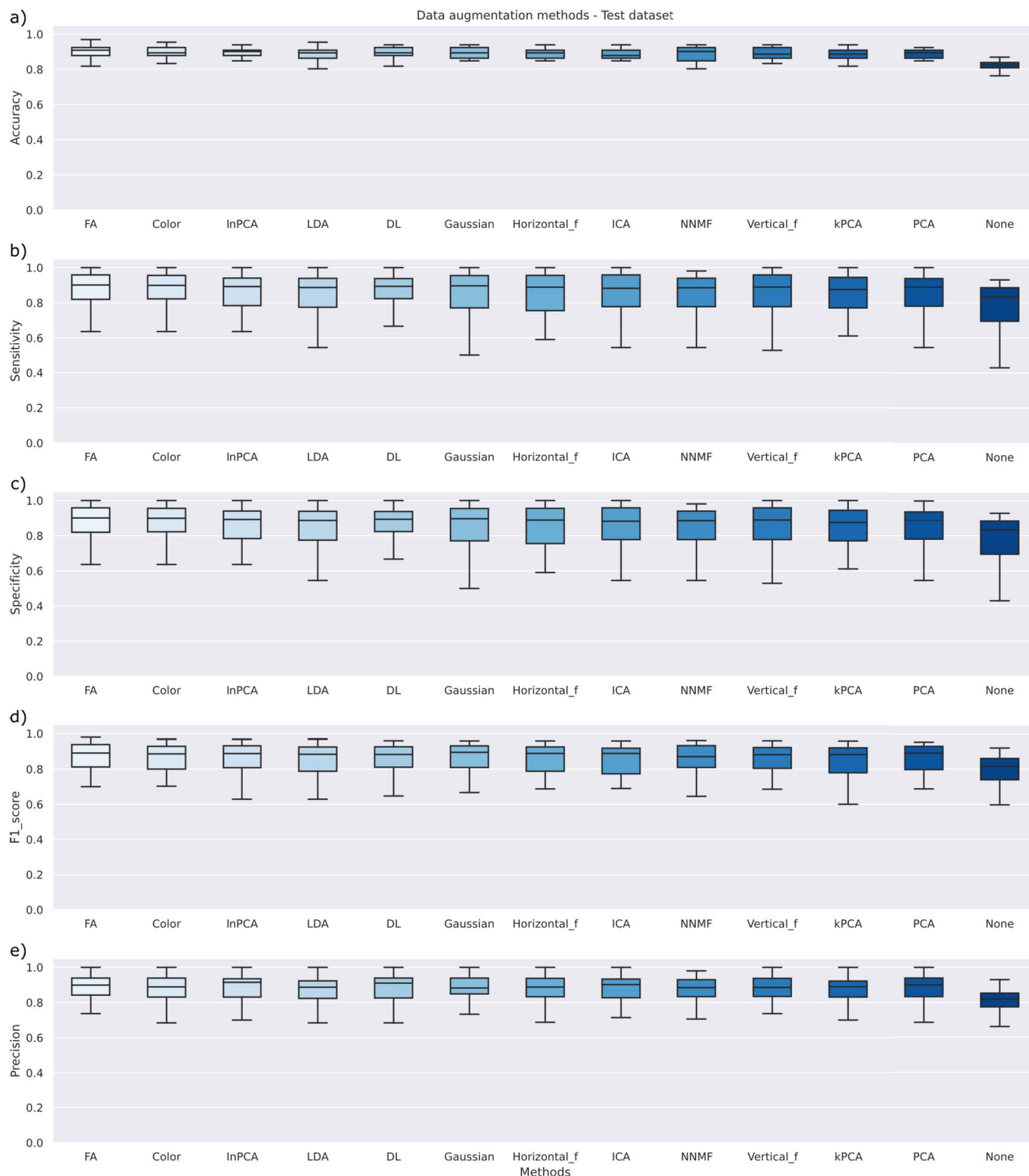


FIGURE 19. Distribution of the scores generated by the data augmentation methods in the ResNet50v2 model trained with learning transfer from the thermographic database for the metrics of a) accuracy, b) sensitivity, c) specificity, d) sensitivity, and e) precision.

C. MODEL TRAINING FOR TRANSFER LEARNING FROM EXTERNAL THERMOGRAPHIC DATA

As mentioned above, it was necessary to train the ResNet50v2 model for transfer learning with the thermal database to

generate the fitted weights and transfer them to the training model with the diabetic foot database. Similarly, to the training discussed in the materials and methods section, this network was trained under the same considerations.

Table 10 shows the mean values with the respective standard deviation of the accuracy, sensitivity, specificity, precision, and F1 score metrics.

Figure 15 presents the distribution of accuracy scores of the five folds in the cross-validation.

The results in Figure 15 show that the model achieved high performance with the test data, ensuring that there was no overtraining of the model. Figure 16, on the other hand, shows the training curve with the accuracy and loss metrics. The curves are the average with the 95% error band. However, the error between training is too small to be visible to the naked eye. In fact, Table 10 also shows this behavior, where the standard deviation for the five folds does not exceed 0.5%.

D. RESULTS OF CROSS-VALIDATION WITH 5-FOLDS

Similarly, the ResNet50v2 network was trained using cross-validation with 5-folds. However, the results do not show significant differences between the different folds. Therefore, it was decided to present these results in this section. Table 11 presents the training results from scratch for the ResNet50v2 model evaluated with 5-fold cross-validation. Similarly, Table 12 and Table 13 show the same results for the model trained with transfer learning. The first one uses the ImageNet database, and the second uses the thermographic database.

Similarly, Figure 17, Figure 18 and Figure 19 show the distribution of scores from the three training conditions, i.e., from scratch, with learning transfer from ImageNet and the thermographic database. The three plots show the distributions in fractional values, with values ranging from 0 to 1 equivalent to 0 and 100%. In addition, the figures show the accuracy, sensitivity, specificity, precision, and F1 score metrics as a function of the 13 training conditions. That is, under the eight newly proposed methods, the four conventional methods, and the model's training without data augmentation

E. GLOSSARY

ACM	Adaptive C Means
AI	Artificial Intelligence
ANN	Artificial Neural Network
BMI	Body Mass Index
CG	Control Group
CNN	Convolutional Neural Network
CSI	Critical Success Index
CTT	Classification Temperature Threshold
DD-DT-CWT	Double Density-Dual Tree-Complex Wavelet Transform
DFU	Diabetic Foot Ulcer
DL	Deep Learning
DM	Diabetes Mellitus
DT	Decision tree
DWT	Discrete Wavelet Transform
ELM	Extreme Learning Machine
ET	Extra Tree

FDR	False Discovery Rate
FOR	False Omission Rate
GB	Gradient Boosting
GBM	Gradient Boosting Machine
GLCM	Gray-Level Co-occurrence Matrix
HOS	Higher Order Spectra
IDF	International Diabetes Federation
IP	Image Processing
KNN	k nearest neighbors
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LR	Logistic regression
LSDA	Locality Sensitive Discriminant Analysis
LTE	Laws Texture Energy
MCC	Mattews Correlation Coefficient
ML	Machine Learning
MLP	Multi-Layer Perceptron
ND	Naïve Bayes
PCA	Principal component analysis
PNN	Probabilistic Neural Network
QDA	Quadratic Discriminant Analysis
RBF	Radial Basis Function
RF	Random Forest
RMSD	Root-Mean-Square Deviation
SVM	Support Vector Machine
TCI	Thermal Change Index
TS	Threat Score
WHO	World Health Organization
XGBoost	Extreme Gradient Boost
ZOL	Zero One Loss

REFERENCES

- [1] C. Gopalan and E. Kirk, "Diabetes mellitus," in *Biology of Cardiovascular and Metabolic Diseases*. Amsterdam, The Netherlands: Elsevier, 2022, pp. 223–243, doi: [10.1016/B978-0-12-823421-1.00015-9](https://doi.org/10.1016/B978-0-12-823421-1.00015-9).
- [2] A. Vaiserman and O. Lushchak, "Developmental origins of type 2 diabetes: Focus on epigenetics," *Ageing Res. Rev.*, vol. 55, Nov. 2019, Art. no. 100957, doi: [10.1016/j.arr.2019.100957](https://doi.org/10.1016/j.arr.2019.100957).
- [3] K. J. Bell, S. Colagiuri, and J. Brand-Miller, "Diabetes and insulin resistance," in *Present Knowledge in Nutrition*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 361–377, doi: [10.1016/B978-0-12-818460-8.00020-4](https://doi.org/10.1016/B978-0-12-818460-8.00020-4).
- [4] L. Ismail, H. Materwala, and J. Al Kaabi, "Association of risk factors with type 2 diabetes: A systematic review," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 1759–1785, Jan. 2021, doi: [10.1016/j.csbj.2021.03.003](https://doi.org/10.1016/j.csbj.2021.03.003).
- [5] A. A. Farooqui, "Insulin resistance, diabetes, and metabolic syndrome," in *Insulin Resistance as a Risk Factor in Visceral and Neurological Disorders*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 71–112, doi: [10.1016/B978-0-12-819603-8.00002-X](https://doi.org/10.1016/B978-0-12-819603-8.00002-X).
- [6] World Health Organization. (2021). *Diabetes*. Accessed: Nov. 16, 2021. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- [7] X. Lin, Y. Xu, X. Pan, J. Xu, Y. Ding, X. Sun, X. Song, Y. Ren, and P.-F. Shan, "Global, regional, and national burden and trend of diabetes in 195 countries and territories: An analysis from 1990 to 2025," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, Art. no. 14790, doi: [10.1038/s41598-020-71908-9](https://doi.org/10.1038/s41598-020-71908-9).
- [8] International Diabetes Federation. (2021). *IDF Diabetes Atlas 10th Edition 2021*. Accessed: Nov. 25, 2021. [Online]. Available: <https://diabetesatlas.org/data/en/>
- [9] K. C. Mekala and A. G. Bertoni, "Epidemiology of diabetes mellitus," in *Transplantation, Bioengineering, and Regeneration of the Endocrine Pancreas*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 49–58, doi: [10.1016/B978-0-12-814833-4.00004-6](https://doi.org/10.1016/B978-0-12-814833-4.00004-6).

- [10] E. L. Feldman, B. C. Callaghan, R. Pop-Busui, D. W. Zochodne, D. E. Wright, D. L. Bennett, V. Bril, J. W. Russell, and V. Viswanathan, "Diabetic neuropathy," *Nature Rev. Disease Primers*, vol. 5, no. 1, pp. 2–12, Dec. 2019, doi: [10.1038/s41572-019-0092-1](https://doi.org/10.1038/s41572-019-0092-1).
- [11] P. Naidoo, V. J. Liu, M. Mautone, and S. Bergin, "Lower limb complications of diabetes mellitus: A comprehensive review with clinicopathological insights from a dedicated high-risk diabetic foot multidisciplinary team," *Brit. J. Radiol.*, vol. 88, no. 1053, Sep. 2015, Art. no. 20150135, doi: [10.1259/bjr.20150135](https://doi.org/10.1259/bjr.20150135).
- [12] Annals of Internal Medicine, "Screening for type 2 diabetes mellitus in adults: Recommendations and rationale," *Ann. Internal Med.*, vol. 138, no. 3, p. 212, Feb. 2003, doi: [10.7326/0003-4819-138-3-200302040-00014](https://doi.org/10.7326/0003-4819-138-3-200302040-00014).
- [13] C. Childs, "Chapter 29—Body temperature and clinical thermometry," in *Thermoregulation: From Basic Neuroscience to Clinical Neurology, Part II*, vol. 157. Amsterdam, The Netherlands: Elsevier, 2018, pp. 467–482, [Online]. Available: <https://doi.org/10.1016/B978-0-444-64074-1.00029-X>, doi: [10.1016/B978-0-444-64074-1.00029-X](https://doi.org/10.1016/B978-0-444-64074-1.00029-X).
- [14] U. Thirunavukkarasu, S. Umopathy, K. Janardhanan, and R. Thirunavukkarasu, "A computer aided diagnostic method for the evaluation of type II diabetes mellitus in facial thermograms," *Phys. Eng. Sci. Med.*, vol. 43, no. 3, pp. 871–888, Sep. 2020, doi: [10.1007/s13246-020-00886-z](https://doi.org/10.1007/s13246-020-00886-z).
- [15] U. Thirunavukkarasu, S. Umopathy, P. T. Krishnan, and K. Janardanan, "Human tongue thermography could be a prognostic tool for pre-screening the type II diabetes mellitus," *Evidence-Based Complementary Alternative Med.*, vol. 2020, pp. 1–16, Jan. 2020, doi: [10.1155/2020/3186208](https://doi.org/10.1155/2020/3186208).
- [16] O. Faust, U. R. Acharya, E. Y. K. Ng, T. J. Hong, and W. Yu, "Application of infrared thermography in computer aided diagnosis," *Infr. Phys. Technol.*, vol. 66, pp. 160–175, Sep. 2014, doi: [10.1016/j.infrared.2014.06.001](https://doi.org/10.1016/j.infrared.2014.06.001).
- [17] R. Vardasca, C. Magalhaes, and J. Mendes, "Biomedical applications of infrared thermal imaging: Current state of machine learning classification," *Multidisciplinary Digit. Publishing Inst. Proc.*, vol. 27, no. 1, p. 46, Oct. 2019, doi: [10.3390/proceedings2019027046](https://doi.org/10.3390/proceedings2019027046).
- [18] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, and F. Ciampa, "Infrared machine vision and infrared thermography with deep learning: A review," *Infr. Phys. Technol.*, vol. 116, Aug. 2021, Art. no. 103754, doi: [10.1016/j.infrared.2021.103754](https://doi.org/10.1016/j.infrared.2021.103754).
- [19] J. Tulloch, R. Zamani, and M. Akrami, "Machine learning in the prevention, diagnosis and management of diabetic foot ulcers: A systematic review," *IEEE Access*, vol. 8, pp. 198977–199000, 2020, doi: [10.1109/ACCESS.2020.3035327](https://doi.org/10.1109/ACCESS.2020.3035327).
- [20] S. B. Gururajarao, U. Venkatappa, J. M. Shivaram, M. Y. Sikkandar, and A. Al Amoudi, "Infrared thermography and soft computing for diabetic foot assessment," in *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*, N. Dey, S. Borra, A. S. Ashour, and D. I. Shi, Eds. New York, NY, USA: Academic, 2019, pp. 73–97.
- [21] M. Adam, E. Y. K. Ng, J. H. Tan, M. L. Heng, J. W. K. Tong, and R. Acharya, "Computer aided diagnosis of diabetic foot using infrared thermography: A review," *Comput. Biol. Med.*, vol. 91, pp. 326–336, Dec. 2017, doi: [10.1016/j.combiomed.2017.10.030](https://doi.org/10.1016/j.combiomed.2017.10.030).
- [22] S. Zahia, M. B. G. Zapirain, X. Sevillano, A. González, P. J. Kim, and A. Elmaghraby, "Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods," *Artif. Intell. Med.*, vol. 102, Jan. 2020, Art. no. 101742, doi: [10.1016/j.artmed.2019.101742](https://doi.org/10.1016/j.artmed.2019.101742).
- [23] R. Usharani and A. Shanthini, "Neuropathic complications: Type II diabetes mellitus and other risky parameters using machine learning algorithms," *J. Ambient Intell. Humanized Comput.*, Mar. 2021, doi: [10.1007/s12652-021-02972-w](https://doi.org/10.1007/s12652-021-02972-w).
- [24] B. J. Petersen, M. Zhou, J. D. Bloom, G. M. Rothenberg, and D. Armstrong, "642-P: Diagnosing peripheral neuropathy with a remote temperature monitoring mat," *Diabetes*, vol. 68, no. 1, p. 642, Jun. 2019, doi: [10.2337/db19-642-P](https://doi.org/10.2337/db19-642-P).
- [25] R. Bayareh, A. Vera, L. Leija, and J. Gutierrez-Martinez, "Programming of a system for the acquisition of images and thermographic data for the diabetic foot analysis," in *Proc. 14th Int. Conf. Electr. Eng., Comput. Sci. Autom. Control (CCE)*, Oct. 2017, pp. 1–7, doi: [10.1109/ICEEE.2017.8108866](https://doi.org/10.1109/ICEEE.2017.8108866).
- [26] S. Sruthi and M. Sasikala, "A low cost thermal imaging system for medical diagnostic applications," in *Proc. Int. Conf. Smart Technol. Manage. Comput., Commun., Controls, Energy Mater. (ICSTM)*, May 2015, pp. 621–623, doi: [10.1109/ICSTM.2015.7225488](https://doi.org/10.1109/ICSTM.2015.7225488).
- [27] N. C. Evangeline, E. Suresh, and S. Srinivasan, "Computer-aided appraisal of plantar foot health in diabetic subjects using infra-red thermograms," in *Proc. 5th Int. Conf. Comput., Commun. Secur. (ICCCS)*, Oct. 2020, pp. 1–5, doi: [10.1109/ICCCS49678.2020.9276816](https://doi.org/10.1109/ICCCS49678.2020.9276816).
- [28] V. Filipe, P. Teixeira, and A. Teixeira, "A clustering approach for prediction of diabetic foot using thermal images," 2020, pp. 620–631, doi: [10.1007/978-3-030-58808-3_45](https://doi.org/10.1007/978-3-030-58808-3_45).
- [29] S. M. Prabhu and S. Verma, "Automatic segmentation of plantar thermograms using adaptive c means technique," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 11, no. 2, p. 1250, Apr. 2021, doi: [10.11591/ijece.v11i2.pp1250-1258](https://doi.org/10.11591/ijece.v11i2.pp1250-1258).
- [30] L. C. Padierna, L. F. Amador-Medina, B. O. Murillo-Ortiz, and C. Villaseñor-Mora, "Classification method of peripheral arterial disease in patients with type 2 diabetes mellitus by infrared thermography and machine learning," *Infr. Phys. Technol.*, vol. 111, Dec. 2020, Art. no. 103531, doi: [10.1016/j.infrared.2020.103531](https://doi.org/10.1016/j.infrared.2020.103531).
- [31] M. Adam, "Automated characterization of diabetic foot using nonlinear features extracted from thermograms," *Infr. Phys. Technol.*, vol. 89, pp. 325–337, Mar. 2018, doi: [10.1016/j.infrared.2018.01.022](https://doi.org/10.1016/j.infrared.2018.01.022).
- [32] M. Adam, E. Y. K. Ng, S. L. Oh, M. L. Heng, Y. Hagiwara, J. H. Tan, J. W. K. Tong, and U. R. Acharya, "Automated detection of diabetic foot with and without neuropathy using double density-dual tree-complex wavelet transform on foot thermograms," *Infr. Phys. Technol.*, vol. 92, pp. 270–279, Aug. 2018, doi: [10.1016/j.infrared.2018.06.010](https://doi.org/10.1016/j.infrared.2018.06.010).
- [33] J. Selle, K. V. M. V. Prakash, G. A. Sai, B. Vinod, and K. Chellappan, "Classification of foot thermograms using texture features and support vector machine," in *Proc. 2nd Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, Aug. 2021, pp. 1445–1449, doi: [10.1109/ICESC51422.2021.9532777](https://doi.org/10.1109/ICESC51422.2021.9532777).
- [34] S. Reddy, G. Mahesh, and N. Preethi, "Exploiting machine learning algorithms to diagnose foot ulcers in diabetic patients," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 7, no. 29, Jul. 2018, Art. no. 170752, doi: [10.4108/eai.24-8-2021.170752](https://doi.org/10.4108/eai.24-8-2021.170752).
- [35] U. Nag, M. Upadhyay, and T. Gupta, "Detecting diabetic foot complications using infrared thermography and machine learning," in *Proc. 5th Int. Conf. Graph. Signal Process.*, Jun. 2021, pp. 41–46, doi: [10.1145/3474906.3474919](https://doi.org/10.1145/3474906.3474919).
- [36] U. Thirunavukkarasu and S. Umopathy, "Classification of prediabetes and healthy subjects in plantar infrared thermal imaging using various machine learning algorithms," 2020, pp. 85–96, doi: [10.1007/978-981-15-2329-8_9](https://doi.org/10.1007/978-981-15-2329-8_9).
- [37] R. Vardasca, C. Magalhaes, A. Seixas, R. Carvalho, and J. Mendes, "Diabetic foot monitoring using dynamic thermography and AI classifiers," in *Proc. QIRT Asia*, 2019, pp. 1–5, doi: [10.21611/qirt.2019.027](https://doi.org/10.21611/qirt.2019.027).
- [38] I. Cruz-Vega, H. Peregrina-Barreto, J. de Jesus Rangel-Magdaleno, and J. M. Ramirez-Cortes, "A comparison of intelligent classifiers of thermal patterns in diabetic foot," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2019, pp. 1–6, doi: [10.1109/I2MTC.2019.8827044](https://doi.org/10.1109/I2MTC.2019.8827044).
- [39] I. Cruz-Vega, D. Hernandez-Contreras, H. Peregrina-Barreto, J. D. J. Rangel-Magdaleno, and J. M. Ramirez-Cortes, "Deep learning classification for diabetic foot thermograms," *Sensors*, vol. 20, no. 6, p. 1762, Mar. 2020, doi: [10.3390/s20061762](https://doi.org/10.3390/s20061762).
- [40] A. Khandakar, M. E. H. Chowdhury, M. B. I. Reaz, S. H. M. Ali, M. A. Hasan, S. Kiranyaz, T. Rahman, R. Alfkey, A. A. Bakar, and R. A. Malik, "A machine learning model for early detection of diabetic foot using thermogram images," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104838, doi: [10.1016/j.combiomed.2021.104838](https://doi.org/10.1016/j.combiomed.2021.104838).
- [41] A. Khandakar, M. E. H. Chowdhury, M. B. I. Reaz, S. H. M. Ali, T. O. Abbas, T. Alam, M. A. Ayari, Z. B. Mahbub, R. Habib, T. Rahman, A. M. Tahir, A. A. A. Bakar, and R. A. Malik, "Thermal change index-based diabetic foot thermogram image classification using machine learning techniques," *Sensors*, vol. 22, no. 5, p. 1793, Feb. 2022, doi: [10.3390/s22051793](https://doi.org/10.3390/s22051793).
- [42] T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Deep learning for diabetes: A systematic review," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2744–2757, Jul. 2021, doi: [10.1109/JBHI.2020.3040225](https://doi.org/10.1109/JBHI.2020.3040225).
- [43] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G. Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017, doi: [10.1109/JBHI.2016.2636665](https://doi.org/10.1109/JBHI.2016.2636665).
- [44] R. Wason, "Deep learning: Evolution and expansion," *Cognit. Syst. Res.*, vol. 52, pp. 701–708, Dec. 2018, doi: [10.1016/j.cogsys.2018.08.023](https://doi.org/10.1016/j.cogsys.2018.08.023).

- [45] I. Castiglioni, L. Rundo, M. Codari, G. D. Leo, C. Salvatore, M. Interlenghi, F. Gallivanone, A. Cozzi, N. C. D'Amico, and F. Sardanelli, "AI applications to medical images: From machine learning to deep learning," *Phys. Medica*, vol. 83, pp. 9–24, Mar. 2021, doi: [10.1016/j.ejmp.2021.02.006](https://doi.org/10.1016/j.ejmp.2021.02.006).
- [46] E. Ahishakiye, M. B. Van Gijzen, J. Tumwiine, R. Wario, and J. Obungoloch, "A survey on deep learning in medical image reconstruction," *Intell. Med.*, vol. 1, no. 3, pp. 118–127, Sep. 2021, doi: [10.1016/j.imed.2021.03.003](https://doi.org/10.1016/j.imed.2021.03.003).
- [47] X. Chen, X. Wang, K. Zhang, K.-M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu, "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102444, doi: [10.1016/j.media.2022.102444](https://doi.org/10.1016/j.media.2022.102444).
- [48] A. Echle, N. T. Rindtorff, T. J. Brinker, T. Luedde, A. T. Pearson, and J. N. Kather, "Deep learning in cancer pathology: A new generation of clinical biomarkers," *Brit. J. Cancer*, vol. 124, no. 4, pp. 686–696, Feb. 2021, doi: [10.1038/s41416-020-01122-x](https://doi.org/10.1038/s41416-020-01122-x).
- [49] M. S. Iqbal, I. Ahmad, T. Khan, S. Khan, M. Ahmad, and L. Wang, "Recent advances of deep learning in biology," 2021, pp. 709–732, doi: [10.1007/978-3-030-77939-9_21](https://doi.org/10.1007/978-3-030-77939-9_21).
- [50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826, doi: [10.1109/CVPR.2016.308](https://doi.org/10.1109/CVPR.2016.308).
- [52] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269, doi: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [53] F. Shamsad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, "Transformers in medical imaging: A survey," 2022, *arXiv:2201.09873*.
- [54] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jun. 2014. [Online]. Available: <http://jmlr.org/papers/v15/srivastava14a.html>
- [55] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 1, Feb. 2015, pp. 448–456.
- [56] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, p. 9, Dec. 2016, doi: [10.1186/s40537-016-0043-6](https://doi.org/10.1186/s40537-016-0043-6).
- [57] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, p. 60, Jul. 2019, doi: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0).
- [58] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *J. Med. Imag. Radiat. Oncol.*, vol. 65, no. 5, pp. 545–563, Aug. 2021, doi: [10.1111/1754-9485.13261](https://doi.org/10.1111/1754-9485.13261).
- [59] A. Anaya-Isaza and L. Mera-Jimenez, "Data augmentation and transfer learning for brain tumor detection in magnetic resonance imaging," *IEEE Access*, vol. 10, pp. 23217–23233, 2022, doi: [10.1109/ACCESS.2022.3154061](https://doi.org/10.1109/ACCESS.2022.3154061).
- [60] G. Szlobodnyik and L. Farkas, "Data augmentation by guided deep interpolation," *Appl. Soft Comput.*, vol. 111, Nov. 2021, Art. no. 107680, doi: [10.1016/j.asoc.2021.107680](https://doi.org/10.1016/j.asoc.2021.107680).
- [61] R. Raj, J. Mathew, S. K. Kannath, and J. Rajan, "Crossover based technique for data augmentation," *Comput. Methods Programs Biomed.*, vol. 218, May 2022, Art. no. 106716, doi: [10.1016/j.cmpb.2022.106716](https://doi.org/10.1016/j.cmpb.2022.106716).
- [62] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Mach. Learn. With Appl.*, vol. 6, Dec. 2021, Art. no. 100134, doi: [10.1016/j.mlwa.2021.100134](https://doi.org/10.1016/j.mlwa.2021.100134).
- [63] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, Dec. 2021, doi: [10.1186/s40537-021-00444-8](https://doi.org/10.1186/s40537-021-00444-8).
- [64] D. A. Hernandez-Contreras, H. Peregrina-Barreto, J. D. J. Rangel-Magdaleno, and F. J. Renero-Carrillo, "Plantar thermogram database for the study of diabetic foot complications," *IEEE Access*, vol. 7, pp. 161296–161307, 2019, doi: [10.1109/ACCESS.2019.2951356](https://doi.org/10.1109/ACCESS.2019.2951356).
- [65] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9. [Online]. Available: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- [66] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [67] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision—ECCV 2016*. Cham, Switzerland: Springer, 2016, pp. 630–645, doi: [10.1007/978-3-319-46493-0_38](https://doi.org/10.1007/978-3-319-46493-0_38).
- [68] D. R. Bull and F. Zhang, "Color spaces and color transformations," in *Intelligent Image and Video Compression*, Amsterdam, The Netherlands: Elsevier, 2021, pp. 107–142.
- [69] Y. Yang, P. Yuhua, and L. Zhaoguang, "A fast algorithm for YCbCr to RGB conversion," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1490–1493, Nov. 2007, doi: [10.1109/TCE.2007.4429242](https://doi.org/10.1109/TCE.2007.4429242).
- [70] W. Chen, Y. Q. Shi, and G. Xuan, "Identifying computer graphics using HSV color model and statistical moments of characteristic functions," in *Proc. IEEE Multimedia Expo. Int. Conf.*, Jul. 2007, pp. 1123–1126, doi: [10.1109/ICME.2007.4284852](https://doi.org/10.1109/ICME.2007.4284852).
- [71] L. Tan and J. Jiang, "Image filtering enhancement," in *Digital Signal Processing*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 649–726.
- [72] F. G. Ashby, "Principal components analysis," in *Statistical Analysis of fMRI Data*. Cambridge, MA, USA: MIT Press 2011, pp. 245–256.
- [73] A. F. M. Alkarkhi and W. A. A. Alqaragholi, "Factor Analysis," in *Applied Statistics for Environmental Science With R*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 151–171, doi: [10.1016/B978-0-12-818622-0.00009-5](https://doi.org/10.1016/B978-0-12-818622-0.00009-5).
- [74] L. P. Fávero and P. Belfiore, "Principal component factor analysis," in *Data Science for Business and Decision Making*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 383–438, doi: [10.1016/B978-0-12-811216-8.00012-4](https://doi.org/10.1016/B978-0-12-811216-8.00012-4).
- [75] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, Jun. 2000, doi: [10.1016/S0893-6080\(00\)00026-5](https://doi.org/10.1016/S0893-6080(00)00026-5).
- [76] J.-T. Chien, "Independent component analysis," in *Source Separation and Machine Learning*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 99–160, doi: [10.1016/B978-0-12-804566-4.00016-4](https://doi.org/10.1016/B978-0-12-804566-4.00016-4).
- [77] J.-T. Chien, "Nonnegative matrix factorization," in *Source Separation and Machine Learning*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 161–229, doi: [10.1016/B978-0-12-804566-4.00017-6](https://doi.org/10.1016/B978-0-12-804566-4.00017-6).
- [78] C. Boutsidis and E. Gallopoulos, "SVD based initialization: A head start for nonnegative matrix factorization," *Pattern Recognit.*, vol. 41, no. 4, pp. 1350–1362, Apr. 2008, doi: [10.1016/j.patcog.2007.09.010](https://doi.org/10.1016/j.patcog.2007.09.010).
- [79] W. Ford, "Vector and matrix norms," in *Numerical Linear Algebra With Applications*. Amsterdam, The Netherlands: Elsevier, 2015, pp. 119–144, doi: [10.1016/B978-0-12-394435-1.00007-7](https://doi.org/10.1016/B978-0-12-394435-1.00007-7).
- [80] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML)*, 2009, pp. 1–8, doi: [10.1145/1553374.1553463](https://doi.org/10.1145/1553374.1553463).
- [81] T. Tong, J. Caballero, K. Bhatia, and D. Rueckert, "Dictionary learning for medical image denoising, reconstruction, and segmentation," in *Machine Learning and Medical Imaging*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 153–181, doi: [10.1016/B978-0-12-804076-8.00006-2](https://doi.org/10.1016/B978-0-12-804076-8.00006-2).
- [82] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
- [83] A. M. Jade, B. Srikanth, V. K. Jayaraman, B. D. Kulkarni, J. P. Jog, and L. Priya, "Feature extraction and denoising using kernel PCA," *Chem. Eng. Sci.*, vol. 58, no. 19, pp. 4441–4448, Oct. 2003, doi: [10.1016/S0009-2509\(03\)00340-3](https://doi.org/10.1016/S0009-2509(03)00340-3).
- [84] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191).
- [85] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowl.-Based Syst.*, vol. 80, pp. 14–23, May 2015, doi: [10.1016/j.knosys.2015.01.010](https://doi.org/10.1016/j.knosys.2015.01.010).

- [86] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," 2014, *arXiv:1409.0575*.
- [87] Q. Ashfaq, U. Akram, and R. Zafar, "Thermal image dataset for object classification," *Mendeley Data*, to be published, doi: [10.17632/btm-rycjbj.1](https://doi.org/10.17632/btm-rycjbj.1).
- [88] A.-M. Šimundić, "Measures of diagnostic accuracy: Basic definitions," *EJIFCC*, vol. 19, no. 4, pp. 203–211, Jan. 2009. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27683318>
- [89] R. Trevethan, "Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice," *Frontiers Public Health*, vol. 5, p. 307, Nov. 2017, doi: [10.3389/fpubh.2017.00307](https://doi.org/10.3389/fpubh.2017.00307).
- [90] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, *arXiv:2010.16061*.
- [91] A. F. M. Alkarkhi, "The observed significance level (P-value) procedure," in *Applications of Hypothesis Testing for Environmental Science*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 79–119, doi: [10.1016/B978-0-12-824301-5.00010-1](https://doi.org/10.1016/B978-0-12-824301-5.00010-1).
- [92] P. E. McKight and J. Najab, "Kruskal–Wallis test," in *The Corsini Encyclopedia of Psychology*. Hoboken, NJ, USA: Wiley, Jan. 2010, doi: [10.1002/9780470479216.corpsy0491](https://doi.org/10.1002/9780470479216.corpsy0491).
- [93] E. Ostertagova, O. Ostertag, and J. Kováč, "Methodology and application of the Kruskal–Wallis test," *Appl. Mech. Mater.*, vol. 611, pp. 115–120, Aug. 2014, doi: [10.4028/www.scientific.net/AMM.611.115](https://doi.org/10.4028/www.scientific.net/AMM.611.115).
- [94] P. V. Tsaklis and N. Tentolouris, "Biomechanics of the diabetic foot: The road to foot ulceration," 2018, pp. 199–217, doi: [10.1007/978-3-319-89869-8_12](https://doi.org/10.1007/978-3-319-89869-8_12).
- [95] R. Reardon, D. Simring, B. Kim, J. Mortensen, D. Williams, and A. Leslie, "The diabetic foot ulcer," *Aust. J. Gen. Pract.*, vol. 49, no. 5, pp. 250–255, 2020. [Online]. Available: <https://search.informit.org/doi/10.3316/informit.110922096526517>, doi: [10.3316/informit.110922096526517](https://doi.org/10.3316/informit.110922096526517).
- [96] C. Setacci, D. Benevento, G. D. Donato, E. Viviani, U. M. Bracale, L. D. Guercio, C. Palasciano, and F. Setacci, "Focusing on diabetic ulcers," *Transl. Med. UniSa*, vol. 21, pp. 7–9, Feb. 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32123673>
- [97] F.-J. Renero-C, "The abrupt temperature changes in the plantar skin thermogram of the diabetic patient: Looking in to prevent the insidious ulcers," *Diabetic Foot Ankle*, vol. 9, no. 1, Jan. 2018, Art. no. 1430950, doi: [10.1080/2000625X.2018.1430950](https://doi.org/10.1080/2000625X.2018.1430950).
- [98] B. G. Childs, H. Li, and J. M. van Deursen, "Senescent cells: A therapeutic target for cardiovascular disease," *J. Clin. Invest.*, vol. 128, no. 4, pp. 1217–1228, Apr. 2018, doi: [10.1172/JCI95146](https://doi.org/10.1172/JCI95146).
- [99] S. A. Bus, D. G. Armstrong, R. W. van Deursen, J. E. A. Lewis, C. F. Caravaggi, and P. R. Cavanagh, "IWGDF guidance on footwear and offloading interventions to prevent and heal foot ulcers in patients with diabetes," *Diabetes/Metabolism Res. Rev.*, vol. 32, pp. 25–36, Jan. 2016, doi: [10.1002/dmrr.2697](https://doi.org/10.1002/dmrr.2697).
- [100] A. J. Singer, A. Tassiopoulos, and R. S. Kirsner, "Evaluation and management of lower-extremity ulcers," *New England J. Med.*, vol. 378, no. 3, pp. 302–303, Jan. 2018, doi: [10.1056/NEJMc1715237](https://doi.org/10.1056/NEJMc1715237).
- [101] B. H. M. Tan, A. Nather, and V. David, "Biomechanics of the foot," in *Diabetic Foot Problems*. Singapore: World Scientific, May 2008, pp. 67–75, doi: [10.1142/9789812791535_0006](https://doi.org/10.1142/9789812791535_0006).

•••