Learning and Nonlinear Models - Journal of the Brazilian Society on Artificial Intelligence (SBIA), Vol. 24, Iss. 1, pp. 60–73, 2026
https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

FOOTSTEP IDENTIFICATION ON PRESSURE SENSITIVE MAT THROUGH CONVOLUTIONAL NEURAL NETWORKS

Natália Soares Girão, Márcia Muller, and Łúcia Valéria Arruda
Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba-Pr 80230-901, Brazil
{ngirao,mmuller,lvrarruda}@utfpr.edu.br

Resumo – Os sistemas de identificação biométrica de usuários reconhecem as características biológicas ou comportamentais únicas de cada pessoa, a fim de identificá-las exclusivamente. Para sistemas com requisitos críticos de segurança, a identificação do usuário requer métodos biométricos robustos e consistentes. Este trabalho desenvolve um método para identificação biométrica da pisada humana baseado em sinais de sensores de pressão instalados em um tapete e redes neurais convolucionais. O objetivo é construir um sistema seguro para acesso a áreas restritas. O equipamento de aquisição de sinais biométricos é um tapete instrumentado com redes de Bragg apto a capturar os sinais óticos associados à distribuição de pressão plantar. Um algoritmo de pré-processamento transforma os deslocamentos dos comprimentos de ondas dos sinais óticos obtidos em um mapa de pressão plantar, extraindo em seguida duas imagens menores, uma para cada pé do usuário. As imagens são usadas para treinar indepedentemente duas redes neurais convolucionais (CNN): uma para o pé esquerdo e outra para o pé direito. Depois de treinadas, as duas CNNs são usadas como entrada para uma rede plenamente conectada (FNN) cujo pesos são treinados para a correta identificação do usuário. O modelo baseado em dois estágios de redes neurais (CNN-FNN) integram um sistema biométrico capaz de identificar usuários a partir dos sinais obtidos sobre o tapete ótico. O sistema desenvolvido alcançou uma acurácia de 98.67% para um pequeno grupo de usuários, mostrando ser uma ferramenta eficiente que pode ser incluida em sistemas biométrios de seguranca.

Palavras-chave – Identificação de pisada, Mapa de pressão, Redes de Bragg, Redes Neurais Convolucionais.

Abstract – Personal identification through biometric systems recognizes a person's unique biological or behavioral characteristics. For high-security systems, user identification requires reliable and robust biometric methods. This work develops a method for footstep biometric recognition, based on pressure-sensor signals and convolutional neural networks, for access to restricted areas. The biometric acquisition device is a platform instrumented with fiber Bragg gratings to capture optical signals associated with the plantar pressure distribution. A preprocessing algorithm transforms the gratings' wavelength shifts in the acquired signals into a plantar pressure map, separating the left and right feet into two smaller images. These small images are used to independently train two convolutional neural networks, one for each foot. After training, both CNNs are used as input to a fully connected neural network (FNN), forming a two-stage neural network (CNN-FNN) model whose weights are trained to identify a person correctly. This CNN-FNN biometric system identifies users from the complete plantar pressure map comprising both feet. The developed system achieves 98.67% accuracy, making it an efficient tool for inclusion in biometric security systems.

Keywords – Footstep Identification, Pressure Map, Fiber Bragg Gratings, Convolutional Neural Network.

1 INTRODUCTION

There is currently a wide range of biometric systems for identifying users in commercial, government, forensics, and security applications. They use electronic devices, such as digital cameras, scanners, cell phones, pressure sensors, and others, to recognize one or more biometric characteristics (fingerprint, iris scan, voice, gait, etc) of a person, depending on the degree of reliability, universality, and performance desired for the identification of the user [1]. Each biometric technology has its strengths and weaknesses, and choosing one or more biometrics for designing an identification system depends on the application. One high-security application is the access control to restricted areas, such as nanotechnology or biological laboratories, silicon industries, nuclear plants, intensive care units, etc. For this access, the biometric systems must authenticate authorized users by comparing captured biometric data with their biometric template(s) stored in the system database, and deny access to individuals not registered in the database [2].

The reliability of biometric security systems is often compromised by challenges such as noisy sensor data, non-universal biometric characteristics, vulnerability to spoof attacks, and enrollment failures. To enhance security, such systems require multiple enrollments of the same biometric or a combination of different biometrics (like iris and voice) [3,4]. However, acquiring multiple biometric characteristics can be tedious and user-unfriendly. Moreover, the need for multiple enrollments for security access complicates user interaction, reducing their comfort with the system. Touchless biometric systems, such as those based on facial or voice recognition, can offer a more user-friendly and hygienic experience.

While convolutional neural networks (CNNs) and machine learning algorithms have improved the performance of biometric systems, their accuracy and efficiency still depend on large-scale training datasets [5,6]. Thus, biometric systems based on these

https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

algorithms often require multiple enrollments, which can be tedious and uncomfortable for users.

In this context, this work proposes developing a biometric access control system based on footstep. Even though forensic science analyzes footwear impressions and barefoot prints at crime scenes to recover footprints and identify criminals, footstep-based biometrics is a considerably newer technique for personal identification in access control systems [7].

This work's biometric access control system uses the sensing platform described in [8] and improves the data analysis method.

Our proposed system's enrollment is user-friendly via a sensing platform, where the user steps in front of the room access door and generates a pressure map, from which two convolutional neural networks (CNNs) (one for each foot) extract features to identify the stepping person. The database used to train the neural network models is expanded using data augmentation techniques to reduce the need for multiple enrollment sessions to create a person's template, mainly to reduce the risk of overfitting and improve its ability to generalize to new examples. Once trained, the two CNNs serve as input to a new sequence of fully connected layers that identify users based on their two-foot match. Checking both feet adds reliability, improving the model's accuracy. The weights of new fully connected layers are trained using stochastic gradient descent. These three fully connected layers form a matching network, responsible for person identification. Then, this two-stage neural network (CNN-FNN) model, composed of two CNNs and the matching network, is pruned as a result of a computational cost optimization study, generating the final user identification model.

We organize this work as follows: Section 2 introduces related work that utilizes footprints and footsteps for identification; Section 3 describes the developed prototype schemes for biometric footstep authentication; Section 4 presents the results, which enable us to evaluate the properties of our biometric system, including accuracy and performance. Section 5 presents the conclusion and outlines future work needed to promote the use of our biometric system for identifying access to security rooms.

2 RELATED WORKS

There are several works in the literature on footprint-based identification, most of which concern forensic applications that require determining a criminal's footprints in environments with heavy foot traffic. In general, these studies perform identification based on the shape of the footprint, minutiae, geometric details, and similar features from an image [9, 10]. The recent work in [11] provides a detailed survey of these studies and compares them. The study in [12] also surveys recent advances in footprint-based individual recognition. All cited works in [11, 12] use footprint images acquired with dactyloscopy or cameras or scanning of a foot and apply various feature extraction and pattern recognition methods. Among these methods, the most used are Euclidean distance, principal component analysis (PCA), independent component analysis (ICA), hidden Markov model (HMM), self-organizing maps (SOM), support vector machine (SVM), binarization, artificial neural networks (ANN), wavelets, etc. Moreover, [12] covers various aspects of improving this biometric modality, including the challenges posed by aging and changes in posture over time.

On the other hand, information about the force exerted by the feet on the ground is also a person's biometric characteristic. Based on this information, biometric recognition is referred to in the literature as Footstep recognition [13]. Footstep recognition generally uses ground force sensors [7, 14–16] or pressure sensors [17–20], as in this work. Recently, vibration sensors and audio sensors have been installed in the floor, enabling footstep recognition [21–23].

Concerning the works using plantar pressure sensing for enrollment, we have unwearable [8, 20, 24] and wearable sensing systems [25–28]. The wearable systems are used to support continual and real-time person verification in the background. They support healthcare and wellness applications and generate a hard-to-manage volume of personal data about userśdaily lives [26]. The unwearable footstep biometric sensing is usually applied for security systems, including airport entry checkpoints, building entry, home-based security systems, and access to restricted areas [29]. Due to the nature of security systems, biometric footstep recognition systems encompass a small number of identified persons, even if the used datasets can be enormous. These datasets can also be constrained by the user not being allowed to change footwear [20, 30]. Moreover, few open-access footstep pressure databases are available in the literature to train and validate recognition systems. Recently, the database UNB StepUP-P150 has been made available, including over 200,000 footsteps from 150 individuals and footwear types (barefoot, standard shoes, and two personal shoes) [31].

Surveying works describing unwearable footstep pressure-sensing systems similar to our proposed biometric system, we can draw a timeline from a small set of works. Nakajima's pioneering work used the BIG-MAT platform (Niita Inc.), instrumented with resistive sensors, to obtain images of footprints from 10 individuals. Once pre-processed and normalized, these footprints served as input for a statistical recognition method that considered two parameters: the Euclidean distance between the centers of mass and the angle formed between the images of the right and left feet. The recognition rate was 85%. We can also cite the work in [32] that uses a commercial pressure-sensing mat to compute the directionally aligned and quantized center of pressure trajectory during the human walk. Person identification uses an HMM and an ANN trained with the Levenberg-Marquardt Learning Algorithm (LMA), achieving 64% accuracy. Despite this low performance, the work's main contribution is the use of dynamic images obtained during a walk. Sutaala and co-workers [17] developed a sensor material, an electromechanical film (EMFi), forming 64 strikes on the floor, producing footstep profiles for 10 persons. They apply five classifiers (Support vector machines - SVM, Multi-layer perceptron - MLP, Radial basis function - RBF, Learning vector quantization - LVQ, and k-Nearest neighbors - kNN) for person identification. SVM and MLP outperformed the other classification methods, reaching 95.33% and 93.67% recognition accuracy. The work in [24] implemented a fuzzy artificial immune system using a high-resolution (5 mm x 7 mm) pressure mat to acquire barefoot and shod data from 10 subjects. The normalized center of pressure (COP) and other features, such as foot width, length, toe-out angle, and total area, were extracted and used for identification. Their system

https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

reached an equal error rate of 3.8%. The work in [18] also extracted COP curves and key point features from a high-resolution pressure mat, forming a 3D COP trajectory over a footstep. It uses Fisher linear discriminant as the classifier and attains an average recognition rate of 92.3% for eleven persons. Connor in [20] used pressure image features and COP features of barefoot and shod data, applying principal component analysis (PCA) to select the most representative features and the nearest neighbor classifier for person identification. He attains a classification rate of 93.2% for the barefoot-based identification of 92 subjects and 90.5% for the identification of 13 persons with different shoes, up to a 100% classification rate if the person maintains the same shoe. An optical pedograph instrumented with four digital cameras (two for each foot) for recording barefoot images and a convolutional neural network trained with deep learning for footprint classification is presented in [33]. The acquisition equipment is mounted in a black acrylic box measuring 40 cm in height and 40 cm in width, forcing the user to stand on this box during enrollment. The color coding result image of foot pressure is used to identify the person, achieving a recognition rate of 92.69% for 13 users. In [34], twenty individuals' footsteps are captured with a grid of high-resolution instrumented floor tiles provided by Stepscan Technologies Inc., generating a database with 19,161 footsteps. The footsteps are obtained with shod (3 types of shoes), socks, and bare feet. Peak pressure images providing detailed shape and texture information from the contact between the sole and the ground are used to train a generative adversarial network-based model with a domain discriminator to improve identification. The model presents around 97% balanced accuracy (considering the shod and unshod feet images).

Our recent work [8] uses a sensing platform instrumented with fiber Bragg gratings (FBG) for capturing plantar pressure signals. The wavelength shifts experienced by FBGs are the features processed by an SVM to recognize the footstep signature of a single foot (left or right), achieving a recognition rate of 86% for twelve persons.

The present work uses the same FBG sensing platform in [8] to acquire plantar pressure signals for both feet. These signals are transformed into a pressure map image and used to train the two-stage neural network (CNN-FNN) model. The main goal is to achieve a user-friendly, reliable, and accurate system for person identification.

3 TWO-STAGE NEURAL NETWORK (CNN-FNN) MODEL FOR FOOTSTEP IDENTIFICATION

The design of our proposed CNN-FNN biometric system followed the steps:

- Data acquisition and data augmentation (data collection): The sensing pressure mat described in [8] is used for person data acquisition. Each user accomplishes only five enrollment sessions, generating 5 pairs of feet. We apply a data augmentation technique to artificially expand the signal database, aiming to increase the CNN model's generalization.

- Generation of templates database (data pre-processing): A signal processing procedure transforms the spectra of optical signals representing the person's footstep obtained with the sensing mat and those obtained by data augmentation into a pressure map corresponding to an image containing two footsteps. Each image is split into two small images corresponding to the left and right footsteps. The complete database contains four hundred pairs of footsteps for each user.

- Feature extraction and classification: Two CNNs (left and right footsteps) are independently trained for feature extraction and person classification. The left-foot CNN model can uniquely identify the user, and the right-foot CNN model can also identify the user. In this step, we have two independent biometric identification models.

- Fusion information and decision for person identification (data post-processing): A fully connected neural network for decision-making is formed by inputting the trained left-foot and right-foot CNN models. This matching network, trained using a stochastic gradient descent algorithm, authenticates the users by identifying their two-footstep images. Thus, applying a pruning technique [35], we reduce the complexity of the CNN-FNN final model and improve the accuracy of the proposed biometric system.

These steps, summarized in Figure 1, are detailed in the next sections.

3.1 Footstep signal acquisition and data augmentation

A successful real-time identification system requires an efficient biometric signal acquisition system, making it hygienic and user-friendly during the enrollment step. This work uses the sensing platform instrumented with 2 sets of FBGs encapsulated in silicone, developed in [8]. The platform comprises a rectangular acrylic base mold, with dimensions (28.0 x 58.0) cm and a height of 5.0 cm, filled with acetic-curing silicone (Selabond). The two sets of fiber segments, containing 7 FBGs each in series, are appropriately positioned on the silicone layer. Then, a second layer of silicone encapsulates the transducers, resulting in a 1 cm high sensor mat. The distribution of the 7 FBGs matches the foot regions corresponding to the forefoot, midfoot, and hindfoot for feet with lengths ranging from 23.0 to 27.0 cm, so that each region of the foot is sensed by at least two FBGs, one for the medial portion and another for the lateral portion. The forefoot region, sensed by 3 FBGs, also matches the metatarsal region. Thus, two sensing regions on the mat correspond to the right and the left foot positions as shown in Figure 2. More information about fiber Bragg gratings recording, platform manufacturing, and sensing characterization can be found in [8].

During enrollment, the user stands on the sensing platform in the areas close to the foot marks for 90 seconds. The enrollment repeats five times, does not require bare feet, and can be done with shoes (except for stilettos). Figure 3 shows the Bragg wavelength shifts ($\Delta\lambda$) for FBG1 to FBG7, corresponding to the left foot when a person steps on the sensor mat.

Learning and Nonlinear Models - Journal of the Brazilian Society on Artificial Intelligence (SBIA), Vol. 24, Iss. 1, pp. 60–73, 2026
https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

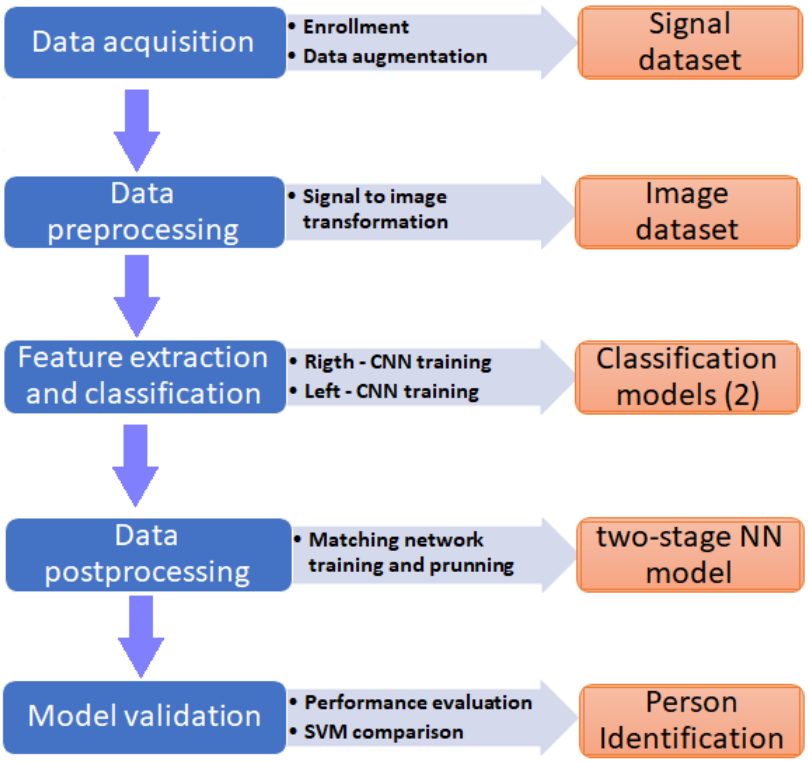


Figure 1: The flowchart of model development and data analysis

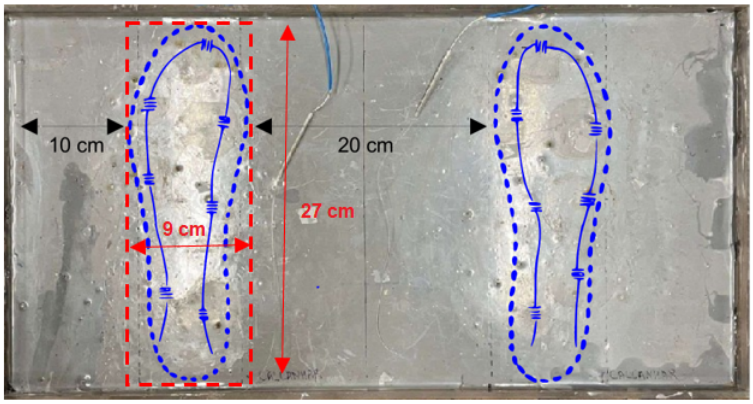


Figure 2: Sensing pressure mat. The dashed lines indicate the sensing regions, whereas the solid lines represent the optical fiber and the approximate location of each Bragg grating. (Adapted from [8])

For each FBG (FBG1 to FBG14), the acquisition algorithm collects Bragg wavelength data and subtracts the first 50 samples from the last 50, generating the measured data. The average of the measurement's absolute values is calculated. These means values correspond to the Bragg wavelength shift ($\Delta\lambda$) due to user plantar pressure and form a measurement vector as given in equation 1 for a person labeled as user $\#1$.

$$\begin{bmatrix}\Delta\lambda_{1,1} & \Delta\lambda_{1,2} & \ldots & \Delta\lambda_{1,13} & \Delta\lambda_{1,14} & label_1\end{bmatrix} \tag{1}$$

The mean and the standard deviation of the Bragg wavelength shifts resulting from the five repetitions are computed for each FBG, corresponding to the average behavior of the response of each sensor. This average behavior is used for data augmentation.

Training convolutional neural networks requires a large amount of data to learn the non-linear mapping between input and output data, thus enhancing their generalization capacity, diminishing overfitting, compensating for class imbalance, and improving classification accuracy. Data augmentation is a method for artificial data generation by creating variations of the original dataset [36]. For signals, data augmentation encompasses techniques like noise injection, slicing, jittering, and scaling. Image data augmentation involves transformations such as rotation, scaling, or translation [37]. In this work, additional users' footsteps are generated using a normal probability distribution characterized by the mean and standard deviation of the Bragg wavelength shifts computed from the original enrollments. The adoption of this procedure is based on the results of the experimental study

https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

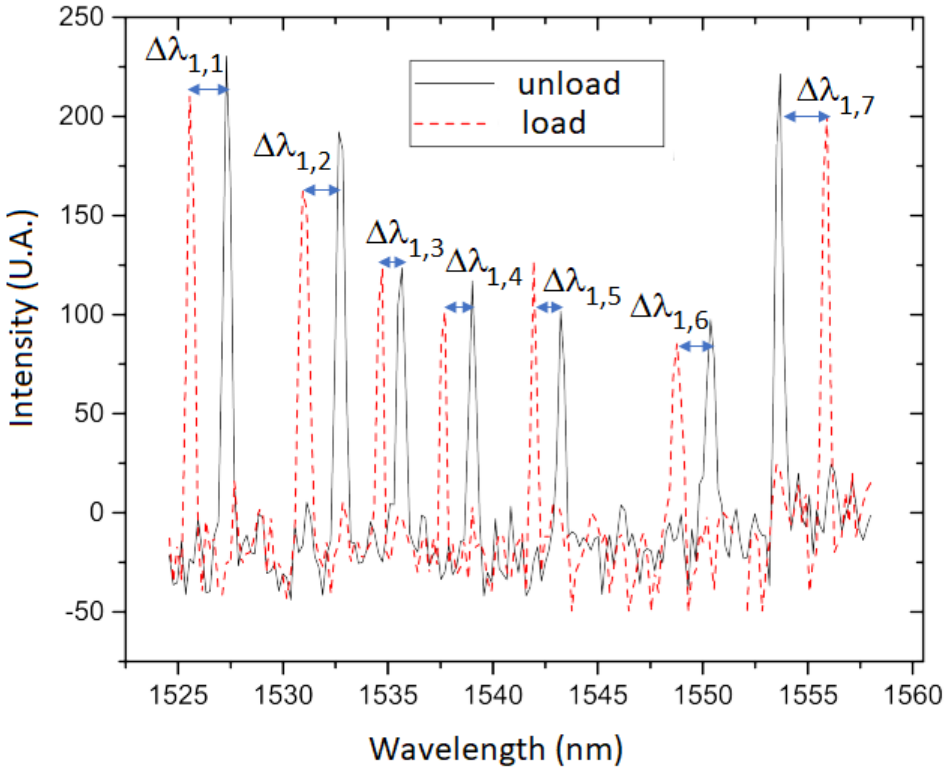


Figure 3: FGB 1 to FGB 7 (left foot) reflected spectra for the unloaded platform and when a person stepped on the platform.(Adapted from [8])

conducted to characterize the sensor mat - as described in [8] - which shows that the wavelength shift signals follow a Gaussian distribution for different feet with extra loads, positioned directly on the platform in the regions delimited by the FBGs. Thus, for each user, a data set containing 400 different Bragg wavelength shift vectors as given in equation 1 emulates different pairs of footprints of the same user.

3.2 Template database

To build the template database, we transform the wavelength shift vectors from equation 1 into a pressure map, which generates an RGB image. The plantar pressure map corresponds to a matrix in which each element represents 0.5 cm of the original dimensions of the sensing platform in Figure 2. The matrix with dimensions of 54 x 108 receives normalized values between 0 and 1, where 0 represents the minimum response, and 1 corresponds to the absolute maximum response observed in the wavelength shift in equation 1. These $\Delta\lambda$ values are converted to a color scale in which the higher the $\Delta\lambda$ value, the more intense the response in the chosen color scale. The FBG position in the matrix corresponds to 9 matrix elements, distributed in a 3 x 3 square shape, where the central component is the value of $\Delta\lambda$ in equation 1 associated with the FBG and the other elements are filled from this value decayed by 5.0% and 10.0%. According to Figure 2, the platform region containing each foot has 27.0 x 9.0 cm dimensions. Thus, the final pressure map has dimensions of 54 x 108, in which the 54 x 54 submatrix on the left is the image of the left foot and the 54 x 54 submatrix on the right is the image of the right foot. Figure 4 illustrates the procedure for obtaining the images. The generated RGB images have PNG extension, with dimensions (54 x 54 x 3) for a foot individually and (108 x 108 x 3) for the complete image of both feet.

Finally, the template database has eight hundred images for each class corresponding to a user. These images constitute two datasets containing 400 left and 400 right foot images, respectively.

3.3 Feature extraction and classification

We choose the CNN as our basic network structure because CNN are a specialized class of artificial neural networks which has been demonstrated to outperform other approaches in signal data transformed into images or spectra [38]. Structurally, CNN comprises convolutional layers, subsampling layers with non-linear activation functions, and fully connected layers. In general terms, the convolutional layer extracts feature maps from the input image, which serve as input for the subsampling layers (Pooling). In turn, the pooling layer's output serves as input for the fully connected layer, whose output is the classification result [39].

In this work, an adaptation of the architecture of LeNet-5, a CNN developed by Lecun and co-workers for numerical character

https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

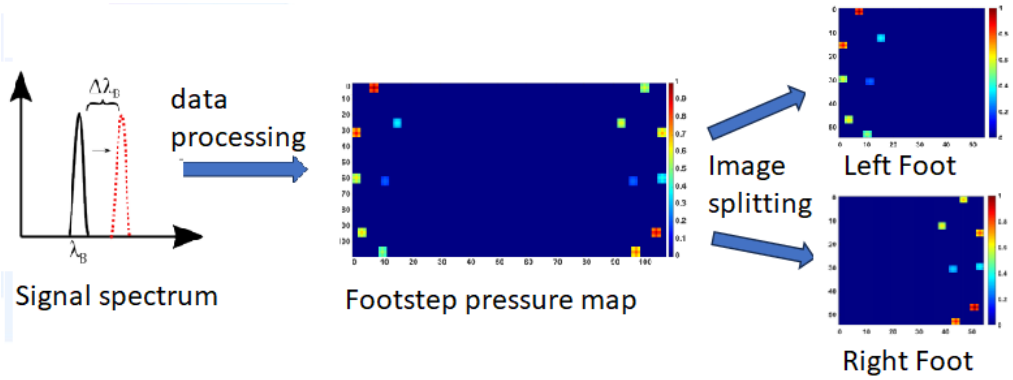


Figure 4: Signal to image transformation.

recognition [40], is used for footstep recognition. The LeNet-5 model can be easily adapted for pattern recognition and classification, from emotion recognition through facial expressions [41] until more recently architectural layouts [42]. The adaptation occurs by changing parameters such as filter dimensions or the fully connected layer's activation function to match the studied problem.

Thus, in this work, the images obtained from the wavelength shift transformation into a graphical representation serve as input for two CNNs inspired by LeNet-5. Each user's feet differ regarding the normalized pressure distribution in the predetermined sensing regions on the platform, as explained in section 3.1, and can be used as independent classifiers. In both CNN models, the structure with the convolutional and subsampling layers remains the same as the original LeNet-5, with only the dimensions of the input images and feature maps varying. Furthermore, the last fully connected layer has n neurons, where n is the number of registered users that should be recognized. The input left and right foot images have dimensions of 54 x 54 x 3. The zero padding and stride parameters are set to 0 and 1, respectively. Figure 5 shows the developed CNN architecture for a single foot image (left or right). The CNN architecture for the other foot image is the same.

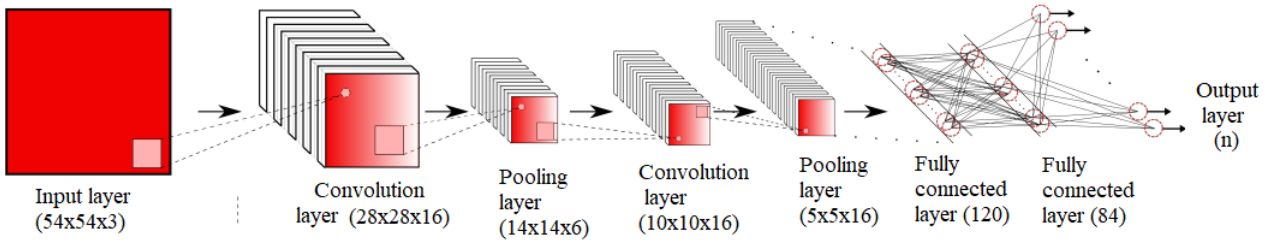


Figure 5: Convolutional Neural Network architecture for a single foot image.

We select twelve users ($n = 12$) from the template database to train both CNN models. Thus, the selected dataset has 4800 images for each foot; from these, 50% (2400) are used for training, 25% (1200) for test e the final 25% (1200) for validation. The trained model is evaluated through a k-fold cross-validation procedure with $k = 5$.

The Left-CNN and Right-CNN are trained with a minimum batch size of 16 for ten epochs. The confusion matrices are given in Figure 6 for the Left-CNN classifier and in Figure 7 for the Right-CNN classifier, which show the CNN performance by class. In a confusion matrix, the main diagonal indicates the data correctly classified in each class. The true classes correspond to the matrix rows, while the matrix columns represent the classes predicted by the model. Thus, the off-diagonal elements in a row correspond to the individuals from one class that are classified as belonging to another (false positive classification). The off-diagonal elements in the columns are the individuals from other classes that are predicted to the wrong class (false negative classification). The algorithm's overall hit rate (accuracy) corresponds to the rate of the sum of the diagonal values divided by the total amount of data. From Figure 6, the CNN trained for the left feet resulted in 79.83% accuracy, while for the right side, the value obtained was 77.58% according to Figure 7.

The confusion matrices indicate that both networks mix classes 10 and 11. In addition, some individuals from class 11 are identified as belonging to class 6, making user #11 the most difficult to recognize. This mistake occurs for both models. This error indicates that classes 10 and 11 are probably similar or that a problem occurred during data acquisition.

3.4 Fusion information and decision for person identification

As expected, neither of the two CNN models in section 3.3 provides 100% accuracy in user classification. Thus, we develop a decision model that integrates information at an abstraction level corresponding to user identification decisions, based on the

https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

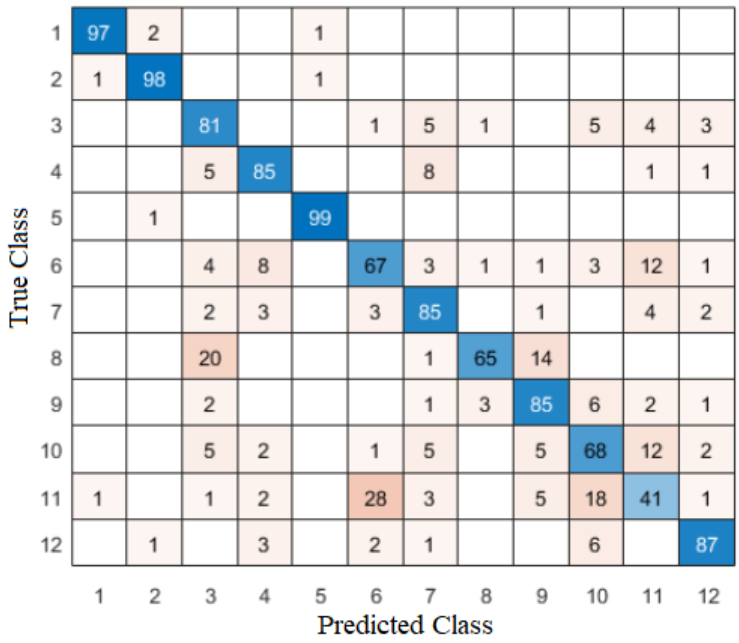


Figure 6: Confusion Matrix for Left-CNN Model

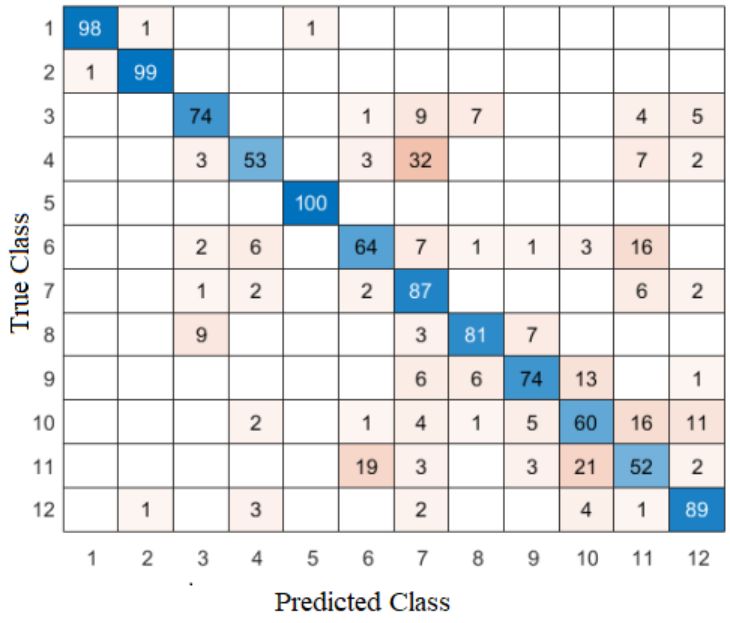


Figure 7: Confusion Matrix for Rigth-CNN Model

decisions of the two previous models. The main goal is to improve system accuracy and reliability. The use of two feet increases system reliability by the expectation of reducing false rejection and/or acceptance rates. For example, if the user steps outside one of the plate markers, the system can validate access using the part of the image corresponding to the second registered foot, thereby avoiding unnecessary read failures. Furthermore, the system's robustness increases, since correctly matching or duplicating the combination of two distinct images is more difficult than doing so for a single image. At the decision level, this can be achieved by combining multiple independently trained classifiers, such as Left-CNN and Right-CNN models [4].

This decision model is a fully connected neural network, matching network, inspired by the solution proposed by [43] to the digit recognition problem, which is solved with a multi-input CNN by dividing the images into upper and lower parts. Each part inputs an independent CNN, forming the base networks whose outputs feed new layers of the matching network. Each base CNN is separately trained and evaluated. The trained base CNNs constitute the first new layers of the matching network. The remaining matching network layers must be trained or fine-tuned. Thus, the matching network has the two CNN models in section 3.3 as base networks, corresponding to the left-side (left footstep) and right-side (right footstep) of the pressure map

https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

image. Four new layers complement the matching network:

- The **Merged layer** concatenates the data from the base CNNs in a parallel way.
- The **Fully connected layer** is the primary building block of the matching network that directly processes the input data from the base CNNs.
- The **Softmax layer** is the decision layer implementing the classification task since the softmax activation function outcomes are mutually exclusive.
- The **output layer** is responsible for the final classification

According to [44], a fully connected neural network estimates conditional posterior probability when the softmax activation is used in the output layer. The decision rule (output layer) chooses the class with the highest probability density, improving accuracy and adding reliability to the biometric system.

The same procedure and data for training the base CNN models in section 3.3 are applied to train and evaluate the matching network. The 4800 images containing a pair of footprints from the same users in section 3.3 are taken; from these, 50% (2400) is used for training, 25% (1200) for testing, and the final 25% (1200) for validation. The weights of the matching network's layers are trained using the stochastic gradient descent algorithm with a minimum batch size of 16 for ten epochs. The final trained model is evaluated through a k-fold cross-validation procedure with $k = 5$. The matching network reaches 96.42% accuracy with validation data.

Figure 8 shows the complete biometrics proposed system (CNN-FNN) operating in parallel mode with two units of the same biometric (left and right footstep image) in the first stage, and a matching network for decision taken in the second stage. Then, to optimize the CNN-FNN architecture and reduce computational costs, a pruning procedure is carried out, in which the network structure is modified by removing layers to observe the resulting network behavior [45]. The pruning procedure was applied only to the fully connected layers of the two CNN branches from the CNN-FNN model. All the convolution and pooling layers in Figure 8 were maintained. Two pruned models are generated. In the first one, the 84-neuron layer was removed, and the 120-neuron layer was connected directly to the 12-neuron output layer (see Figure 8). The second pruned model has the 120-neuron layer removed, and the output of the pooling layer (5x5x16) was connected to the fully connected layer (84). Both pruned models were retrained to compute their new weights, using the same procedures and training data as the previous training of CNN-FNN. The performance of the architecture in Figure 8 and the two pruned networks was compared on the validation set for class-specific and overall accuracy, as described in [45]. As a result, the retained architecture is similar to Figure 8 in which the 120-neuron fully connected layer is pruned from both base CNNs. The final CNN-FNN achieves 98.67% accuracy, as detailed in the next section.

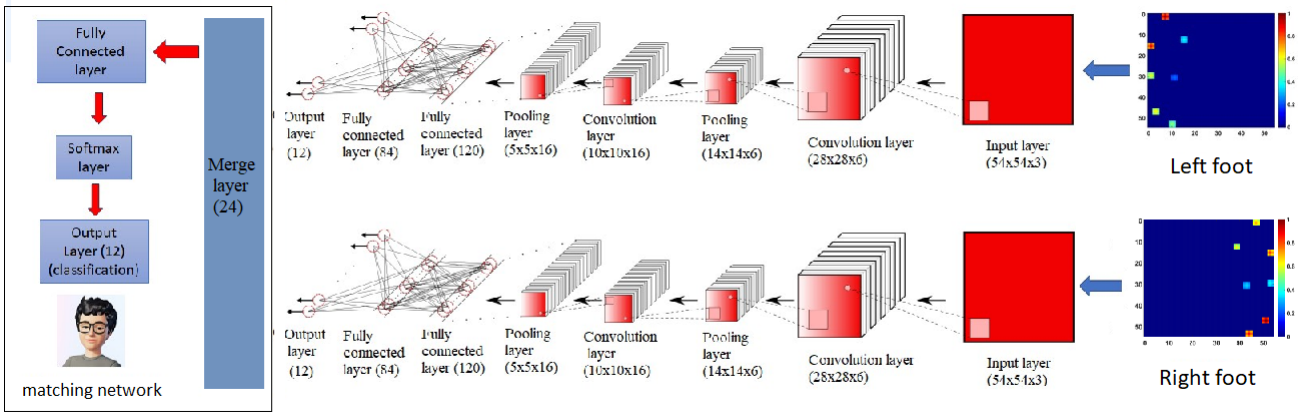


Figure 8: Proposed biometric system composed of a two-stage neural network (CNN-FNN) model.

4 RESULTS

All algorithms in this work are coded with MATLAB software, release R2023a, and run in a Dell notebook with 16.0 GB of RAM and an Intel(R) Core(TM) i7-7500U processor.

We compare the performance of the proposed CNN-FNN with an SVM classifier adapted to solve the multiclass classification problem of recognizing a user from the pressure plantar map images. The use of SVM as a baseline method is supported by the results presented in [8] and [17] and discussed in Section 2 above. These works report SVM achieves accuracy exceeding 80% in a pressure-sensor-based footprint identification systems.

Nowadays, multiclass SVM methods using kernel functions are a traditional method for statistical learning and classification [46]. This well-known machine learning method works by constructing hyperplanes to separate data into different classes. An

https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

SVM with an appropriate cost function can balance computational cost and classification efficiency for multiclass problems. Moreover, SVM models can achieve accuracy similar to deep learning models for the same problem [47], [12].

The non-linear polynomial kernel function in the equation 2 maps the RGB images in section 3.2 into a larger sampling space in which they are clustered, allowing user identification.

$$G(x_j, x_k) = (1 + x_j^T x_k)^3 \tag{2}$$

One of the drawbacks of using SVM is that the method does not compute probability estimates for classes. However, using a 5-fold cross-validation procedure, as done in this work, circumvents this drawback.

The same template database, with 12 users and 400 images per user, totaling 4800 images of footprint pairs, is used in this comparative study. The partition of the data into 50% for training, 25% for testing, and 25% for validation remains. That is, each user is trained with 200 pairs of footprints; 100 other pairs are used for testing, and the remaining 100 pairs are used for validation.

To evaluate and compare the performance of SVM and CNN-FNN identification models, we use the metrics *Precision*, *Recall*, and *F1-Score*, which are widely used in the literature, as recommended in [48]. Knowing the true positive rate (TPR), which computes the rate of correctly identified users, the false positive rate (FPR) corresponds to footprints that are indicated to belong to one user but actually belong to others, and the false negative rate (FNR) indicates the footprints of the same user identified as belonging to other users. These metrics are computed as follows.

$$Precision = \frac{TPR}{TPR + FPR} \tag{3}$$

$$Recall = \frac{TPR}{TPR + FNR} \tag{4}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

For data validation (100 selected and not used in training pairs of footprints for each user), the obtained confusion matrix for the SVM classifier is shown in Figure 9 and in Figure 10 for the CNN-FNN model.

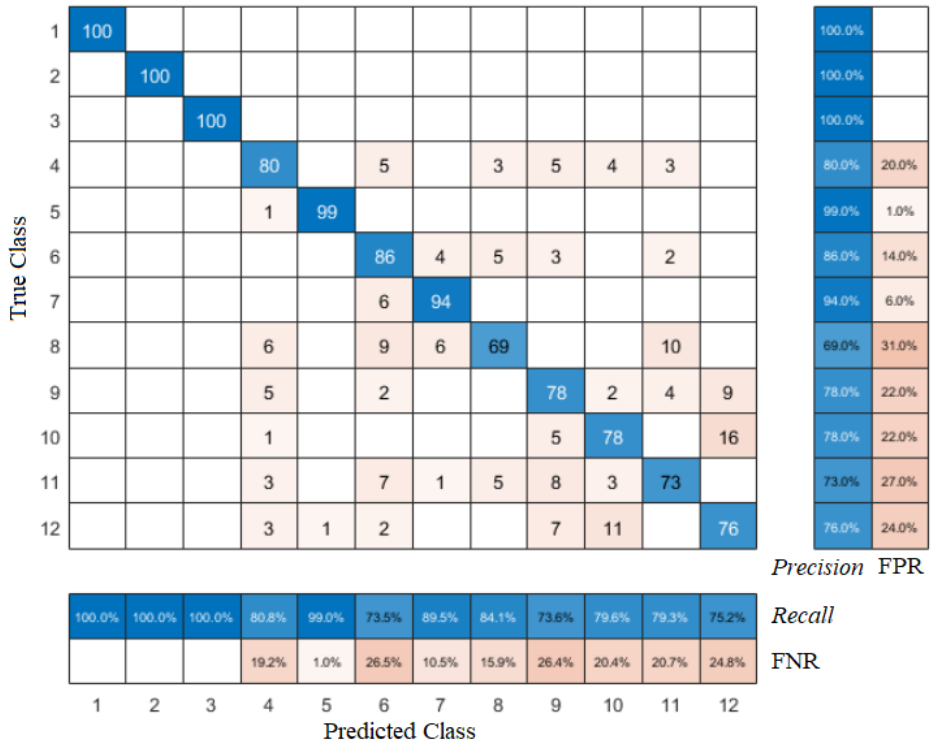


Figure 9: Confusion Matrix for SVM Classifier

The two columns on the right side of the confusion matrix in Figures 9 and 10 show the *Precision* metrics (blue column) and the FPR rate (white column). In turn, the two rows below the confusion matrix bring the *Recall* (blue row) and FNR rate (white row). Table 1 highlights these metric values.

https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

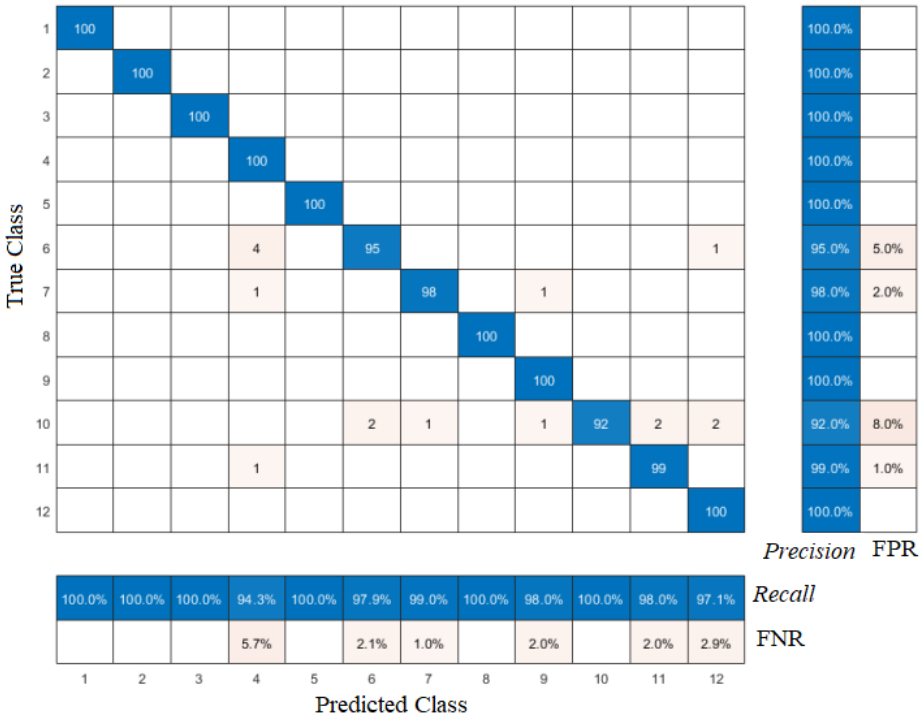


Figure 10: Confusion Matrix for CNN-FNN Model

Users	SVM		CNN-FNN	
(#)	*Precision*	*Recall*	*Precision*	*Recall*
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	0.800	0,808	1	0.943
5	0.990	0.990	1	1
6	0.860	0.735	0.950	0.979
7	0.940	0.895	0.980	0.990
8	0.690	0.841	1	1
9	0.780	0.736	1	0.980
10	0.780	0.796	0.920	1
11	0.730	0.9793	0.990	0.980
12	0.760	0.752	1	0.971

Table 1: *Precision* and *Recall* values for validation data

Users (#)	SVM	DCNN
1	1	1
2	1	1
3	1	1
4	0.8040	0.9707
5	0.9900	1
6	0.7926	0.9643
7	0.9169	0.9850
8	0.7581	1
9	0.7574	0.9899
10	0.7879	0.9583
11	0.7602	0.9850
12	0.7560	0.9853

Table 2: F1-Score Results

Using these values from Table 1, we calculate the *F1-Score* shown in Table 2, which demonstrates that the CNN-FNN model outperforms or equals the SVM model for all classes. In addition, the models' accuracy, that is, the ratio of the sum of the confusion matrix diagonal values to the total number of data points, is 86.08% for the SVM and 98.67% for the CNN-FNN.

Looking more closely at Figure 10 and Table 1, we see that the lowest *precision* value of the CNN-FNN is 0.92 for class 10, indicating that eight individual samples in this class were wrongly classified (FPR). Regarding *recall*, which represents the classifier's sensitivity, the lowest value was 0.943, indicating that six individuals were improperly classified as belonging to class 4 (FNR). There were no false-positive classifications for users 1, 2, 3, 4, 5, 8, 9, and 12, nor false-negatives for classes 1, 2, 3,

https://doi.org/10.21528/lnlm-vol24-no1-art5 © Brazilian Society on Artificial Intelligence

5, 8, and 10. On the contrary, for the SVM classifier Figure 9 and Table 1 show that except for classes 1 to 3, which had 100% hits, and class 5 with 99% hits, all other classes presented *precision* in the interval 0.69 to 0.86 and *recall* in the interval 0.73 to 0.89. For example, 31 individual samples from class 6 and 28 individual samples from class 9 were wrongly classified (FPR), and 31 other individuals were improperly classified as belonging to class 4 (FNR). The CNN-FNN's superior performance can be attributed to its ability to learn hierarchical features directly from the images. In contrast, the SVM relies on a priori defined features that may not capture the full complexity of the image patterns. Note that the results in this study were obtained using the same approach for SVM and CNN-FNN during feature extraction, which may affect SVM performance. Furthermore, the CNN's deeper architecture enables a more detailed representation of the image data, likely contributing to its superior performance.

It should also be noted that using an image with two feet increased the reliability of the results compared to using only one foot image, as discussed in section 3.4. Table 3, derived from the confusion matrices for the right and left feet, and figure 10 show the reduction in false-negative and false-positive rates in the system that uses the two-foot image. Furthermore, the system's accuracy improved from 79.83% (left foot, best-case scenario) to 98.67% (both feet).

Users	**Left-CNN**		**Rigth-CNN**		**CNN-FNN**	
(#)	*FPR*	*FPN*	*FPR*	*FPN*	*FPR*	*FPN*
1	0.03	0.020	0.02	0.010	0	0
2	0.02	0.039	0.01	0.020	0	0
3	0.19	0.325	10.26	0.169	0	0
4	0.15	0.175	0.47	0.197	0	0.057
5	0.01	0.020	0	0.01	0	0
6	0.33	0.34.3	0.36	0.289	0.05	0.021
7	0.15	0.241	0.13	0.431	0.02	0.01
8	0.35	0.071	0.19	0.156	0	0
9	0.15	0.234	0.26	0.178	0	0.02
10	0.32	0.358	0.40	0.406	0.08	0
11	0.59	0.461	0.48	0.490	0.01	0.02
12	0.13	0.112	0.11	0.205	0	0.029

Table 3: *False Positive Rate* and *False Negative Rate* values for validation data

Compared with other studies in the literature, especially those in [33], the proposed biometric system offers advantages, including a low-complexity, hygienic, and user-friendly enrollment device. The data augmentation procedure improves generalization capability and reduces the risk of overfitting during training, which is common in CNN-based models. An additional advantage is that with this procedure, the biometric system does not require excessive user enrollment sessions as in [34]. The CNN-FNN model achieves performance equal to or better than that of these studies and is therefore suitable for a restricted-access security system.

As with most neural network-based recognition pattern models, adding new classes to the developed biometric system presented in this work requires retraining the algorithms. However, access control systems are characterized by a small number of allowed users. Moreover, if necessary, other CNN-FNN models, trained for 12 to 20 users, can be combined through an ensemble learning technique, as proposed in [49], which can ensure the safe expansion of our biometric access system when including new users.

Regarding computational cost, the SVM classifier took 2 min to train, whereas the CNN-FNN model took approximately 60 min. Furthermore, the design of CNN-FNN is more complex, requiring training of the base networks, training of the matching network, and a pruning procedure with recovery training. In contrast, SVM is ready to use and more viable for constrained scenarios. However, it is worth noting that CNN-FNN training ran on a single-core machine without parallel or distributed computing. Following the recommendations of [50], distributed processing on GPUs and multicore machines can reduce the computational cost of training. In particular, using a GPU to train convolutional layers reduces training time by 40x [50]. Furthermore, the Stochastic Gradient Descent algorithm used to train the fully connected layers, especially the recovery training of the target network, can be distributed across multiple cores, thereby speeding up the training process.

5 CONCLUSIONS

This work presented a biometric system for person identification composed of a user-friendly platform instrumented with two sets of FBGs and a CNN-FNN network resulting from the fusion of two CNN models, one for each foot of a user, with a fully connected neural network. The platform is hygienic and user-friendly, allows enrollment with shoes, and requires only five acquisitions per user. The five measurements per user yield 400 distinct records through data augmentation. Data augmentation helped reduce the risk of overfitting and improved the classifier's ability to generalize to new examples. The template database is built by converting these recorded signals into images for classifier training. The system has two classifiers (Left-CNN and Right-CNN), one for each user's foot, which were independently trained. These networks serve as inputs to a matching network (CNN-FNN), the final pattern-recognition model. This matching network is trained on images containing both feet, resulting in a system that is more reliable and accurate than one composed solely of a foot CNN classifier. This is due to the reduction in false rejection and acceptance rates achieved using the combined image of two feet. The biometric CNN-FNN system was also compared with a standard SVM-based classifier. The CNN-FNN outperformed the SVM across the Precision, Recall, and F1-Score metrics. In particular, its accuracy reached 98.87%, whereas the SVM's accuracy was 86.08%. Overall, the CNN-FNN

demonstrated solid, reliable performance in identifying all 12 users. However, CNN-FNN required more computational resources and longer training times than SVM, which can limit its application when simplicity is prioritized and computational resources are constrained.

As future work, we are adapting the proposed model to handle non-square images, thereby reducing image size, model complexity, and training computational time. We are also considering using non-Gaussian distributions to build the augmented database and analyzing how different PDFs can affect system performance. Furthermore, we work on distributed computation by parallelizing the algorithms on GPUs and multicore machines, as discussed in Section 4. We are also investigating the performance of other neural network architectures, such as vision transformers (ViTs) and generative adversarial networks (GANs), compared with the present system, which is based on convolutional networks (CNNs). After this step, we plan to migrate the system to edge computing devices without Internet connectivity, reducing the risk of spoof attacks and increasing security.

REFERENCES

[1] N. S. Latman and E. Herb. "A field study of the accuracy and reliability of a biometric iris recognition system". *Science & Justice*, vol. 53, no. 2, pp. 98–102, 2013.

[2] V. Dhir, S. Amarpreet, K. Rakesh and S. Gurpreet. "Biometric recognition: a modern era for security". *International Journal of Engineering Science and Technology*, vol. 2, Aug. 2010.

[3] D.-J. Kim, K.-W. Chung and K.-S. Hong. "Person authentication using face, teeth and voice modalities for mobile device security". *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2678–2685, 2010.

[4] N. Bala, R. Gupta and A. Kumar. "Multimodal biometric system based on fusion techniques: a review". *Information Security Journal: A Global Perspective*, vol. 31, no. 3, pp. 289–337, 2022.

[5] A. Parashar, A. Parashar, M. Shabaz, D. Gupta, A. K. Sahu and M. A. Khan. "Advancements in artificial intelligence for biometrics: A deep dive into model-based gait recognition techniques". *Engineering Applications of Artificial Intelligence*, vol. 130, pp. 107712, 2024.

[6] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun and D. Zhang. "Biometrics recognition using deep learning: a survey". *Artificial Intelligence Review*, vol. 56, no. 8, pp. 8647 – 8695, 2023.

[7] L.-C. Zhao, T. Zhou, S.-D. Chang, H.-X. Zou, Q.-H. Gao, Z.-Y. Wu, G. Yan, K.-X. Wei, E. M. Yeatman, G. Meng and W.-M. Zhang. "A disposable cup inspired smart floor for trajectory recognition and human-interactive sensing". *Applied Energy*, vol. 357, pp. 122524, 2024.

[8] N. S. Girao, M. Muller and L. V. R. de Arruda. "A New Biometric Identification System Based on Plantar Pressure". *IEEE Sensors Journal*, vol. 23, no. 15, pp. 16900–16906, 2023.

[9] V. D. A. Kumar and M. Ramakrishnan. "Legacy of footprints - a review". *International Journal of Computer Applications*, vol. 35, no. 11, pp. 9–16, Dec. 2011.

[10] R. Kushwaha, N. Nain and S. K. Gupta. "Person Identification on the Basis of Footprint Geometry". In *12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 164–171, Naples, Italy, 2016. IEEE.

[11] S. Basheer, K. K. Nagwanshi, S. Bhatia, S. Dubey and G. R. Sinha. "FESD: An Approach for Biometric Human Footprint Matching Using Fuzzy Ensemble Learning". *IEEE Access*, vol. 9, pp. 26641–26663, 2021.

[12] R. Khokher and R. C. Singh. "Footprint Identification: Review of an Emerging Biometric Trait". *Macromolecular Symposia*, vol. 397, no. 1, pp. 2000246, 2021.

[13] R. V. Rodriguez, N. W. D. Evans and J. S. D. Mason. "Footstep Recognition". In *Encyclopedia of Biometrics*, edited by S. Z. Li and A. Jain, pp. 550–557. Springer US, Boston, MA, 2009.

[14] M. Addlesee, A. Jones, F. Livesey and F. Samaria. "The ORL active floor [sensor system]". *IEEE Personal Communications*, vol. 4, no. 5, pp. 35–41, 1997.

[15] M. Derlatka and M. Bogdan. "Ensemble kNN classifiers for human gait recognition based on ground reaction forces". In *2015 8th International Conference on Human System Interaction (HSI)*, pp. 88–93, 2015.

[16] Y.-J. Lee, Y.-S. Wu and P.-C. Lin. "Utilization of two types of feature datasets with image-based and time series deep learning models in recognizing walking status and revealing personal identification". *Advanced Engineering Informatics*, vol. 62, pp. 102729, 2024.

[17] J. Suutala and J. Roning. "Methods for person identification on a pressure-sensitive floor: Experiments with multiple classifiers and reject option". *Information Fusion*, vol. 9, no. 1, pp. 21–40, 2008. Special Issue on Applications of Ensemble Methods.

[18] G. Qian, J. Zhang and A. Kidane. "People Identification Using Floor Pressure Sensing and Analysis". *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1447–1460, 2010.

[19] S. Zheng, K. Huang and T. Tan. "Evaluation framework on translation-invariant representation for cumulative foot pressure image". In *2011 18th IEEE International Conference on Image Processing*, pp. 201–204, 2011.

[20] P. C. Connor. "Comparing and combining underfoot pressure features for shod and unshod gait biometrics". In *2015 IEEE International Symposium on Technologies for Homeland Security (HST)*, pp. 1–7, 2015.

[21] S. Drira, S. G. Pai, Y. Reuland, N. F. Olsen and I. F. Smith. "Using footstep-induced vibrations for occupant detection and recognition in buildings". *Advanced Engineering Informatics*, vol. 49, pp. 101289, 2021.

[22] S. Wu, S. Huang, Z. Liu, Q. Zhang and J. Liu. "AFPILD: Acoustic footstep dataset collected using one microphone array and LiDAR sensor for person identification and localization". *Information Fusion*, vol. 104, pp. 102181, 2024.

[23] Y. Zhou, J. K. Yeoh, Y. E. Li and W. Solihin. "Large scale indoor occupant tracking using distributed acoustic sensing and machine learning". *Building and Environment*, vol. 247, pp. 111005, 2024.

[24] T. Takeda, K. Taniguchi, K. Asari, K. Kuramoto, S. Kobashi and Y. Hata. "Biometric personal authentication by one step foot pressure distribution change by fuzzy artificial immune system". In *International Conference on Fuzzy Systems*, pp. 1–6, 2010.

[25] T. Yamakawa, K. Taniguchi, K. Asari, S. Kobashi and Y. Hata. "Biometric personal identification based on gait pattern using both feet pressure change". In *2010 World Automation Congress*, pp. 1–6, 2010.

[26] K.-H. Yeh, C. Su, W. Chiu and L. Zhou. "I Walk, Therefore I Am: Continuous User Authentication with Plantar Biometrics". *IEEE Communications Magazine*, vol. 56, no. 2, pp. 150–157, 2018.

[27] V. V. Tselishchev, A. V. Yupashevskiy, A. G. Samokhin, A. S. Kazmina, K. A. Metsler and A. V. Markov. "The Development of Pedograph and Methods for Diagnosing Gait Disorders". In *2022 IEEE 23rd International Conference of Young Professionals in Electron Devices and Materials (EDM)*, pp. 534–537, 2022.

[28] N. S. Girao, D. Lugarini, L. V. R. de Arruda, J. L. Fabris and M. Muller. "Multiplexed in-Series Macro-Bend Fiber Sensors for Personal Authentication Through Foot Recognition". *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, vol. 23, no. 02, pp. 1–12, 2024.

[29] O. Costilla-Reyes, R. Vera-Rodriguez, P. Scully and K. B. Ozanyan. "Analysis of Spatio-Temporal Representations for Robust Footstep Recognition with Deep Residual Neural Networks". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 285–296, 2019.

[30] E. Macdonald, R. Larracy, A. Phinyomark and E. Scheme. "Underfoot Pressure-Based Left and Right Foot Classification Algorithms: The Impact of Footwear". *IEEE Access*, vol. 11, pp. 137937–137947, 2023.

[31] R. Larracy, A. Phinyomark, A. Salehi, E. MacDonald, S. Kazemi, S. S. Bashar, A. Tabor and E. Scheme. "A dataset of high-resolution plantar pressures for gait analysis across varying footwear and walking speeds". *Scientific Data*, vol. 12, no. 1, 2025.

[32] J.-W. Jung, T. Sato and Z. Bien. "Dynamic footprint-based person recognition method using a hidden markov model and a neural network". *International Journal of Intelligent Systems*, vol. 19, no. 11, pp. 1127–1141, 2004.

[33] T. Keatsamarn and C. Pintavirooj. "Footprint Identification using Deep Learning". In *2018 11th Biomedical Engineering International Conference (BMEiCON)*, pp. 1–4, 2018.

[34] R. Larracy, S. S. Bashar, A. Phinyomark and E. Scheme. "StepGAN: Gait Feature Extraction Using Generative Adversarial Networks for Footstep Recognition". In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–6, 2024.

[35] R. Meyes, M. Lu, C. W. de Puiseau and T. Meisen. "Ablation Studies in Artificial Neural Networks". *arXiv*, 2019.

[36] A. Al-Kababji, F. Bensaali, S. P. Dakua and Y. Himeur. "Automated liver tissues delineation techniques: A systematic survey on machine learning current trends and future orientations". *Engineering Applications of Artificial Intelligence*, vol. 117, pp. 105532, 2023.

[37] Z. Gao, H. Liu and L. Li. "Data Augmentation for Time-Series Classification: An Extensive Empirical Study and Comprehensive Survey". *Journal of Artificial Intelligence Research*, vol. 83, pp. 5.1–5.46, 2025.

[38] H. Alaskar. "Convolutional neural network application in biomedical signals". *Journal of Computer Science and Information Technology*, vol. 6, no. 2, pp. 45–59, 2018.

[39] I. J. Goodfellow, Y. Bengio and A. Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. http://www.deeplearningbook.org.

[40] 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, 1998.

[41] C.-J. Lin, C.-H. Lin, S.-H. Wang and C.-H. Wu. "Multiple Convolutional Neural Networks Fusion Using Improved Fuzzy Integral for Facial Emotion Recognition". *Applied Sciences*, vol. 9, no. 13, 2019.

[42] C. Cai and B. Li. "Training deep convolution network with synthetic data for architectural morphological prototype classification". *Frontiers of Architectural Research*, vol. 10, no. 2, pp. 304–316, 2021.

[43] I. Kenta. "Image Classification using CNN with Multi-Input using MATLAB". Disponivel em: https://github.com/KentaItakura/Image-Classification-using-CNN-with-Multi-Input-using-MATLAB/releases/tag/2.0, 2024. Acessado em: 20 nov. 2022.

[44] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 11 1995.

[45] A. Hussain, S. U. Khan, N. Khan, M. Shabaz and S. W. Baik. "AI-driven behavior biometrics framework for robust human activity recognition in surveillance systems". *Engineering Applications of Artificial Intelligence*, vol. 127, pp. 107218, 2024.

[46] L. Sunitha and M. B. Raju. "Multi-class classification for large datasets with optimized SVM by non-linear kernel function". *Journal of Physics: Conference Series*, vol. 2089, no. 1, pp. 012015, nov 2021.

[47] P. Liu, K. R. Choo, L. Wang and F. Huang. "SVM or deep learning? A comparative study on remote sensing image classification". *Soft Computing*, vol. 21, no. 23, pp. 7053–7065, 2017.

[48] M. El-Abed, C. Charrier and C. Rosenberger. "Evaluation of Biometric Systems". In *New Trends and Developments in Biometrics*, edited by J. Yang and S. J. Xie, chapter 7, pp. 149–170. IntechOpen, Rijeka, 2012.

[49] A. Mohammed and R. Kora. "A comprehensive review on ensemble deep learning: Opportunities and challenges". *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 757–774, 2023.

[50] V. Hegde and S. Usmani. "Parallel and distributed deep learning". Technical report, Stanford University, 2016.