IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 8, 2020, accepted July 28, 2020, date of publication August 17, 2020, date of current version August 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3016970

Identity Recognition by Walking Outdoors Using Multimodal Sensor Insoles

KAMEN IVANOV[1,2], (Associate Member, IEEE), ZHANYONG MEI[3], MARTIN PENEV[4], LUDWIG LUBICH[4], OMISORE OLATUNJI MUMINI[5], (Member, IEEE), SAU NGUYEN VAN[2,6], (Student Member, IEEE) YAN YAN[1,2], (Member, IEEE), AND LEI WANG[1], (Member, IEEE)

[1]Shenzhen Key Laboratory for Low-cost Healthcare, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[2]Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China
[3]College of Cyber Security, Chengdu University of Technology, Chengdu 610059, China
[4]Faculty of Telecommunications, Technical University of Sofia, 1000 Sofia, Bulgaria
[5]CAS Key Laboratory for Health Informatics, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[6]Center for Biomedical Information Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

Corresponding author: Lei Wang (wang.lei@siat.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 71531004, Grant U1505251, Grant U19A2086, and Grant 61701049; in part by the Shenzhen Project for Overseas High-level Talents under Grant KQJSCX20170731163142116; in part by the Shenzhen Public Technology Service Platform Improvement Project of Biomedical Electronics; in part by the Key Project of the Science and Technology Department of Sichuan Province under Grant 2017GZ0304; and in part by the Young and Middle-aged Backbone Teachers' Program of Chengdu University of Technology.

ABSTRACT Recently, gait attracts attention as a practical biometric for devices that naturally possess walking pattern sensing. In the present study, we explored the feasibility of using a multimodal smart insole for identity recognition. We used sensor insoles designed and implemented by us to collect kinetic and kinematic data from 59 participants that walked outdoors. Then, we evaluated the performance of four neural network architectures, which are a baseline convolutional neural network (CNN), a CNN with a multi-stage feature extractor, a CNN with an extreme learning machine classifier using sensor-level fusion and CNN with extreme learning machine classifier using feature-level fusion. The networks were trained with segmented insole data using 0%, 50%, and 70% segmentation overlap, respectively. For 70% segmentation overlap and both-side data, we obtained mean accuracies of 72.8% ±0.038, 80.9% ±0.036, 80.1% ±0.021 and 93.3% ±0.009, for the four networks, respectively. The results suggest that multimodal sensor-enabled footwear could serve biometric purposes in the next generation of body sensor networks.

INDEX TERMS Gait recognition, smart insole, plantar pressure, wearable sensors, sensor fusion.

I. INTRODUCTION

Personal wearable devices take diverse roles in daily life, allowing for communication, entertainment, sports activity tracking, and vital signs monitoring. As they handle personal data, the aspects of security are a primary concern. Wearable nodes interconnect with one another dynamically within Body Sensor Networks (BSN), through diverse signal transmission mediums (Figure 1) and could thus become an easy target to imposter attacks. Furthermore, tiny sensor nodes do not possess enough computational power to process in real time the complex security tokens that traditionally ensure adequate levels of protection. For user access, wearable devices often do not possess traditional user interfaces to allow entering a password, nor can they use established biometrics such as face image or fingerprint. Thus, new methods that have the potential to meet the specific demands for security in BSNs become necessary. As human-body generated signals are unique to different individuals and available for wearable collection, they are extensively studied as candidates for biometric traits in wearable devices.

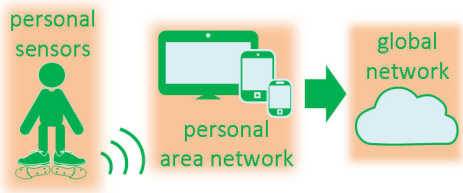

FIGURE 1. Transmission pathway from sensors to cloud.

The associate editor coordinating the review of this manuscript and approving it for publication was Fuhui Zhou.

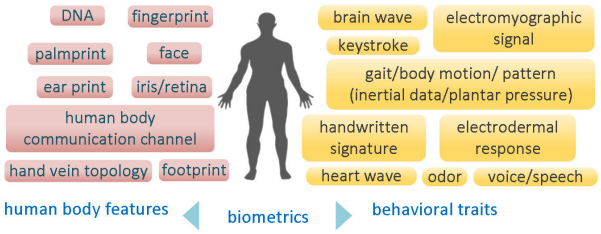

FIGURE 2. Human body characteristics serving as biometrics.

When evaluating a BSN biometric, a primary concern is the level of security that it allows. Some human body characteristics are considered more secure than others (Figure 2). For instance, time-evolving traits, such as heart, brain, and gait signals, when obtained at a single time instant, do not possess enough statistics for identity recognition. Thus, they are considered challenging to mimic. Another vital issue is whether the biometric requires explicit user input. Not only may traditional user interfaces be absent, but schemes, where the cooperation of the user is required may be perceived as obtrusive; thus, it may not be acceptable for the user to be disturbed. The diverse aspects of wearable-device biometric recognition were discussed in surveys [3]–[5].

Among the human body traits, gait meets the requirements for a highly secure biometric that does not require explicit user cooperation [4]. It is attractive for application in wearable systems capable of sensing walking characteristics, such as mobile phones, and fitness trackers. However, mobile applications involving user identification based on gait are practically absent since the gait as a biometric is still challenging, and the feasibility of reliable gait-based person identification based on data collected from wearable devices is yet to be proved.

Motivated by the fact that sensor footwear is expected to become prevalent in the foreseeable future, we focus on exploring the feasibility of person recognition based on data acquired from a multimodal sensor insole developed by us and intelligent processing using a 1D convolutional neural network (CNN). Our contributions lie in the following aspects:

- We collected a sensor insole dataset captured from 59 subjects during outdoors walking that includes records with duration of at least ten minutes of both kinetic and kinematic information. To our knowledge, so far, no other studies involved multimodal sensor insole data with comparable parameters.
- We explored the feasibility of person identification using the collected multimodal sensor insole data; for that, we developed a set of CNN models using various segmentation window overlaps. Especially, we tested an extreme learning machine classifier, a technique with potential performance benefits in implementations requiring model training in real time.

With the suggested extensive battery of experiments over large multimodal insole dataset, we extend the results of existing studies on person recognition using smart footwear.

II. BACKGROUND

The first practical, objective, and technical-tool-aided gait recognition methods were based on video tracking through cameras and restricted to laboratories [5]–[8]. These confirmed the technical feasibility of gait recognition but were not suited to application in natural settings. The advancement in sensor technologies made it possible to capture data reflecting gait through miniature inertial and force sensors paving the way to using gait recognition in BSN-applications. An extensive survey [4] covers in detail the specifics of gait analysis through inertial sensors. In wearable biometric applications, the user and the owner of the wearable device are typically the same person, and the trait of interest, such as gait is viewed in the context of continuous authentication, where information of the user is continuously collected to re-confirm the identity [2], [10]. So far, most successful studies that allowed capturing continuous gait information in natural settings relied on using inertial sensors integrated into mobile phones. That stems from the fact that mobile phones provide an easy, unobtrusive, continuous signal collection [10]–[18]. However, some limitations in the performance of mobile phone-based gait recognition arise from the limited number and types of available sensors and the lack of fixed location and alignment of sensors towards the human body and joint axes [17], [19], [20]. Recently, sensor-enabled footwear becomes sophisticated and practical enough for daily use. In our previous work [9], we demonstrated a multimodal sensor insole that allows capturing kinetic and kinematic information reflecting the foot dynamic characteristics. Figure 3 illustrates several domains where smart footwear enables mobile and pervasive applications involving health monitoring, remote medical diagnostics, and sports enhancement. Unlike mobile phones, smart footwear allows for fixed sensor locations and alignment, and a higher number, and different types of sensors. A combination of inertial and force sensors forms a multimodal system that reflects both kinematic and kinetic characteristics associated with the gait and functioning of lower limbs. Multiple modalities complement each other and provide richer information about gait patterns, determining better overall recognition performance than a single modality, as elaborated in survey [3].

Numerous studies on wearable-device based gait recognition utilized only kinematic variables obtained by inertial sensors [4]. Those that explored the combination of kinetic and kinematic data collected through a wearable device are fewer [1], [2]. Yeh *et al.* [2] demonstrated a method of continuous authentication based on plantar data from a smart insole. They used 14 samples and applied naïve Bayes and support vector machines with Gaussian radial basis function, achieving accuracies as high as 96.6 %. However, they did not use multimodal data. Choi *et al.* [1] suggested a method

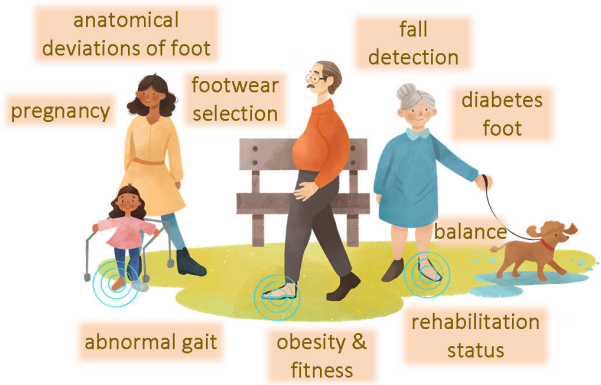

FIGURE 3. Various uses of smart footwear with potential implications of gait recognition.

for user identification utilizing both force and accelerometric data. They collected 3-minute recordings of walking from 14 subjects and used a non-linear discriminant analysis technique to map the input vector space to a lower-dimensionality one, thus avoiding the problems arising from the small sample size. Then they supplied the reduced-dimensionality vectors to one nearest neighbor classifier and achieved recognition accuracy of 95%. A common issue in these works is the small sample size. However, state-of-the-art gait recognition algorithms rely on deep learning methods that can offer superior performance but require large datasets from many users for model training. Also, as gait is a dynamic trait [4], and a weak biometric [21], obtaining personal gait features requires long-duration and long-term recordings, in contrast with traits such as fingerprint that may only need a single scan. In terms of inertial sensing, there were some successful attempts to address this problem. Neverova *et al.* [18] collected daily gait data from 1500 users over several months by inertial sensors incorporated into mobile phones. They then applied a Long Short-Term Memory network for identity recognition. Zou *et al.* [22] used data from 118 subjects collected through mobile phones to supply a deep learning method and achieved an accuracy of 93.5% for person identification and 93.7% for person authentication, respectively. Gadaleta and Rossi [23] collected gait data by mobile phones from 50 users for six months and implemented an orientation-invariant algorithm IDNet involving convolution neural networks. They reported a misclassification rate of less than 0.15% when supplying less than five walking cycles. Ngo *et al.* [24] provided the first large publicly available inertial sensor dataset collected from 744 subjects, intended to supply person recognition studies and balanced by gender and age. However, large datasets containing plantar pressure data from sensor insoles, or multimodal ones are yet to be constructed and made available to the community.

To make positioning of this work in gait recognition studies clear, we recall some basic definitions. Gait recognition could refer to distinguishing between normal and pathological gait [25], [26], pathological gaits, evaluation of the gait efficiency, or identifying an individual by gait [4], [5]. The latter subdivides into identification, verification, and authentication [4], [27]. When the recognition is based on machine learning, a recognition model is trained [28]. In the verification task, the user claims identity to the recognition system explicitly, and it is a one-class classification problem. Authentication is to continuously verify whether the claimed user remains the same during the access session. In the identification task, the system identifies the user without user's identity being claimed. As a multi-class classification problem, with no hint of identity provided, identification is most challenging [3]. When all identities in the test dataset were used in training, they are familiar to the system, and the recognition is of closed-set type; otherwise, it is an open-set recognition [29], [30]. We focus on person identification by gait of closed-set type.

III. SYSTEM DESIGN AND METHODOLOGY

A. HARDWARE IMPLEMENTATION

1) SENSOR INSOLE PROTOTYPE

The sensor insole used in the present study is designed, implemented, and tested by us. It inherits our previous design described in [9]. The components of the system are shown in Fig. 4. It involves a control module affixed on the frontal part of the shoe with an incorporated inertial sensor of type BMI160 (Bosch Sensortec, Germany). By a tiny cable, the module connects a sole insert with nine force sensors attached to a thin, flexible printed circuit board and allocated under the main weight-bearing areas of the foot. These are the big toe (T1), the five metatarsal heads (M1-M5), the midfoot (MF1), and the heel (LH1, MH1). To read force sensor signals, we applied channel multiplexing. It allowed for reliable operation, small control module, and low power consumption of less than 3 mA in an active mode of operation. In contrast with our previous design, we used a force sensor of type RP-C-10 (FilmSensor, China) with a lower range of 100 N and higher sensitivity, respectively. Besides, for convenient and reliable data collection outdoors, we used a new BSN data logger described below.

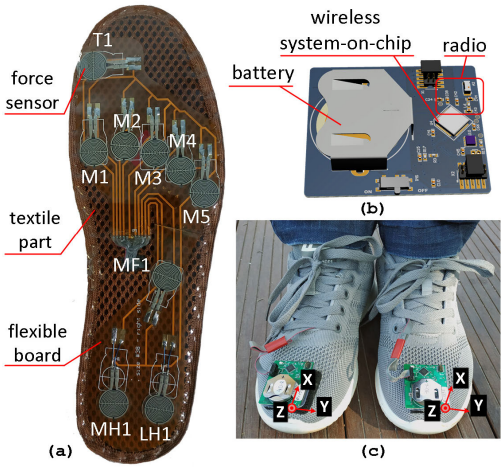

FIGURE 4. Components of the system (a) sensor insole prototype (b) control module (c) experimental setting.

The sampling rates for all sensors were 100 Hz. The sampling interval accuracy of force sensors was ensured by a crystal oscillator incorporated in the control board, whereas the sampling of the inertial sensor was controlled by its built-in RC oscillator, and its output rate accuracy was in the range of approximately ±1.5%. The range of accelerometer was set to 16 G, and the range of gyroscope was set to 2000 DPS.

2) BODY SENSOR NETWORK DATA LOGGER

For the present study, we developed a standalone, bare-metal data logger, illustrated in Fig. 5. It is provided with Bluetooth Low Energy connectivity and an SD card to store data. Compared to smartphones, it is not limited by operating system or resource sharing with other applications, and allows full real-time control of the data reception. Thus, it was easy to automate and simplify the experimental procedure. To ensure data integrity, a series of checks were performed continuously (Fig. 5b). Also, the external antenna enabled stable signal reception. For preventing massive data loss in case of system failure, data were stored on the SD card in sequentially - numbered files, each containing information of a 5-minute fraction of the complete recording. During the offline processing, these files were concatenated. Data packets were recorded entirely, including the overhead containing start and end of packets, sample counter, and a checksum. This kind of self-documentary recording allows identifying problems with the transportation of the packets during offline processing.

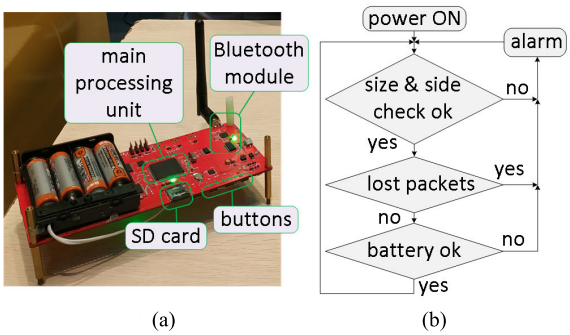

FIGURE 5. (a) custom datalogger (b) continuous checks.

B. EXPERIMENTAL PROCEDURE AND DATA ACQUISITION

We recruited 59 volunteers in total, from the Chengdu University of Technology and Shenzhen Institute of Advanced Technologies. The experimental procedure was approved by the ethics committees of both institutions and conformed to the Declaration of Helsinki. All volunteers were introduced to the experimental procedure and safety precautions, and consent for participation was obtained from each of them. Upon the time of the experiment, volunteers were in normal general health with no known foot anatomical or functional deficits, aged between 20 and 40, with foot sizes of 39, 40 or 44, body weight between 50 and 76 kg, and body height between 165 and 188 cm. Each participant put à pair of

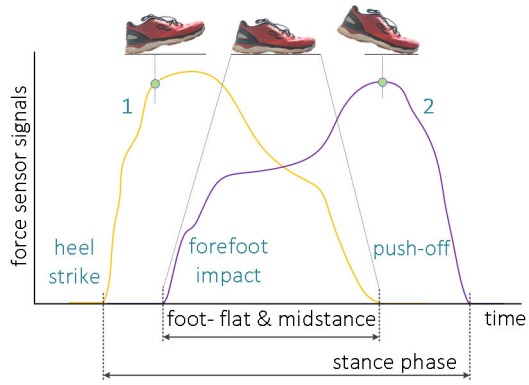

FIGURE 6. Foot contact phases reflected by insole signals; 1 – heel sensor signal; 2 – forefoot sensor signal.

custom instrumented shoes of appropriate size and was asked to walk outdoors freely at a self-selected speed for at least ten minutes. The outdoor environment of data collection were parks with a horizontal walking surface, allowing for mainly straight-line walking. During the experiment, each participant was accompanied by an operator who wore the data logger, instructed the participant about the route of walking, and ensured proper data collection. The duration of each session was indicated by the data logger. The recorded multimodal signals reflect spatial and temporal features of individual gait. Figure 6 shows how insole force sensor signals reflect foot contact phases. Upon heel contact, the heel force sensors are almost simultaneously activated. The progression of the foot contact ends with the toe-off when sensors under the metatarsal heads become active. Thus, force sensors reflect relatively accurately the temporal parameters, involving the stance and swing phase, and cadence. As the number of force sensors is low, the spatial resolution is also low. Thus, plantar pressure distribution and its derivative parameters, such as the path of the center of pressure are reflected only to a certain extent. Such a partial picture of the plantar pressure pattern raises the question of whether wearable force sensing could provide enough discriminative features for person identification. As to kinematic information, it is reflected by the inertial sensor. In that, characteristic peaks in the signals reflect each gait cycle; accelerometer signals also reflect the orientation towards the Earth, and gyroscope signals reflect the main axis of the rotational motion of the foot. Fig. 7 shows a representative set of signals from all modalities.

C. DATASET PREPARATION

1) INITIAL DATA PREPARATION

The processing pipeline accepted in this study is illustrated in Fig. 8. We aimed at collecting recordings with a duration of at least ten minutes to be used for training and validation of neural network models. An initial check was performed to ensure the validity of recorded signals, which included visual observation and file testing. During this process, we disregarded all files that contained incompletely received

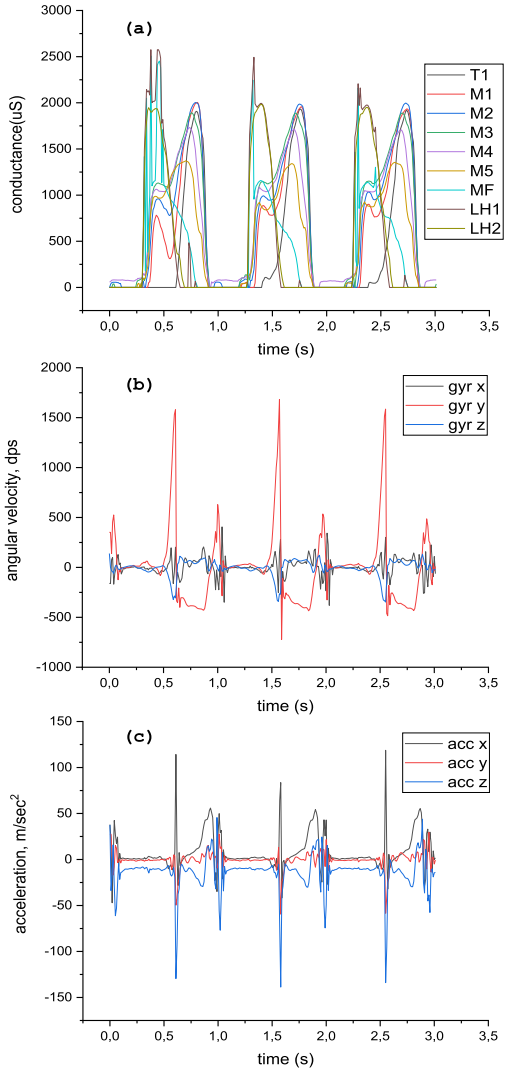

FIGURE 7. Representative signals of the different sensing modalities of the insole. (a) Force sensor signals (b) gyroscopic signals (c) accelerometric signals.

packets and ensured that each subject had at least one unimpaired 10-minute recording. For thirty-eight of the subjects, the recording sessions exceeded the ten-minute duration significantly. For each of these subjects, the additional data reflecting a regular walking pattern were used to form a test set.

Extracted signal time series for each participant were stored into a comma-separated file. For each participant, we obtained the raw sensor signal from thirty sensors (i.e., channels), sampled at 100 Hz. The sensors were a 3-axis accelerometer, 3-axis gyroscope, and nine force sensors, for the left and the right insole, respectively.

2) SENSOR SIGNAL SEGMENTATION

Neural networks allow raw signals to be fed to them directly, without explicitly defining features. For that, the input series can be segmented by individual gait cycles, or into frames with a fixed length [4], [22], [23]. Gait cycle detection is sensitive to failures and irregularities in the walking pattern and suffers from inter-cycle phase misalignment [31]. Hence, it is desired to devise algorithms that are independent of gait cycle detection and related features. In this work, we aimed to use convolutional neural networks, and as these can extract discriminative features from frames, we considered using frame segmentation. For the frame length, we have chosen 500 samples (5 seconds), thus ensuring that most frames will contain at least one gait cycle. Of all 59 recruited subjects, 21 subjects completed one good recording (i.e., no lost packets during data transmission and mostly straight-line walking) with a duration of exactly ten minutes. The other 38 subjects executed much longer recording sessions. As shown in Fig. 8, we used a ten-minute recording from each of the 59 subjects to create a large dataset for training and validation. As to testing, it is acceptable to be performed with just a part of the subjects. Thus, we used the additional good recordings beyond the 10-minute data of 38 subjects to create a testing set, forming a ratio of 67:33 between the "training+validation" set and the testing set. For applications utilizing real-time training, it is essential to set a proper balance between accuracy and training resources. It can be achieved by adjusting the segment overlap. Also, the information from the two insoles in a pair could be redundant, and using a single-insole data might be enough for satisfactory performance. To obtain insights about optimal choices of overlap and single/double side data, we explored the performance for overlaps of 0%, 50%, and 70%, respectively, for one and two insoles. Table 1 shows the number of segments for different overlaps.

TABLE 1. Number of segments for different overlaps.

overlap, %	0%	50%	70%
training and validation	7021	14042	23364
testing	3626	7252	12098

IV. CLASSIFICATION OF MULTIMODAL INSOLE DATA FOR GAIT RECOGNITION

With a new dataset, the first task is to identify whether data contain enough discriminative information. Convolutional neural networks and long-short term memory networks are most appropriate to learn features from raw sensor data. Hence, for initial evaluation, we adopted a CNN. For the construction of the CNN, there are no definite rules, and the optimal structure was determined empirically. The architecture was chosen as a trade-off between simplicity and discriminative power. The structure of the proposed neural network is given in Fig. 9. Feature vectors were obtained through a 1D convolutional neural network. Each feature vector (signature) is a numeric representation of the individual gait patterns. We chose a baseline CNN architecture consisting of three convolution layers, two max-pooling layers, and one

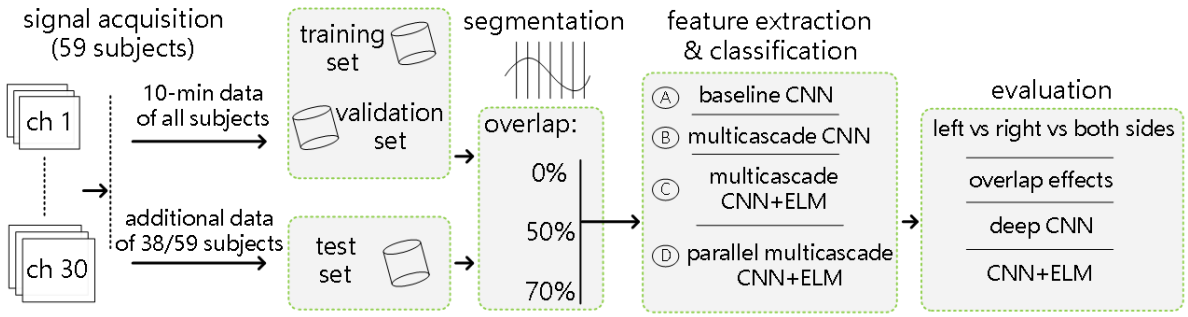

FIGURE 8. Data processing flow accepted in this study: signal acquisition, validation and test sets formation, segmentation, and classification.

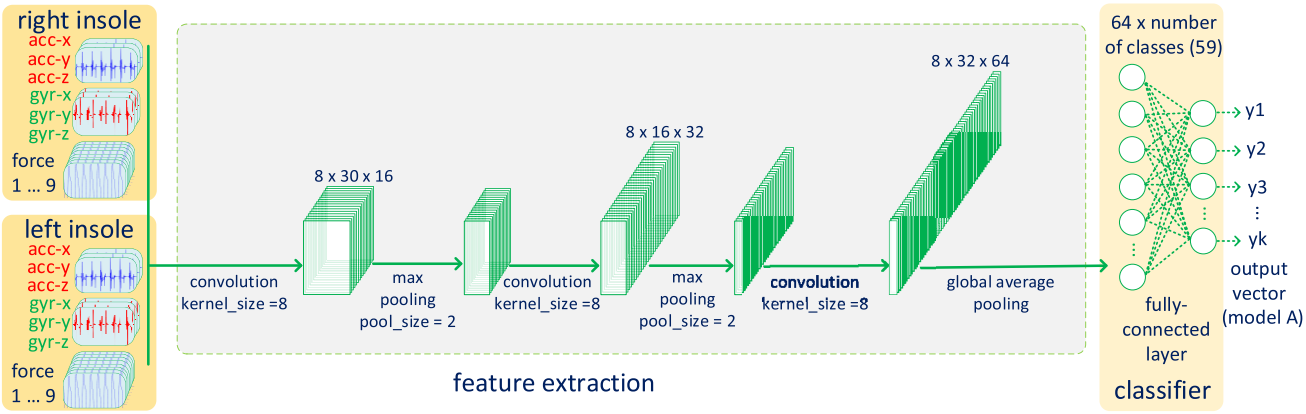

FIGURE 9. Baseline CNN (model A).

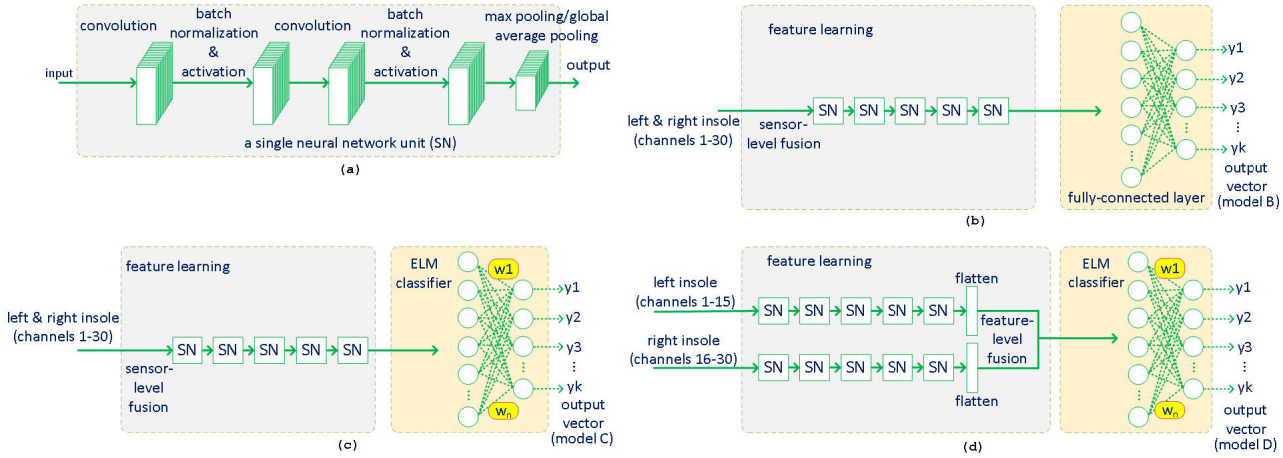

FIGURE 10. Neural network structures – the single unit (a) and models B, C, and D.

global average pooling layer. The input vector had a length of 500 x 30 channels. Before supplying features to the neural network, they were scaled in the range of $-1.0 \ldots +1.0$ to avoid an explosion of gradients. The output of the feature extraction block is a feature vector of size 64. The feature extraction block is followed by a fully-connected layer with a softmax operation. To reduce the overfitting, a dropout layer was added before the final dense layer as a regularization technique. The neural network is trained for a classification task. Each class represented a unique identity from the training dataset. We assigned a class label between 0-58 to each subject. For the baseline model A, we executed 100 train

TABLE 2. Parameters of the 5-stage cascade for models B-D.

layer	stage 1	stage 2	stage 3	stage 4	stage 5
1.Conv1D	filters = 32 kernel size = 8	filters = 64 kernel size = 3	filters = 128 kernel size = 3	filters = 256 kernel size = 3	filters = 256 kernel size = 3
2. BatchNormalization					
3. ReLU					
4. Conv1D	filters = 32 kernel size = 8	filters = 64 kernel size = 3	filters = 128 kernel size = 3	filters = 256 kernel size = 3	filters = 256 kernel size = 3
5. BatchNormalization					
6. ReLU					
7. MaxPooling1D/ Global Average Pooling 1D	pool size = 5	pool size = 3	pool size = 2	pool size = 2	Global Average Pooling 1D

epochs, using a batch size of 128, and categorical cross-entropy for the loss function with Adam optimizer.

As a subsequent step, we aimed at exploring a different structure of the feature extractor and the classifier. The main questions were whether a more complex feature extractor would improve the accuracy and how the segmentation overlap affects the performance for each architecture. For the feature extractor, we considered a cascade of five units. Each unit consisted of two repeating stages of convolution, followed by batch normalization and activation. The last layer of the first four units was max pooling, whereas the last layer of the fifth unit was global average pooling. The structure of the single unit is shown in Fig. 10a. The parameters of the single 5-stage cascade are given in Table 2. Then, as shown in Fig. 10b, we used the feature extractor to process all 30 channels, relying on sensor-level fusion [3], [4], [32].

Person recognition technologies are mostly necessary for real-time applications that need fast model training. One possible solution is adopting the Extreme learning machine (ELM) proposed by Huang *et al.* The foundations of this technique were explained in [33]. ELMs do not need tuning of the hidden node parameters. The latter are randomly initialized without subsequent updates. The hidden node output weights are learned in a single step. Thanks to these specifics, ELMs learn much faster compared to networks using backpropagation; at the same time, they preserve a comparable generalization performance. As ELM can be very suitable for real-time training in security applications, it motivated us to adopt it as one of the methods explored in this work. We replaced the last layer of the CNN with an extreme learning machine classifier, as shown in Fig. 10c. Finally, having the given ELM classifier, we implemented a more complex feature extractor that involved two parallel cascades, each accepting data from one insole. The outputs of the two cascades were concatenated to implement feature level fusion. The proposed structure is given in Fig. 10d as model D. For models B-D, the loss function was also categorical cross-entropy.

V. RESULTS

All reported accuracies are based on training models with data of all 59 samples and the given number of segments determined by the selected overlap. For each model, the number of correctly classified samples was divided by the total number of tested samples (i.e., 38). Each model training was executed five times, and the average accuracy was calculated. In most cases, the accuracy and validation losses stabilized before reaching the 20th epoch. A representative case is shown in Fig. 11. We initially performed a one-sample Kolmogorov-Smirnov test on accuracies data, and it showed that the requirement for normal distribution was not satisfied ($p<0.001$). Therefore, we performed statistical testing using a Kruskal-Wallis test. There was no significant difference among accuracies for different segmentation overlaps ($p=0.930$). However, statistically significant differences were exhibited in testing by model ($p<0.001$) and foot combination ($p<0.001$). Fig. 12 shows the results of statistical testing, giving insights on how accuracy depends on network architecture, segmentation overlap, and insole combination. Table 3 shows results for single-side and both-side data, for three segmentation overlap settings, for models A and B, respectively. Table 4 shows the performance when using an ELM classification with the single-cascade feature extractor; the results for the two-cascade feature extractor are given in Table 5.

VI. DISCUSSION

A. ACCURACY EVALUATION ON GENERAL TERMS

Table 6 shows a brief comparison of our results with others. The recognition performance is likely to drop with a significant increase in the number of classes. In [18], Google Abacus Dataset was used that contained multimodal data of 1500 mobile phone users collected for several months. By applying a standard convolutional neural network, they obtained an accuracy of 37% for the case when the claimed identity was in the top 5% of the classes. As they used

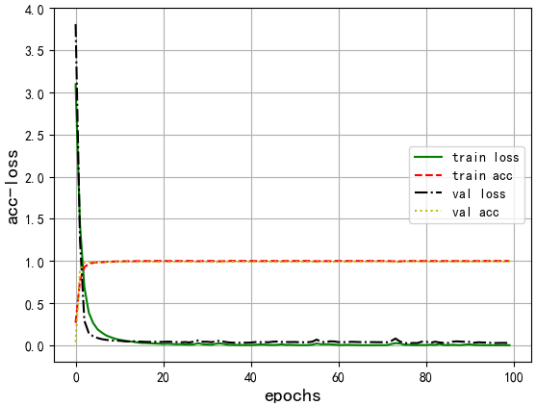

FIGURE 11. Training and validation losses by epochs using data of both feet segmented without overlap for model B.

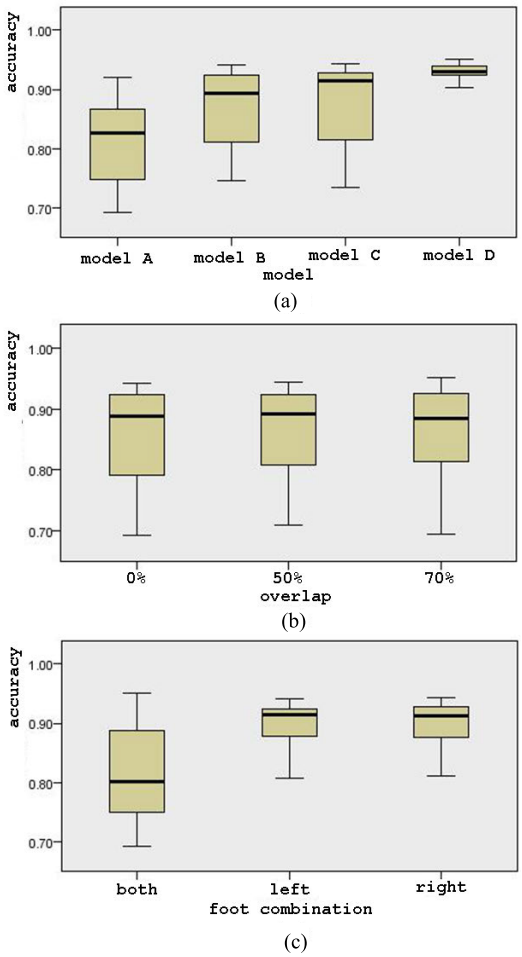

FIGURE 12. Statistical testing for model selection (Kruskal-Wallis test) by: (a) neural network architecture (b) segment overlap (c) insole combination.

big data, it allowed for using 6 102 137 parameters, which would not be possible with small datasets. With deep neural networks, the amount of data for training is the leading factor for obtaining high accuracy. However, despite reporting high accuracies in the order of more than 90%, existing studies, e.g. [22, 25], confirm that lack of enough training data hinders the algorithm results.

TABLE 3. Performance for different overlaps and sides - model A and model B.

insole side	model	overlap 0% (avg.acc., avg.loss, std)		
both	A	73.2%	1.8	0.028
both	B	77.5%	1.2	0.029
left	A	83.6%	0.9	0.031
left	B	92.2%	0.4	0.016
right	A	85.7%	0.8	0.034
right	B	90.7%	0.5	0.016

insole side	model	overlap 50% (avg.acc., avg.loss, std)		
both	A	74.0%	2.5	0.018
both	B	78.4%	1.4	0.020
left	A	86.0%	0.9	0.045
left	B	90.9%	0.5	0.017
right	A	87.0%	0.9	0.027
right	B	91.3%	0.5	0.018

insole side	model	overlap 70% (avg.acc., avg.loss, std)		
both	A	72.8%	3.9	0.038
both	B	80.9%	1.2	0.036
left	A	83.3%	1.9	0.035
left	B	90.8%	0.5	0.022
right	A	85.1%	1.4	0.036
right	B	92.1%	0.5	0.011

TABLE 4. Performance for different overlaps and sides - model C.

insole side	overlap 0% (avg.acc., std)	
both	77.4%	0.027
left	91.5%	0.011
right	92.9%	0.008

insole side	overlap 50% (avg.acc., std)	
both	81.4%	0.0095
left	92.1%	0.0081
right	92.7%	0.0111

insole side	overlap 70% (avg.acc., std)	
both	80.1%	0.021
left	92.8%	0.006
right	91.4%	0.012

TABLE 5. Performance for different overlaps - model D, both sides.

overlap 0% (avg.acc., std)		overlap 50% (avg.acc., std)		overlap 70% (avg.acc., std)	
92.9%	0.007	92.8%	0.016	93.3%	0.009

B. SINGLE-SIDE VS BOTH-SIDE ACCURACY

Surprisingly, models A-C showed much higher accuracy when single-insole data were supplied. This result could be attributed to several possible factors. First, system bias might take place, due to the imperfect synchronization between the acquisition of left and right insole data, as well as the inaccuracies between sampling rates of the left and right motion sensor. As a result, temporal relationships between

TABLE 6. Comparison of recent studies for wearable gait recognition.

study	modalities	subjects	sessions	accuracy
our study	9 force sensors, accelerometer, gyroscope	59	10 min	93.3%, identification
Choi et al, 2019 [1]	8 force sensors and accelerometer	14	3 min	95%, identification
Yeh et al, 2018 [2]	6 force sensors	14	-	99.6% for authentication

left- and right-side features could be distorted. As evidence of this hypothesis, the accuracies shown by model D are significantly higher than both-side accuracies of models A-C. In that, the feature-level fusion between the left and right insole data makes the classifier insensitive to the lack of sensor-level synchronization between the two insoles. Possibly, synchronization at a signal sample level would lead to improved recognition. Such synchronization, however, may not be practical in terms of energy preservation in the wearable device. Also, synchronizing the timestamp clocks of independent sensor nodes communicating over wireless interface could be challenging due to the unpredictable delays typical for wireless transmission. Another likely reason behind the lower both-side accuracies is that when having the double-limb data, the number of features doubles while the number of segments of training data remains the same. Thus, some part of the features between the left and right sides become redundant, which works against accuracy; a possible significant increase in the volume of the training set could lead to the opposite effect.

C. EVALUATION OF MODELS AND WINDOW OVERLAP

Intuitively, using a more complex feature extractor leads to improved accuracy. Thus, model B consisting of five cascaded single units for feature extraction, shows generally better performance than model A that contains a baseline CNN. Also, the two-cascade structure of model D shows a higher accuracy compared to model C and B. As expected, ELM based algorithms show similar or better accuracy compared to models B and A. The fact that the results of the 50% segmentation overlap are, in some cases, higher than those of the 70% overlap can be attributed to instabilities of accuracy with small datasets.

D. OPEN QUESTIONS

In this study we demonstrated the application of CNN for wearable-device-based person recognition. Despite the promising results, the way neural networks interpret signals is considered a black box and it was therefore not possible to explain what gait features were important. Brute-force procedures with hand-crafted features could shed light on the most important features; however, these methods might be biased and lacking accuracy. Instead, Layer-Wise Relevance Propagation is a new technique that could reveal what variables at exact time instances of the gait cycle are responsible for the output [34].

As to accuracy, several measures can help to improve recognition performance: (1) more data from more subjects; with the achievements in generative models, it might be possible to generate large synthetic datasets for training under the open-set gait recognition paradigm; (2) apply activity recognition to filter out the non-walking segments of the signal; (3) transform the signal into a new orientation-independent reference system; (4) use cycle detection to segment the signal and normalize it through fixed length, zero mean and unit variance vectors; (5) ensure accurate sampling rates for all sensors, precise synchronization at sample level, avoid saturation; (6) model the temporal transitions with recurrent connections, for example by adding a Long Short-Term Memory layer; (7) apply augmentations on the training dataset to avoid overfitting.

Among limitations of the current study was the lack of synchronization between left and right insoles, as well as between force and inertial sensors. Also, we did not design the system to adapt dynamically for new users without complete re-training of the model. In application aspect, intuitively, the next step would be to explore the feasibility of gait recognition with longitudinal data. These matters are to be addressed in future studies.

VII. CONCLUSION

In this work, we explored the feasibility of person identification by level walking outdoors through foot-mounted inertial and force sensors. Compared to previous studies [1], [2], we made a step further by having collected a large multimodal sensor insole dataset. We used a custom sensor insole and logger to collect walking data from 59 participants outdoors. Because of the higher number of subjects, long recording sessions and multiple modalities reflecting both kinetic and kinematic characteristics, the presented dataset is appropriate to supply machine learning methods. We explored the performance of four neural network architectures for different segmentation overlaps. Results confirm that identity recognition through multimodal sensor insoles is a viable option for increasing the security in Body Sensor Networks.

ACKNOWLEDGEMENT

The authors would like to thank Gergana Shehtova for help in preparing the illustration of Fig. 3.

REFERENCES

[1] S.-I. Choi, J. Moon, H.-C. Park, and S. T. Choi, "User identification from gait analysis using multi-modal sensors in smart insole," *Sensors*, vol. 19, no. 17, p. 3785, Aug. 2019.

[2] K.-H. Yeh, C. Su, W. Chiu, and L. Zhou, "I walk, therefore i am: Continuous user authentication with plantar biometrics," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 150–157, Feb. 2018.

[3] J. Blasco, T. M. Chen, J. Tapiador, and P. Peris-Lopez, "A survey of wearable biometric recognition systems," *ACM Comput. Surv.*, vol. 49, no. 3, pp. 1–35, Dec. 2016.

[4] S. Sprager and M. Juric, "Inertial sensor-based gait recognition: A review," *Sensors*, vol. 15, no. 9, pp. 22089–22127, 2015.

[5] P. Connor and A. Ross, "Biometric recognition by gait: A survey of modalities and features," *Comput. Vis. Image Understand.*, vol. 167, pp. 1–27, Feb. 2018.

[6] S. A. Niyogi and E. H. Adelson, "Analyzing and recognizing walking figures in XYT," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1994, pp. 469–474.

[7] J. Han and B. Bhanu, "Statistical feature fusion for gait-based human recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun./Jul. 2004, p. 2.

[8] Y. Liu, J. Zhang, C. Wang, and L. Wang, "Multiple HOG templates for gait recognition," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 2930–2933.

[9] K. Ivanov *et al.*, "Design of a sensor insole for gait analysis," in *Intelligent Robotics and Applications. ICIRA* (Lecture Notes in Computer Science), vol. 11743, H. Yu, J. Liu, L. Liu, Z. Ju, Y. Liu, and D. Zhou, Eds. Cham, Switzerland: Springer, 2019. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-27538-9_37, doi: 10.1007/978-3-030-27538-9_37.

[10] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Process. Mag.*, vol. 33, no. 4, pp. 49–61, Jul. 2016.

[11] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. Makela, and H. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Mar. 2005, pp. ii/973–ii/976.

[12] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *Proc. 6th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Oct. 2010, pp. 306–311.

[13] F. Juefei-Xu, C. Bhagavatula, A. Jaech, U. Prasad, and M. Savvides, "Gait-ID on the move: Pace independent human identification using cell phone accelerometer dynamics," in *Proc. IEEE 5th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2012, pp. 8–15.

[14] C. Nickel, C. Busch, S. Rangarajan, and M. Mobius, "Using hidden Markov models for accelerometer-based biometric gait recognition," in *Proc. IEEE 7th Int. Colloq. Signal Process. Appl.*, Mar. 2011, pp. 58–63.

[15] H. M. Thang, V. Q. Viet, N. Dinh Thuc, and D. Choi, "Gait identification using accelerometer on mobile phone," in *Proc. Int. Conf. Control, Autom. Inf. Sci. (ICCAIS)*, Nov. 2012, pp. 344–348.

[16] M. Muaaz and R. Mayrhofer, "An analysis of different approaches to gait recognition using cell phone based accelerometer," in *Proc. 11th Int. Conf. Adv. Mobile Comput. Multimedia (MoMM)*, Vienna, Austria. New York, NY, USA: Association for Computing Machinery, Dec. 2013. [Online]. Available: https://dl.acm.org/doi/proceedings/10.1145/2536853, doi: 10.1145/2536853.2536895.

[17] Y. Zhong, Y. Deng, and G. Meltzner, "Pace independent mobile gait biometrics," in *Proc. IEEE 7th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Arlington, VA, USA, 2015, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/7358784, doi: 10.1109/BTAS.2015.7358784.

[18] N. Neverova, C. Wolf, G. Lacey, L. Fridman, D. Chandra, B. Barbello, and G. Taylor, "Learning human identity from motion patterns," *IEEE Access*, vol. 4, pp. 1810–1820, 2016.

[19] Y. Zhong and Y. Deng, "Sensor orientation invariant mobile gait biometrics," in *Proc. IEEE Int. Joint Conf. Biometrics*, Sep. 2014, pp. 1–8.

[20] R. Subramanian and S. Sarkar, "Evaluation of algorithms for orientation invariant inertial gait matching," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 304–318, Feb. 2019.

[21] E. Mordini and D. Tzovaras, *Second Generation Biometrics: The Ethical, Legal and Social Context*. Amsterdam, The Netherlands: Springer, 2012.

[22] Q. Zou, Y. Wang, Q. Wang, Y. Zhao, and Q. Li, "Deep learning-based gait recognition using smartphones in the wild," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3197–3212, Apr. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9056812, doi: 10.1109/TIFS.2020.2985628.

[23] M. Gadaleta and M. Rossi, "IDNet: Smartphone-based gait recognition with convolutional neural networks," *Pattern Recognit.*, vol. 74, pp. 25–37, Feb. 2018.

[24] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication," *Pattern Recognit.*, vol. 47, no. 1, pp. 228–237, Jan. 2014.

[25] J. Gao, P. Gu, Q. Ren, J. Zhang, and X. Song, "Abnormal gait recognition algorithm based on LSTM-CNN fusion network," *IEEE Access*, vol. 7, pp. 163180–163190, 2019.

[26] Z. Mei, K. Ivanov, L. Lubich, and L. Wang, "Recognition of pes cavus foot using smart insole: A pilot study," in *Intelligent Robotics and Applications. ICIRA* (Lecture Notes in Computer Science), vol. 11742, H. Yu, J. Liu, L. Liu, Z. Ju, Y. Liu, and D. Zhou, Eds. Cham, Switzerland, 2019. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-27535-8_58, doi: 10.1007/978-3-030-27535-8_58.

[27] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to Biometrics*. New York, NY, USA: Springer, 2011.

[28] P. Terrier, "Gait recognition via deep learning of the center-of-pressure trajectory," *Appl. Sci.*, vol. 10, no. 3, p. 774, Jan. 2020.

[29] S. Z. Li and A. K. Jain, *Encyclopedia of Biometrics*, 2nd ed. Boston, MA, USA: Springer, 2015. [Online]. Available: https://link.springer.com/referencework/10.1007/978-1-4899-7488-4

[30] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 212–220, doi: 10.1109/CVPR.2017.713.

[31] Y. Zhang, G. Pan, K. Jia, M. Lu, Y. Wang, and Z. Wu, "Accelerometer-based gait recognition by sparse representation of signature points with clusters," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1864–1875, Sep. 2015.

[32] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, May 2017.

[33] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[34] F. Horst, S. Lapuschkin, W. Samek, K.-R. Müller, and W. I. Schöllhorn, "Explaining the unique nature of individual gait patterns with deep learning," *Sci. Rep.*, vol. 9, no. 1, p. 2391, Dec. 2019.

KAMEN IVANOV (Associate Member, IEEE) received the bachelor's degree in electronics engineering from the Technical University of Sofia, Sofia, Bulgaria, in 2003. He is currently pursuing the Ph.D. degree in biomedical electronics engineering with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. His research interests include low-power sensors and hardware, sparse signal recovery, compressed sensing, machine learning, and their applications in wearable medical devices.

ZHANYONG MEI received the Ph.D. degree from the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, in 2013. He is currently a Lecturer at the Chengdu University of Technology. His research interests include biomedical signal processing, digital image processing, and gait analysis.

MARTIN PENEV received the M.Sc. degree from the Technical University of Sofia, Bulgaria, in 2000, where he is currently pursuing the Ph.D. degree with the Department of Radio Communications and Video Technologies. His research interests include digital image processing and pattern recognition.

LUDWIG LUBICH received the M.S. degree in electronic engineering and the Ph.D. degree in wireless receivers from the Technical University of Sofia, in 1994 and 2015, respectively. He is currently an Assistant Professor in radio communication devices and systems with the Technical University of Sofia.

OMISORE OLATUNJI MUMINI (Member, IEEE) received the B.Tech. and M.Tech. degrees in computer science and the Ph.D. degree in pattern recognition and intelligent systems from the University of Chinese Academy of Sciences. He was a Software Engineer, a System Analyst, and a University Lecturer. He has nine years of research experience. He currently holds a postdoctoral position with the CAS Key Laboratory for Health Informatics, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. His research interests include constraints control of flexible and MIS robotic systems, computational intelligence, and digital libraries with specialty in data and knowledge mining.

SAU NGUYEN VAN (Student Member, IEEE) received the B.S. and M.S. degrees in computer science from the VNU University of Engineering and Technology, in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. He has been a Lecturer with the University of Fire Fighting and Prevention, Vietnam, since 2012. His research interests include deep learning, natural language processing, digital image processing, biomedical information technology, and the Green Internet of Things.

YAN YAN (Member, IEEE) received the B.Eng. and M.Sc. degrees in instrument engineering from the Harbin Institute of Technology, in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree in computer science. He was a Research Assistant with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, from 2012 to 2014. He has been an Honorary Research Assistant with the Department of Computer Science, University of Liverpool. His research interests include biomedical signal processing, machine learning, dynamical systems, and topological data analysis.

LEI WANG (Member, IEEE) received the B.Eng. degree in information and control engineering and the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 1995 and 2000, respectively. He was with the University of Glasgow and Imperial College London, from 2000 to 2008. He is currently a Full Professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, as a Professor and the Deputy Director of the Institute of Biomedical and Health Engineering. He has published over 200 scientific articles and authored four book chapters and holds 60 patents. His current research interests include body sensor networks, digital signal processing, and biomedical engineering.