RESEARCH ARTICLE

Gender Detection Based on Gait Data: A Deep Learning Approach With Synthetic Data Generation and Continuous Wavelet Transform

ERHAN DAVARCI[ID] AND EMIN ANARIM[ID]

Electrical and Electronics Engineering Department, Boğaziçi University, 34342 Istanbul, Turkey

Corresponding author: Erhan Davarci (erhan.davarci@boun.edu.tr)

ABSTRACT Smart devices equipped with various sensors enable the acquisition of users' behavioral biometrics. These sensor data capture variations in users' interactions with the devices, which can be analyzed to extract valuable information such as user activity, age group, and gender. In this study, we investigate the feasibility of using gait data for gender detection of users. To achieve this, we propose a novel gender detection scheme based on a deep learning approach, incorporating synthetic data generation and continuous wavelet transform (CWT). In this scheme, the real dataset is first divided into training and test datasets, and then synthetic data are intelligently generated using various techniques to augment the existing training data. Subsequently, CWT is used as the feature extraction module, and its outputs are fed into a deep learning model to detect the gender of users. Different deep learning models, including convolutional neural network (CNN) and long short-term memory (LSTM), are employed in classification. Consequently, we evaluate our proposed framework on different publicly available datasets. On the BOUN Sensor dataset, we obtain an accuracy of 94.83%, marking a substantial 6.5% enhancement over the prior highest rate of 88.33%. Additionally, we achieve 86.27% and 88.15% accuracy on the OU-ISIR Android and OU-ISIR Center IMUZ datasets, respectively. Our experimental results demonstrate that our proposed model achieves high detection rates and outperforms previous methods across all datasets.

I. INTRODUCTION

With the rapid advancement of technology, smart devices such as smartphones, wearable watches, and tablets have become an integral part of daily life. Consequently, studies on users' interaction patterns and usage habits with these devices have significantly increased. Motion sensors, like accelerometers and gyroscopes embedded in these smart devices, have emerged as valuable sources of information in such studies. The widespread utilization of these sensors can be attributed to two primary factors: firstly, collecting data from these sensors is both convenient and cost-effective. Secondly, these sensors enable the extraction of various characteristics and behavioral biometrics associated with the users.

The associate editor coordinating the review of this manuscript and approving it for publication was Vincenzo Conti[ID].

Several studies have highlighted the diverse applications of motion sensors, encompassing various domains such as demographics, activity, behavior detection, user authentication, keystroke, and text inference. For instance, in [1], the authors successfully determined user gender by analyzing smartphone accelerometer sensor data. Accelerometer sensor data have also been leveraged to detect behaviors associated with individuals' stress levels in [2]. Additionally, smartphone motion sensor data has been utilized in multiple studies to recognize and classify human daily activities [3], [4], [5]. On the other hand, the authors have employed sensor data from smart devices to enhance user identification and authentication systems in [6] and [7].

Gait analysis is a significant biometric feature that facilitates human identification and provides insights into physical and medical conditions. Due to the unique nature of an individual's gait, which reflects their walking style

and physical abilities, it becomes challenging to mimic the gait pattern of others [8]. Thus, gait analysis finds applications across various domains, including security, sports, surveillance, and the medical field [9], [10]. For instance, in [11], the authors underscored using sensor-based gait analysis in clinical applications for monitoring and diagnosing conditions such as Parkinson's disease.

Sensor data originating from touchscreen interactions and device handling style can be utilized to extract behavioral biometrics when the user is stationary. However, their effectiveness diminishes when the user is in motion due to the impact of movement [12]. Hence, gait data analysis is a promising candidate for biometrics extraction, like age group and gender in mobile scenarios.

The identification of gender holds significant importance in various usage scenarios [13]. For instance, customizing screens or applications on a device based on the user's gender can enhance user interaction and experience [14], [15]. Gender information can also improve personalized recommendations, enable more relevant search results, and facilitate targeted advertisements by applying gender-specific filters to users. In healthcare, knowledge of a user's gender can contribute to more accurate and tailored health support [16]. Furthermore, authentication mechanisms can leverage soft biometric traits, such as gender, to enhance performance [17], [18].

In this study, we propose a model for gender detection by analyzing the sensor data derived from users' gait patterns. Our primary objective is to explore the distinct gait characteristics between female and male users, which can be effectively captured through sensor data processing. Recent studies show that the utilization of deep neural network (DNN) models as classifiers yields superior performance compared to traditional machine learning (ML) algorithms when analyzing sensor data [19]. This excellent performance is attributed to the enhanced capability of DNN models to learn complex and nonlinear relationships from data. Thus, we train a DNN model to discriminate between female and male users.

Specifically, we employ deep learning models like convolutional neural network (CNN) and long short-term memory (LSTM) as the DNN classifier. In these models, the CNN layers extract dimensional relationships within input data samples, whereas LSTM performs sequence prediction [20]. Furthermore, we explore and compare various CNN, LSTM, and hybrid models and compare their performances. Our results indicate that the 3-Layer CNN + LSTM hybrid model outperforms the alternatives, with superior performance in detecting gender with sensor data.

During gait analysis, both time-domain and frequency-domain analyses are commonly utilized. Time-domain analysis helps us understand how data changes over time, but it has limitations in identifying underlying patterns and causes of gait behavior. On the other hand, frequency-domain analysis breaks down time-series data into its frequency

components, revealing the frequencies where gait vibrations are most prominent. Gait consists of sequential cycles, each comprising a series of events. In addition, studies [21] and [22] show that age and gender affect movement styles and various gait parameters like walking speed, step length, cycle frequency, and toe-off angle. Changes in walking patterns also impact the frequency content of gait data. For example, older individuals may show noticeable vibrations at lower frequencies in their walking, while younger individuals tend to have vibrations at higher frequencies. These reasons make frequency analysis useful for capturing distinct movement patterns during walking.

Subsequently, we adopt a feature engineering approach rather than directly feeding raw sensor data into the DNN model. This approach enhances the detection performance and captures the distinctive patterns inherent in the sensor data. In this context, we employ continuous wavelet transform (CWT) as a feature extraction technique to reveal frequency domain characteristics. Unlike the Fourier transform (FT), which primarily focuses on identifying energy distribution in different frequency bands, CWT offers a more comprehensive analysis by revealing the frequency content and their respective occurrence times in the corresponding time-series. Therefore, given the temporal nature of our data, we position the CWT before the classification. The primary objective in employing CWT is to extract trends, periodicities, and temporal changes within time-series sensor data that may not be readily apparent in the time-domain.

In DNN frameworks, the classification model's performance may be hindered by overfitting, especially when the dataset is limited. To mitigate this problem, we employ data augmentation techniques to expand the existing datasets. Specifically, we utilize traditional data augmentation, synthetic minority oversampling technique (SMOTE), and auxiliary classifier generative adversarial network (AC-GAN) approaches. Traditional data augmentation involves applying transformations such as jittering and scaling to the sensor data. In contrast, SMOTE and AC-GAN estimate the underlying data distribution and generate synthetic data accordingly. Our experimental results indicate that the synthetic data generated using AC-GAN closely resemble the original dataset.

Furthermore, we employ two strategic approaches in the synthetic data generation process. Firstly, we balance the data of different age groups within the two gender classes, improving performance. Secondly, we validate the generated synthetic data and exclude samples that belong to the opposite gender class.

Moreover, we evaluate the proposed model on multiple datasets and compare its performance against previous methods. Whereas k-fold cross-validation (CV) is commonly employed in gender detection studies, its results can be misleading due to the large number of data samples from each user in the dataset. Thus, to ensure robust evaluation and

avoid any group leakage [23], we use the leave-one-group-out CV (LOGOCV) method in performance evaluation.

The major contributions of the paper are as follows:

- We introduce a novel approach to gender detection based on gait data, employing advanced signal processing and deep learning techniques.
- We incorporate CWT as a pre-processing step, enhancing the system's performance by extracting detailed time and frequency domain information.
- To expand real datasets and eliminate overfitting, we apply various synthetic data generation techniques to time-series data and compare their performance. We also emphasize balancing age groups within each class and verification of generated data.
- We design and evaluate multiple DNN models, including a hybrid architecture combining the CNN and LSTM layers.
- We implement LOGOCV to ensure reliable performance assessment, highlighting its significance in this field.
- We evaluate our proposed model on three diverse datasets, including the largest gait databases based on sensor data, and provide a comprehensive evaluation. The experimental results indicate that our proposed model achieves high detection rates and performs better than previous methods.

The rest of this paper is structured as follows. Section II provides the literature review. Section III presents a brief theoretical background on synthetic data generation and CWT. Section IV introduces the methodology of our proposed gender detection scheme. The performance evaluation of our model on various datasets is presented in Section V. Finally, Section VI concludes this paper.

II. LITERATURE OVERVIEW

In previous studies, gender prediction on smart devices often relied on voice recordings and facial images [24], [25]. Furthermore, studies in the literature explore age and gender estimation using image-based gait information [26], [27], [28]. For instance, the authors investigated the application of computer vision and gait analysis in gender classification for forensics in [28]. The study employed video sequences captured through various modalities. Then, gender classification was performed using body keypoints extracted from these video sequences. However, these approaches require explicit user consent for camera or microphone usage, and their performance can be compromised by external factors, leading to inaccurate estimation of gait characteristics. In contrast, sensor-based gait approaches mitigate these limitations and offer implicit and more robust detection models that operate in the background [8], [29], [30].

As discussed in [31], floor sensor-based and wearable sensor-based techniques can be used in gait analysis. Floor sensor-based methods involve equipping the floor with specialized pressure and weight sensors to collect gait data. However, due to the implementation complexity and limited

performance, this approach is rarely employed in gait analysis studies. On the contrary, wearable sensor-based techniques have gained popularity in recent years, mainly because of their widespread availability in smart devices and ease of use. Accelerometers are among the most commonly used sensors in this approach. They are preferred for their high sensitivity, minimal susceptibility to external factors, and accurate data output. Conversely, studies employing gyroscopes for gait analysis have reported inconsistent results. Therefore, gyroscopes are considered supplementary information sources to enhance accelerometer-based gait analysis.

Studies that detect gender using gait analysis yields diverse findings [16]. For instance, in [21], the authors investigated the impact of age, gender, and walking speed on adults. They observed that gender differences in gait performance were significant, whereas age-related effects varied for both genders across different parameters. Additionally, a study by [22] examined the gait characteristics of 112 adults and concluded that the aging process affects males and females differently.

In their study [32], the authors presented a methodology for gender recognition by utilizing behavioral biometrics on smartphones. This research focused on gender identification using gait data extracted from the smartphone's embedded accelerometer and gyroscope sensors. The proposed approach involved calculating the curvature of the gait signals. The data collection phase involved subjects walking with the smartphone placed in their trouser pocket, acquiring 252 gait data samples from 42 individuals. In performance evaluation, a 5-fold CV was used, and the bagging classifiers yielded accuracy rates ranging from 73% to 77% across different walking scenarios. Another study [33] explored gender classification by analyzing the human gait cycle based on accelerometer signals. The authors utilized the OU-ISIR dataset, comprising data from 744 users, and divided it into a 70% training set and a 30% test set. The study reported 68.2% and 65% accuracy rates using a logistic regression (LR) classifier for different walking sequences.

In a research study focusing on gender detection using a deep learning approach, the model's effectiveness was evaluated across distinct age groups [34]. With data from 640 users, the study achieved an overall accuracy of 82.8% using the inter-subject monte-carlo CV technique. Another study [35] demonstrated the feasibility of recognizing gender, age, and height attributes using a single inertial sensor with a sample size of 26 subjects. The classification utilized a random forest (RF) classifier, achieving accuracy rates of up to 85.5% for gender prediction in a subject-wise CV scenario.

Furthermore, in a comprehensive analysis of gait-based age and gender estimation approaches conducted by [16], it was demonstrated that the most promising outcome for gender estimation, with an accuracy rate of 75.8%, was achieved using a temporal convolution network. In this study, the OU-ISIR dataset was utilized for training data, whereas a distinct and separate test set was employed

for evaluation. Another work [36] introduced a method for gender prediction based on activity, utilizing a dataset collected from smartphones carried in users' pockets. The authors attained accuracy rates of up to 95% using various classifiers on the MotionSense dataset, encompassing data from 24 users. However, their dataset's age range needed to be expanded, and no information was provided regarding the potential overlap of user samples between the training and test sets. Similarly, [37] explored gender recognition using gait data of 109 subjects and reported the highest accuracy of 96.3% using the bagged tree classifier with CV. However, CV may not reflect real-world performance, as the training and test sets can contain data from the same user.

Several works in the literature utilize motion sensor data and gait analysis for gender estimation. However, these works are subject to certain notable limitations. Typically, the gait data are examined in small groups of subjects, often lacking a balanced representation of both females and males and focusing on narrow and specific age ranges [38]. Due to using private datasets, it becomes impractical to compare the findings or validate the models using related datasets. Furthermore, performance evaluation often relies on non-robust assessment methods, and more details should be provided regarding the separation of training and test datasets.

Moreover, certain studies aimed at analyzing gait data first involve identifying gait cycles, then extracting features from each cycle. This complex approach yields precise outcomes solely when gait cycles are accurately identified. Nevertheless, factors such as varying walking speeds among users and utilizing large datasets spanning diverse age ranges can impose challenging limitations on gait cycle identification.

To overcome these limitations, we develop a robust gender detection scheme and evaluate it using publicly available datasets characterized by a large number of subjects, balanced gender distributions, and diverse age ranges. We aim to ensure a more robust and comprehensive performance evaluation when comparing different models by utilizing such datasets and suitable evaluation methods. To the best of our knowledge, our work is the first to employ a hybrid approach for gender detection, integrating well-established techniques such as synthetic data generation, CWT, and DNN models. Whereas CWT and CNN have been utilized in previous studies for time-series classification tasks, such as human activity recognition [39], our research uniquely applies these methodologies in sensor-based gender detection.

III. BACKGROUND INFORMATION

This section provides the theoretical background of the proposed gender detection scheme. Initially, we describe synthetic data generation techniques, followed by a comprehensive explanation of the CWT.

A. SYNTHETIC DATA GENERATION

The performance of ML models is significantly influenced by the quality and size of the dataset utilized [40]. Deep learning models, in particular, require substantial amounts of high-quality data to train effectively, as inadequate data can lead to overfitting. However, real-world datasets are often constrained in size and diversity due to the inherent challenges and costs associated with data collection.

To overcome this challenge, researchers often utilize data augmentation and synthetic data generation techniques to expand the size of the dataset by generating additional samples. Although these two methods are sometimes evaluated under the umbrella term of data augmentation, they differ in the approach used to create the additional data. Data augmentation involves applying transformations, such as adding noise or rotating, to existing data samples to form further examples. On the other hand, synthetic data generation entails creating entirely new artificial data samples that exhibit similar characteristics and statistical properties to the original data.

Consequently, the expansion of time-series datasets using these techniques enhances the performance of deep learning models through improved generalization, reduced overfitting, facilitated feature learning, and increased robustness. By providing a more diverse and representative training set, these techniques enable the model to learn and generalize more effectively to unseen data.

In the subsequent subsections, we provide a concise overview of data augmentation and synthetic data generation techniques employed in our study.

1) TRADITIONAL DATA AUGMENTATION TECHNIQUES

This section provides an overview of traditional augmentation techniques commonly employed in various fields, such as image processing, and adapted for use in time-series data [41].

Jittering, which involves adding noise to data, is one of the most widely used augmentation techniques in the time-series context. As sensor data often exhibit noise, jittering leverages the existing noise in the data to simulate and generate new samples. Typically, Gaussian noise is added to each time step during the jittering process.

Scaling involves altering the magnitude of a time-series signal while preserving its overall shape. **Magnitude Warping** is a scaling technique that applies variable scaling to different samples within the time-series.

Another scaling technique, known as **Time Warping**, involves stretching and shortening the time intervals of the time-series signal. Unlike magnitude warping, which alters the magnitude of the time-series, time warping modifies the temporal location of the time-series. On the other hand, **Rotation** can be applied to time-series data by utilizing a rotation matrix with a specified angle.

These techniques and their definitions and parameters are summarized in Table 1. Applying these techniques to time-series sensor data can enhance its robustness and introduce various interpretations. For example, jittering can be regarded as a means of simulating additive sensor noise,

whereas scaling can simulate walking motions of different sizes. Since the placement of sensors and smart devices can impact the obtained sensor readings, rotation can be seen as a way to simulate different sensor placements.

One crucial consideration when applying these techniques to time-series data is carefully selecting parameters and probabilities for each method. Excessive manipulation of the data may distort the time-series information to the extent that the class information is lost. Additionally, other augmentation techniques, such as cropping and permutation, are not considered in this work. These techniques can significantly disrupt the gait patterns in the time-series data, potentially resulting in the generation of invalid synthetic samples.

2) SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE

SMOTE is a widely used approach to address class imbalance in ML datasets. It mainly aims to increase the representation of the minority class by generating synthetic data [42]. The fundamental process of SMOTE involves selecting a random data sample from the minority class and identifying its k nearest neighbors. Subsequently, one of these neighbors is randomly chosen, and a synthetic data instance is created by interpolating between the selected sample and its neighbor. Mathematically, SMOTE can be defined as follows:

$$x_{new} = x_i + r(x_i - x_j) \quad for \ r \in [0, 1], \qquad (1)$$

where x_{new} is the generated synthetic sample, x_i is the sample from minority class, x_j is the chosen nearest neighbor, r is a random number between 0 and 1.

The procedure mentioned above can generate the desired number of synthetic data samples for the minority class. This approach proves effective as it develops new synthetic samples near existing instances of the same class in the feature space. However, in scenarios where the minority class exhibits significant overlap with the majority class, the performance improvement achieved by SMOTE may be limited, as synthetic data are generated without considering the majority class. Thus, modified versions of SMOTE, including Borderline-SMOTE and ADASYN, are proposed to increase SMOTE performance.

Borderline-SMOTE specifically targets the generating synthetic samples near the decision boundary between classes [43]. Focusing on these boundary regions aims to enhance the discriminatory ability of the generated synthetic data. On the other hand, Adaptive Synthetic Sampling (ADASYN) dynamically adjusts the number of synthetic samples to be generated based on the density of different regions in the feature space to provide a more balanced representation of the data [44].

3) AUXILIARY CLASSIFIER GENERATIVE ADVERSARIAL NETWORK

Generative adversarial networks (GANs), introduced in [45], are a prominent deep learning approach widely used for generating synthetic data in various domains. The primary objective of GANs is to replicate a given data distribution by synthesizing new data samples that closely resemble the input data distribution. GANs have two main network components: the Generator (G) and the Discriminator (D) networks.

The generator network aims to generate synthetic (fake) data samples that closely resemble the actual data distribution. In contrast, the discriminator network distinguishes between real and synthetic data. These two networks are trained in an adversarial manner, where the G tries to deceive the discriminator, and the D aims to classify the data accurately. This training process involves iterative updates to the weights of both networks using backpropagation to reach an equilibrium point. Ideally, the generator should be able to generate synthetic samples that follow the same distribution as the real data, and the discriminator should no longer be able to differentiate between real and synthetic samples.

From a mathematical perspective, G and D engage in a zero-sum game inspired by game theory to generate entirely new data. The D is trained to maximize the log-likelihood corresponding to the real data and minimize the log-likelihood corresponding to the generated data, as defined in (2). On the other hand, the G is trained to minimize the second term in (2), aiming to increase the probability of the generated data being classified as real.

$$L = E[\log D(x)] + E[\log(1 - D(G(z)))], \qquad (2)$$

where z represents randomly generated noise vector that serves as input to G and $X_{generated} = G(z)$ is a generated sample by G. On the other hand, $D(x)$ denotes the probability that the input x originates from the real data rather than being generated.

AC-GAN is a class-conditional extension of the GANs framework that incorporates an additional auxiliary classifier in the discriminator network [46]. This modification allows AC-GAN to generate synthetic data samples that capture the real data's statistical properties and follow the class distribution. The objective function of AC-GAN consists of two components: the log-likelihood of the correct data source, L_S, and the log-likelihood of the right class label, L_C. These components are defined as follows:

$$L_S = E[\log D(x)] + E[\log(1 - D(G(z)))] \qquad (3)$$
$$L_C = E[\log D(x, c)] + E[\log(1 - D(G(z, c)))] \qquad (4)$$

In the AC-GAN, the G takes a random noise vector z and a class label c as input, generating a synthetic sample $X_{generated} = G(z, c)$. The discriminator network is trained to maximize $L_C + L_S$, whereas the generator network is trained to maximize $L_C - L_S$. Fig. 1 shows the typical AC-GAN architecture.

The advantages of AC-GAN include the following:

- The potential to learn more discriminative representations.
- The ability to generate diverse and high-quality synthetic data.
- The capability to control the class distribution of the generated samples.

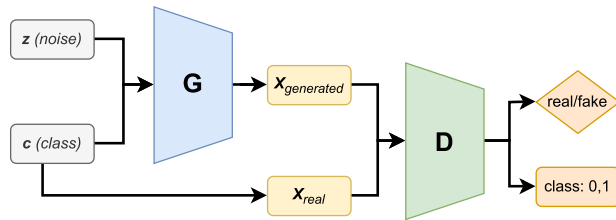| Augmentation Technique	Definition	Parameters
Jittering	$x'(n) = \{x_1 + n_1, ..., x_t + n_t\}$	n_i : noise added at each sample
Scaling	$x'(\alpha) = \{\alpha x_1, ..., \alpha x_t\}$	α : scaling coefficient
Magnitude Warping	$x'(\alpha) = \{\alpha_1 x_1, ..., \alpha_t x_t\}$	α_i: variable scaling coefficients
Time Warping	$x'(\tau) = \{x_{\tau(1)}, ..., x_{\tau(t)}\}$	τ : magnitude of the warp
Rotation	$x'(R) = \{Rx_1, ..., Rx_t\}$	R : rotation matrix

FIGURE 1. AC-GAN architecture.

However, it is essential to note that AC-GAN models require substantial data for practical training, and ensuring stability during the training process can be challenging. Our study employs the AC-GAN model and carefully fine-tunes its parameters to achieve practical training and generate diverse synthetic time-series data.

B. CONTINUOUS WAVELET TRANSFORM

The analysis of time-series data typically involves two distinct approaches: time-domain and frequency-domain analysis. Time-domain analysis examines data variation over time, allowing for an understanding of how the time-series changes or evolves. On the other hand, frequency-domain analysis transforms the data into its frequency components, providing insights into the specific frequencies in the time-series. By decomposing the time-series into its frequency components, frequency domain analysis reveals additional information that may not be readily apparent in the time-domain.

FT is a widely used transformation method that converts a time-series into its frequency components [47]. However, FT treats time and frequency as fixed entities, disregarding any temporal information about the frequency components. Whereas FT accurately reveals the frequency content of stationary signals whose frequencies stay the same over time, it falls short when analyzing non-stationary time-series data with frequency characteristics that vary over time. This limitation becomes particularly relevant when examining signals with dynamic frequency behavior.

CWT serves as an alternative analysis method to FT by offering a simultaneous representation of time and frequency information. This capability allows for a time-frequency localization of the time-series, providing a more comprehensive understanding of its characteristics [48]. The CWT of a

signal $x(t)$ can be mathematically defined as follows:

$$CWT(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t)\psi^*(\frac{t - \tau}{s})dt, \qquad (5)$$

where τ and s are the transition and scale parameters, respectively. The $\psi(t)$ is the wavelet function, also called the mother wavelet, and the symbol (*) denotes the operation of the complex conjugate. The transition parameter, τ, is associated with shifting the mother wavelet across the time-series, enabling the movement of differently-scaled wavelets from the beginning to the end of the time-series. On the other hand, the scale parameter, s, determines the extent of scaling applied to the time-series and is inversely proportional to frequency. Smaller scales correspond to compressed time-series, capturing high-frequency components, whereas larger scales correspond to stretched-out time-series, highlighting low-frequency components.

In calculating the CWT, the process begins by selecting the mother wavelet. The analysis starts with $s = 1$ and proceeds by incrementing s, moving from high to low frequencies. The mother wavelet is initially positioned at the beginning of the time-series, corresponding to time $t = 0$. The wavelet function at scale $s = 1$ is multiplied by the time-series and integrated over all time points. Subsequently, the wavelet is shifted by τ units to the right, and this process is repeated until the wavelet reaches the end of the time-series. Finally, the CWT of the time-series is obtained by repeating this procedure for each value of s.

In CWT, choosing a suitable mother wavelet is crucial, as it directly impacts the effectiveness of the analysis. Various types of mother wavelets are available, and the selection process is typically guided by the similarity between the time-series signal under investigation and the mother wavelet [49]. Based on this, our study aims to identify the most appropriate mother wavelet for our specific data by considering different options and comparing their performance.

Consequently, the CWT offers a comprehensive analysis of non-stationary time-series by examining the relationship between time and frequency through wavelets of different scales. Given that motion sensor data captured by smart devices exhibits dynamic changes, making it non-stationary, we employ CWT in our study. Our proposed method involves utilizing CWT to transform the one-dimensional (1D) sensor data into two-dimensional (2D) scalogram images, which in turn allows for improved results due to the enhanced

predictive capacity of 2D neural networks. Subsequently, the outputs derived from the CWT are employed in deep learning algorithms for classification.

IV. GENDER DETECTION SCHEME USING SYNTHETIC DATA GENERATION AND CWT

This section presents a gender detection scheme based on synthetic data generation and CWT. The overall architecture of the proposed method is depicted in Fig. 2. The proposed approach comprises three main components: (i) the *data generation* module, (ii) the *feature extraction* module, and (iii) the *gender detection* module.

Synthetic data are generated in the *data generation* module to augment the existing dataset. The *feature extraction* module employs CWT on the expanded dataset to extract relevant features. Subsequently, the 2D outputs from the *feature extraction* module are fed into a DNN model within the *gender detection* module to classify the gender as either female or male accurately.

A. DATA GENERATION

After acquiring the raw sensor data, the real-world dataset is initially divided into training and test datasets. Then, synthetic data are generated to enhance the overall performance of the training process. For this purpose, a *data generation* module is utilized.

Given that the class distributions in the datasets utilized in this study are balanced, we initially generate an equal amount of synthetic data for each class to expand the training data. Subsequently, we conduct a sensitivity analysis to examine the impact of different age groups on overall performance. This analysis shows that age groups with limited data, referred to as minority classes, exhibit lower performance than overall performance. This disparity can be attributed to the insufficient representation of these groups in the dataset, hindering the DNN model from effectively capturing their distinctive characteristics. Consequently, to address this problem, we modify the synthetic data generation method to make it more intelligent and targeted.

The new data generation method operates based on **Algorithm 1**. Initially, female and male users are categorized into specific age groups based on their age. For instance, if there are n age groups, we obtain $2n$ distinct subclasses by treating each age group within the female and male data as a separate class. Next, we assess the data available for each subclass and generate additional data until an equal distribution is achieved. We successfully balance the age group distributions within the female and male classes by generating more synthetic data for subclasses with limited data. We anticipate that this approach, which ensures equal representation of each subclass in the dataset, will yield improved performance, particularly for datasets encompassing a wide range of age groups.

One of the common challenges in synthetic data generation is the possibility of the generated samples belonging to incorrect classes. If time-series is distorted excessively, it can

---

Algorithm 1 Data Generation

Input: Real training dataset X_{real};
Output: Augmented training dataset $X_{augmented}$;
1: Categorize users into specific *sub_classes* based on their age;
2: Calculate the amount of data for each sub-class c, and set the maximum value to K_{max};
3: Import the pre-trained classifier C, which is trained on real training dataset X_{real};
4: **for** $i = 1$ to $len(sub\_classes)$ **do**
5:      Set the quantity of data for i^{th} sub-class c_i to K_i;
6:      **while** $K_i \leq K_{max}$ **do**
7:          Generate a synthetic data sample x_g^i for c_i;
8:          Utilize classifier C to predict the class of x_g^i as c_p;
9:          **if** $c_i = c_p$ **then**
10:             Add the synthetic sample x_g^i to the generated dataset $X_{generated}$;
11:             K_i ++;
12:         **else**
13:             Exclude the synthetic sample x_g^i;
14:         **end if**
15:     **end while**
16: **end for**
17: Concatenate X_{real} with $X_{generated}$ to obtain $X_{augmented}$;
18: **return** $X_{augmented}$

---

lead to a loss of class information, or a newly developed data sample from one class in the AC-GAN model may be closer to the distribution of another class. Although we carefully adjust and fine-tune the parameters of the data generation algorithms, we also employ another mechanism to address this problem: the verification of generated synthetic data. Following the generation of synthetic data, each newly developed sample is subjected to a verification process. A pre-trained classifier is utilized to predict the class label of the generated samples. If the predicted class does not match the class assigned to the generated data, the sample is discarded, and a new sample is generated. This approach ensures that the dataset does not include poor-quality synthetic samples, which can negatively impact training and performance. This verification process continues until an equal number of data samples is achieved for all subclasses.

Following the data generation step, we augment our training dataset by combining the newly produced data with our existing training dataset. The block diagram in Fig. 3 illustrates this procedure. The resulting augmented training dataset is then passed to the subsequent module for feature extraction.

B. FEATURE EXTRACTION

This section presents the *feature extraction* module, which is responsible for extracting features from the augmented time-series dataset. In our previous work [50], we explored
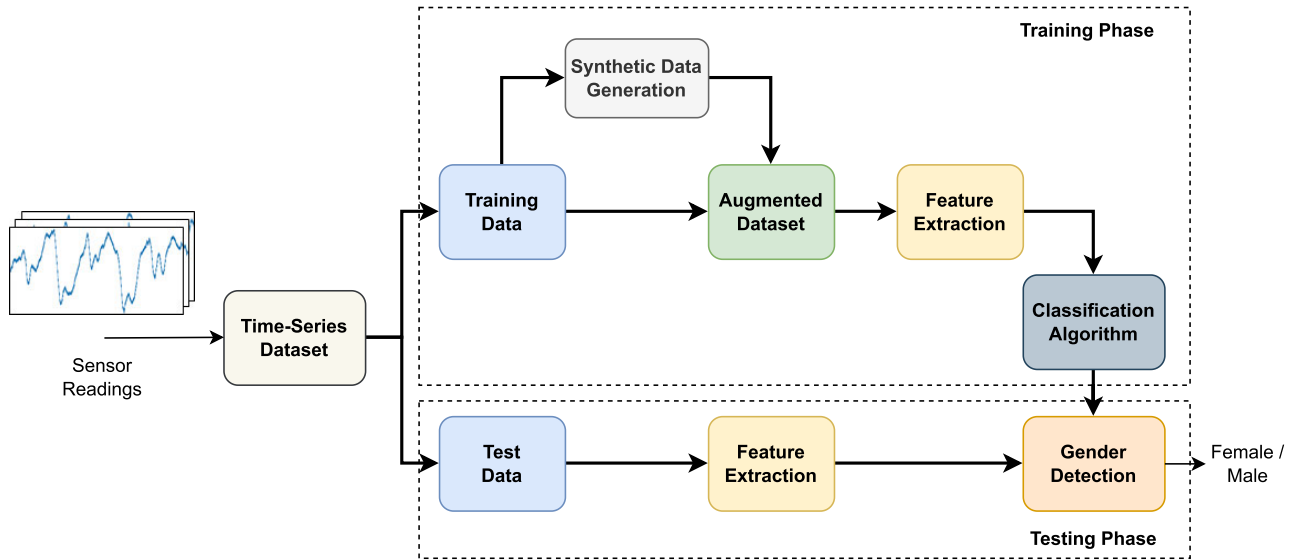
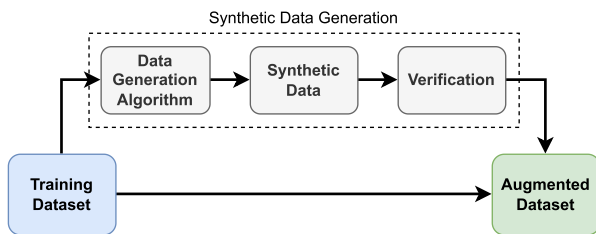**FIGURE 2.** Overall architecture of proposed gender detection scheme.

FIGURE 3. Block diagram of data generation module.

different approaches for gender detection from walking data: a traditional ML approach and a CNN-based approach.

In the traditional ML approach, we extracted statistical features from the sensor readings and applied various traditional ML algorithms for gender detection. In the CNN-based approach, we directly used the raw sensor data without any feature engineering and trained a CNN model for gender detection.

In this study, we take raw sensor data and extract specific features from them, which we then input into the DNN model. To achieve this, we prefer using frequency-domain features over time-domain ones, as time-domain features alone might not be enough to understand the patterns and reasons behind gait data fully. Additionally, since the sensor data in our study are non-stationary and follow a repeating pattern, they are well-suited for frequency-domain analysis. While FT is commonly used for frequency analysis, it cannot show the timing of the frequency components. On the other hand, the CWT provides a more complete analysis by revealing both the timing and frequency components. Utilizing CWT enables the localization of power variations, breakpoints, and transient peaks within gait data attributed to distinct walking

behaviors, which can be challenging to discern in the time-domain.

Therefore, we adopt the CWT as the feature extraction method before classification. The pseudo-code for the feature extraction process is presented in **Algorithm 2**.

Motion sensors like accelerometers and gyroscopes provide readings of three dimensions (x, y, z). Acknowledging that the sensor's orientation influences these sensor readings, any device rotation can impact these readings [31]. To address the effects of rotation, we utilize the magnitude vector obtained by computing the sum of squares of the sensor vectors. This approach helps minimize the impact of external rotations and enables a better capture of the corresponding changes in sensor readings. For example, in the case of the accelerometer, we calculate the sum of accelerations as follows:

$$A_{sum} = |A|^2 = A_x^2 + A_y^2 + A_z^2, \tag{6}$$

where, A_x, A_y, and A_z represent the acceleration values in the x, y, and z dimensions, respectively. If there are other sensor readings, such as those from a gyroscope, we similarly calculate their sum. We then add this as a fourth dimension to the sensor readings and normalize each attribute vector before further processing. A representative example of the accelerometer sensor readings is given in Fig. 4.

While specific studies first identify gait cycles and then extract features from each cycle, inaccurate identification caused by varying walking speeds and step sizes can lead to suboptimal performance. Hence, we opt for an approach that utilizes fixed-time windows for feature extraction, ensuring consistent and reliable results. For this purpose, we employ the sliding window technique to segment the sensor readings collected from different types of sensors into time windows. By using this technique, the time-series is divided into

Algorithm 2 Feature Extraction

Input: Time-series data X, sliding window size w, overlap ratio *overlap_ratio*;

Output: 2D feature matrix X_{2D};

1: Calculate the sum of squares of the attribute vectors on the x, y, z axes for each sensor and add it as a fourth dimension;

2: Normalize each *attribute_vector*;

3: Initialize an empty data variable $X_{windowed}$;

4: **for** each *attribute_vector* **do**

5:      **for** $i = 0$ to $len(attribute\_vector) - w$ **do**

6:        Extract a time-series segment x_i by applying sliding window with w and *overlap_ratio*;

7:        Add x_i to $X_{windowed}$;

8:        $i = i + (w \times (1 - overlap\_ratio))$;

9:      **end for**

10: **end for**

11: Apply CWT on $X_{windowed}$ to obtain X_{2D};
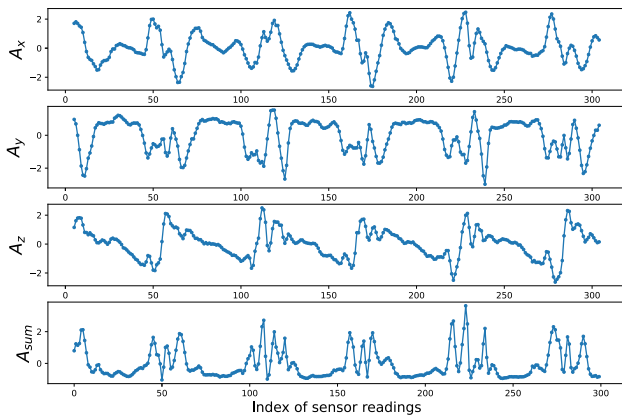
12: **return** X_{2D}

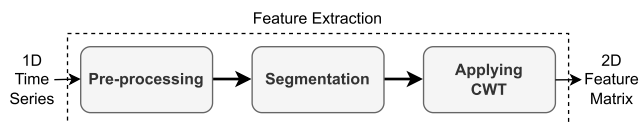**FIGURE 4.** A typical example of the sensor readings.

FIGURE 5. Block diagram of feature extraction module.

segments of uniform length using a predetermined window size and a specified overlap ratio.

Subsequently, CWT is applied to each segmented time-series to extract features within each sliding window. The output of the *feature extraction* module is a four-dimensional variable for each sensor type. It is important to note that this output is in the form of 2D data, resembling image data, and is well-suited for classification using 2D DNN models such as 2D CNN. The block diagram of the operations conducted in the *feature extraction* module is depicted in Fig. 5.

C. GENDER DETECTION

In the last module of the proposed method, gender recognition is conducted. Instead of relying on traditional ML algorithms that heavily rely on handcrafted feature extraction, which is limited by human domain expertise, recent studies have turned to deep learning approaches for their capability of automatically extracting features from raw sensor data [51]. Deep learning models offer multiple layers that can simultaneously learn various patterns in the data and handle complex problems. Hence, we employ DNN models as a classifier in our proposed method.

This module follows the procedure outlined in **Algorithm 3**. For each windowed time-series, the 2D feature data extracted by the feature extraction module is inputted into the DNN classifier. The DNN classifier is trained using labeled data from each gender class to learn the patterns associated with gender. Subsequently, the classifier outputs the probability of the time-series sample belonging to a female or male user. Finally, we employ a soft voting approach to determine the final gender prediction based on the overall input data from the user. In the soft voting approach, we calculate the cumulative probabilities for each class label and select the class label with the highest cumulative probability. The block diagram of the gender detection module is illustrated in Fig. 6.

Algorithm 3 Gender Detection

Input: 2D feature matrix X_{2D} includes coefficients for n time-series segments;

Output: Gender prediction $gender \in \{female, male\}$;

1: Load trained DNN classifier model $model_{DNN}$ from training phase;

2: **for** each $segment \in X_{2D}$ **do**

3:      Utilize $model_{DNN}$ to predict the class probabilities p_{female} and p_{male};

4:      Add p_{female} and p_{male} to the prediction set $P_{predictions}$;

5: **end for**

6: Sum up the prediction probabilities in $P_{predictions}$ as $\sum(p_{female})$ and $\sum(p_{male})$;

7: **if** $\sum(p_{female}) > \sum(p_{male})$ **then**

8:      $gender =$ female;

9: **else**

10:      $gender =$ male;

11: **end if**

12: **return** $gender$

Our proposed method involves utilizing CWT to transform the 1D sensor data into 2D scalogram images, which enhances the predictive capacity of 2D neural networks. Thus, in this module, we employ 2D DNN models. For the DNN classifier, we explore different deep learning models, including CNN and LSTM. CNN is well-suited for capturing dimensional relationships within input data samples, whereas LSTM is capable of sequence prediction and temporal feature extraction [20]. We design hybrid architectures combining
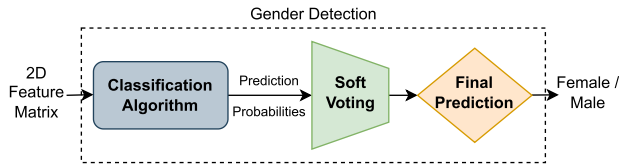
FIGURE 6. Block diagram of gender detection module.

CNN and LSTM layers to achieve a robust classifier. The different DNN models used in our study can be summarized as follows:

2-Layer CNN: This base architecture comprises two stacked CNN layers that facilitate learning patterns in the data.

CNN-LSTM: This architecture combines the CNN layer for the feature extraction process of input data with the LSTM layer to support sequence forecasting (1-Layer CNN + LSTM).

ConvLSTM: This model extends the CNN-LSTM approach by integrating the convolutional operations within the LSTM layer. This combination, known as Convolutional LSTM or ConvLSTM, effectively captures spatial and temporal dependencies in the data.

3-Layer CNN + LSTM: This is a more complex CNN-LSTM model incorporating three convolutional layers for extracting features from the input data.

In all DNN architectures above, a dropout layer is employed to mitigate overfitting, a fully connected layer interprets the learned features, and an output layer is used for making predictions.

V. PERFORMANCE EVALUATION

In this section, we present the performance evaluation of the proposed scheme. Firstly, we conduct a complexity analysis of the algorithms utilized in this study. Subsequently, we introduce the datasets employed for testing our proposed method. Furthermore, we define the performance metrics and evaluation methods used for assessing the system's performance. Finally, we present and discuss the experimental results.

A. ALGORITHM COMPLEXITY ANALYSIS

This section provides a brief overview of the computational complexity analysis of the proposed algorithms. The proposed model consists of three main modules, each comprising various algorithms. For instance, when traditional data augmentation techniques are employed in the *data generation* module, the complexity is $O(n)$. This is because traditional methods typically involve a constant number of operations, each performed n times. On the other hand, the computational complexity of SMOTE is $O(n \log n)$, whereas Borderline-SMOTE and ADASYN have a complexity of $O(n^2)$ [52].

Additionally, when the AC-GAN model is used for data generation, the complexity depends on various operations in

neural network architecture, such as matrix multiplication, convolution, and pooling. To assess the overall complexity of the model, we can focus on the convolution operation, as it typically exhibits the highest computational complexity. Since 1D convolution is utilized for time-series data, its complexity is approximately $O(nk)$, where n is the length of the data, and k is the kernel size.

In the *feature extraction* module, we utilize CWT to obtain a 2D feature matrix, which has a complexity of $O(n \log n)$ [53]. The *gender detection* module involves a DNN classifier and soft voting. Our proposed model incorporates a 3-Layer CNN + LSTM architecture. Since the CNN utilized in this module is 2D, its computational complexity is approximately $O(n^2 k^2)$, where nxn represents the image dimensions and kxk is the kernel size. Conversely, the LSTM layer is local in space and time, resulting in a complexity of $O(w)$, where w is the number of weights [54]. Additionally, the soft voting algorithm has a $O(n)$ complexity. Consequently, the overall complexity of this module is approximately $O(n^2)$, which is naturally higher than the computational complexity of a 1D CNN model.

B. DATASETS

Although several datasets available for gait analysis consist of sensor data, they exhibit certain limitations. Firstly, many of these datasets involve a limited number of subjects [9]. Secondly, data distribution between female and male gender classes is often unequal in these datasets. Lastly, the age ranges represented in these datasets can be narrow.

To address these limitations and evaluate the performance of our proposed method, we utilize publicly available datasets that do not possess the drawbacks mentioned above. Specifically, we employ the BOUN Sensor dataset [55] and the OU-ISIR dataset [56].

A mobile application was developed in BOUN Sensor dataset to collect data from smartphone users. The data collection process involved users walking and playing games while holding their smartphones. The application recorded accelerometer data on the x, y, and z axes to capture the users' movements. This way, data were collected from each user for approximately two minutes. During the data collection phase, the users had complete freedom regarding holding the phone and walking styles, and no specific guidance was provided. Commonly used Samsung and LG smartphones were employed in the experiment, and the sampling frequency was set to 100 Hz.

This dataset includes data from 60 female (average age=35.6, min=18, max=57), and 60 male (average age=30.3, min=17, max=57) users aged 17-57. The distribution of users in terms of age and gender is illustrated in Fig. 7. Whereas the gender distribution is evenly balanced in this dataset, there is variability in the age distributions for both genders. Specifically, the 20-24 and 35-39 age ranges exhibit more male users, whereas the 25-29 age range shows
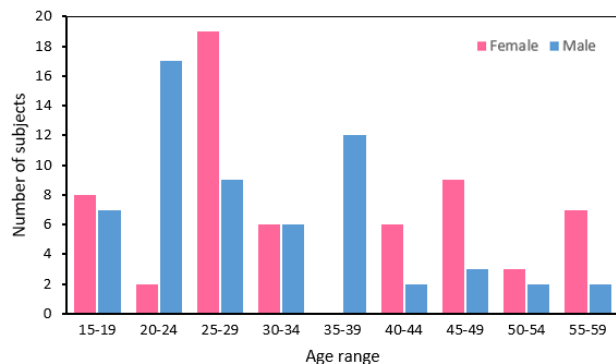
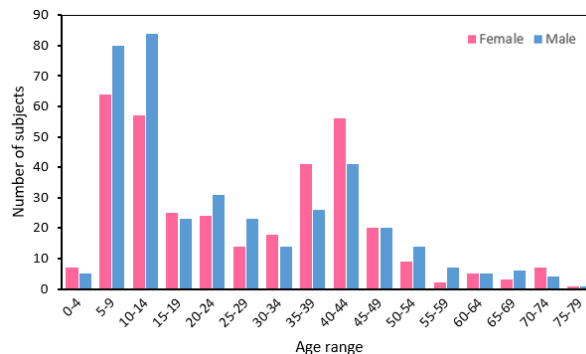**FIGURE 7.** Age and gender distribution of users in the BOUN Sensor dataset.

FIGURE 8. Age and gender distribution of users in the OU-ISIR Center IMUZ dataset.

FIGURE 9. Age and gender distribution of users in the OU-ISIR Android dataset.

the highest number of female users. Remarkably, no female users are in the 35-39 age range.

On the other hand, the OU-ISIR dataset, collected by the Institute of Scientific and Industrial Research (ISIR) at Osaka University (OU), is the largest available dataset based on inertial sensors. This dataset consists of two subsets: OU-ISIR Center IMUZ and OU-ISIR Android. In the first subset, level walk data from 744 subjects aged between 2 and 78 years were captured using the center IMUZ (Inertial Measurement Units) sensor. 355 of the subjects in this subset are female (average age=27.1, min=3, max=77), and 389 are male (average age=24.8, min=2, max=78). Fig. 8 visually represents the age and gender distribution of users in this dataset. Upon analysis, it is evident that the gender distribution between males and females is fairly consistent. However, the age distribution displays significant variation for both genders. Age ranges of 5-14 and 35-44 encompass a larger number of users. Conversely, there are fewer participants in the age groups of 0-4 and those over 55 years old.

The IMUZ sensors positioned at the center back waist of the subjects were used for obtaining the inertial signals. Each IMUZ sensor includes a triaxial accelerometer and a triaxial gyroscope and operates at 100Hz. Within this subset, two different walk sequences were extracted for each subject. In other words, data were collected while subjects walked the same designed path and returned.

In the second subset, data were collected from 408 subjects using the Motorola ME860 smartphone. In this subset, there are 189 female (average age=28.6, min=6, max=77), and 219 male (average age=24.7, min=2, max=78) users. Fig. 9 depicts the age and gender distribution within this subset. Similar to the first subset, the distribution concerning age demonstrates considerable variation for both genders. There is a notable concentration of subjects within the age ranges of 5-24 and 35-44. In contrast, the number of users is low in all other age ranges.

This dataset includes only triaxial accelerometer data. In this subset, there are four walk sequences with different labels, two corresponding to walking on flat ground and two
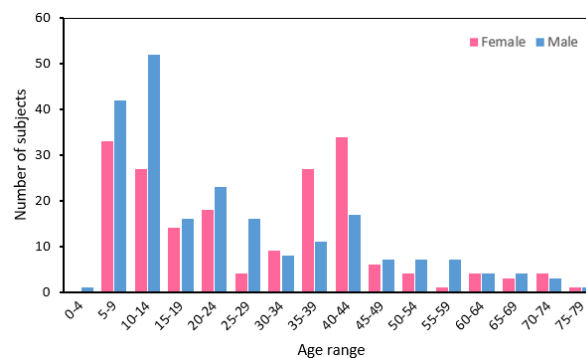
corresponding to walking on sloping ground. For our work, we focus on two walking sequences on flat ground. In the experiment, Motorola smartphones equipped with a 100 Hz accelerometer sensor were positioned at the subjects' lower back midsection. Subsequently, the sensor data were recorded as the subjects were instructed to walk along a designated path multiple times.

The non-uniformity observed in the age range of all these datasets poses a significant constraint and introduces a challenge. However, to mitigate this concern within our model, we group specific age ranges and take measures to balance the distribution among these groups.

C. PERFORMANCE METRICS
In classification problems, several commonly used performance metrics exist, such as accuracy, precision, recall, and F1-score. These metrics are derived from assessing correct and incorrect predictions. Taking females as the positive class, true positive (TP) and true negative (TN) denote the accurate classification of female and male subjects, respectively. False positive (FP) signifies male subjects incorrectly identified as female, whereas false negative (FN) represents female subjects misclassified as male. In our study,

the evaluation metrics are computed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1\text{-}score = \frac{2TP}{2TP + FP + FN} \quad (10)$$

Accuracy mainly measures the proportion of correct predictions made by the model. Precision quantifies the proportion of correct positive predictions. Conversely, recall quantifies the number of positive instances that the classifier accurately predicted among all the positive instances. The F1-score corresponds to the harmonic mean of precision and recall.

Given that the datasets employed in this study exhibit balanced classes, with each gender class holding equal significance, we primarily utilize accuracy as the main performance metric for assessing diverse models and optimizing their components for gender classification. Furthermore, we present the outcomes of our final model with all relevant metrics. Reporting these metrics allows us to demonstrate the variations in accuracy for each gender and enables potential comparisons with other studies. It is important to note that, as the positive class can be chosen as either female or male, we report the metrics for females and males separately, apart from accuracy.

D. EVALUATION METHODS

K-fold CV is a widely used method for evaluating the performance of sensor-related studies. In k-fold CV, the dataset is randomly divided into k equal-sized subsets or folds. The model is then trained on $k - 1$ folds, with the remaining fold used for testing. Leave-one-out CV (LOOCV) is a particular case of k-fold CV where k equals the number of samples in the dataset. In LOOCV, each sample is used for testing, while the rest are used for training in each fold. Although k-fold CV and LOOCV are commonly employed in similar works, their results may be misleading because of the possible overlap of data instances from the same user between the training and test sets.

To address this problem, a modified k-fold CV, known as a leave-one-user-out CV (LOUOCV) [57], has been proposed. In LOUOCV, the classifier is trained using all but one user's data, repeating this process for each user. This method evaluates the algorithm's generalization capability for unseen user data during training. However, when applying deep learning models to datasets with many users, LOUOCV may not be operationally efficient, as training a new deep learning model for each user can be time-consuming.

To mitigate computational complexity, we employ the LOGOCV method. In LOGOCV, the dataset is first divided into groups, and then one group is left out while the remaining groups are used for model training. The model is then evaluated on the omitted group. This process is repeated for each group in the dataset, and the evaluation results are averaged to obtain the final performance score.

In our study, we use 5-fold LOGOCV and randomly partition the users into five non-overlapping groups of equal size, taking into account class information to ensure balanced class distributions across groups. This means the training dataset constitutes 80% of the entire dataset, whereas the test dataset comprises 20%. For instance, when using the BOUN Sensor dataset, which encompasses 120 users, the training set includes 96 users, and the test set includes 24 users, evenly split between females and males. In the OU-ISIR datasets, the number of male users slightly outweighs female users. Consequently, we randomly exclude the surplus male users to establish balanced class distributions in these datasets. Furthermore, each user is allocated an equal amount of data during the testing phase. Consequently, both our training and test datasets are balanced not only in terms of gender labels but also in the quantity of data samples for each gender.

Numerous studies in the literature employed the k-fold CV and LOUOCV methods, comparing their respective performances. For instance, in [58], researchers investigate gender recognition from keystroke dynamics data and touchscreen swipes. They evaluate classification outcomes using 10-fold CV and LOUOCV, demonstrating that only the latter method is suitable for classifying unseen user data. Similarly, in [59], the authors explore the impact of the subject CV on the performance of human activity recognition. Their findings indicate that k-fold CV tends to overestimate system performance by approximately 16% when overlapping windows are utilized. To ensure robust and reliable performance evaluation in our study, we adopt and execute the LOGOCV method. In addition, to ensure the generalizability of the results, we employ the LOGOCV method five times. We then calculate the average accuracy results from these iterations. All the results presented in the following section are obtained using this approach.

E. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed scheme using three datasets: BOUN Sensor dataset and two subsets of the OU-ISIR dataset. Firstly, we test the proposed model and its subcomponents on the BOUN Sensor dataset. We compare different methods for each module of the proposed model and select the best one. For instance, we choose the best model in the *data generation* phase and tune its parameters. We also analyze the effects of balancing the age group distribution across classes and using validation. In the *gender detection* stage, we compare various classification models and select the most suitable one. Subsequently, we evaluate the best end-to-end model on the other two datasets step-by-step and analyze the results.

First, we employ CWT as a feature extraction method and a 2D CNN model as a classifier and compare our results with the work [50]. We compare our results with the work [50] because it also predicts gender using sensor data related to gait behavior and utilizes the BOUN Sensor dataset. In our approach, the time-series X is fed into the *feature extraction* module to extract discriminative features using **Algorithm** 2. For segmenting time-series data, we evaluate different window sizes (w) and identify $w = 128$ as the optimal choice, corresponding to a 1.28 seconds length of sensor data. It was reported in [60] that the natural cadences of human walking fall between 90 and 130 steps per minute, with a gait cycle consisting of 2 steps. Therefore, we consider an entire gait cycle in each window by setting $w = 128$. Additionally, we apply a 75% overlap ($overlap\_ratio = 0.75$) between consecutive time windows to comprehensively capture the entire gait characteristics.

Then, CWT is applied to each segmented time-series to obtain a 2D feature matrix X_{2D}, which serves as the input for the 2-Layer CNN model in the *gender detection* module. Fig. 10 presents example scalograms obtained as the output of the CWT for different genders and ages. Upon examining these scalograms, it is evident that teenage males exhibit more pronounced high-frequency components compared to teenage females. When considering middle age, males have slightly higher frequencies than females, with differences in time location. Furthermore, both males and females display decreasing frequencies with advancing age. Therefore, we can say that these scalograms exhibit distinct patterns that can be utilized for classification purposes.

Furthermore, in CWT operation, we explore various continuous wavelets, including Gaussian derivative (Gauss), complex Gaussian derivative (C-Gauss), Morlet, complex Morlet (C-Morlet), Mexican Hat, Shannon, and frequency B-spline (Fbsp), to select the most suitable mother wavelet for our data. The accuracy results for the employed wavelet functions in this study are summarized in Table 2.

As seen in Table 2, the frequency B-spline wavelet achieves the highest performance rate of 91.37%. Therefore, we use the frequency B-spline wavelet in the subsequent steps. In the study [50], both traditional ML methods and CNN approaches were applied to the BOUN Sensor dataset for gender detection, resulting in accuracy rates of 83.33% and 88.33%, respectively. When comparing the performance obtained here with the methods applied in [50], we see that the 2-Layer CNN with CWT outperformed the other two techniques.

This result is expected as CWT allows for detailed information extraction from time-series data in both the time and frequency domains. Furthermore, by providing a 2D output, CWT enables us to utilize a 2D CNN model in the classification stage, which has a higher classification capacity than a 1D CNN model.

On the other hand, despite tuning the parameters of the utilized 2D CNN model, we observe that overfitting occurs when we try to improve the model's performance

TABLE 2. The accuracy results for continuous mother wavelets.

Wavelet	Accuracy (%)	Wavelet	Accuracy (%)
Gauss1	80.53	C-Gauss4	90.41
Gauss2	89.70	C-Gauss5	88.15
Gauss3	85.06	C-Gauss6	90.66
Gauss4	88.87	C-Gauss7	88.49
Gauss5	84.70	C-Gauss8	89.70
Gauss6	78.87	Morlet	86.41
Gauss7	85.53	C-Morlet	88.78
Gauss8	89.27	Mexican Hat	90.56
C-Gauss1	80.35	Shannon	90.81
C-Gauss2	90.58	**Fbsp**	**91.37**
C-Gauss3	88.90		

by trying more iterations. This is because the 2D CNN has a higher capacity but requires more data for efficient training. To address this problem, we introduce a *data generation* module before applying CWT to expand the dataset. We experiment with different synthetic data generation methods in the *data generation* module, including traditional augmentation techniques, SMOTE methods, and the AC-GAN approach.

Firstly, we apply various traditional augmentation techniques to the dataset, such as jittering, scaling, magnitude warping, time warping, rotation, and different combinations of these methods. Fig. 11 illustrates the original version of an example time-series segment and its augmented versions obtained by applying these techniques. As an alternative to traditional augmentation methods, we explore commonly used versions of the SMOTE technique: SMOTE, Borderline-SMOTE, and ADASYN. Whereas SMOTE methods are typically employed to address class imbalance situations, in this study, we utilize them to expand the already balanced dataset further.

Subsequently, as a third alternative, we employ the AC-GAN model to generate synthetic data and augment our dataset. AC-GAN is a well-known deep learning approach used in various domains. The fundamental idea behind AC-GAN is to learn and mimic the distribution of the existing dataset, thereby generating new data examples that closely resemble the realized distribution.

In addition to the 2-Layer CNN with CWT approach used in the previous step, we apply these three data generation methods to the dataset and compare the results with those obtained in the last step. Table 3 illustrates the impact of synthetic data generation on the results. As shown in Table 3, expanding the dataset by generating new data samples synthetically generally enhances the overall performance. The primary reason for this improvement is that deep learning models, especially 2D deep learning algorithms, require high-quality data for efficient training. Therefore, augmenting the dataset prevents the model from overfitting and enhances its generalization and robustness.
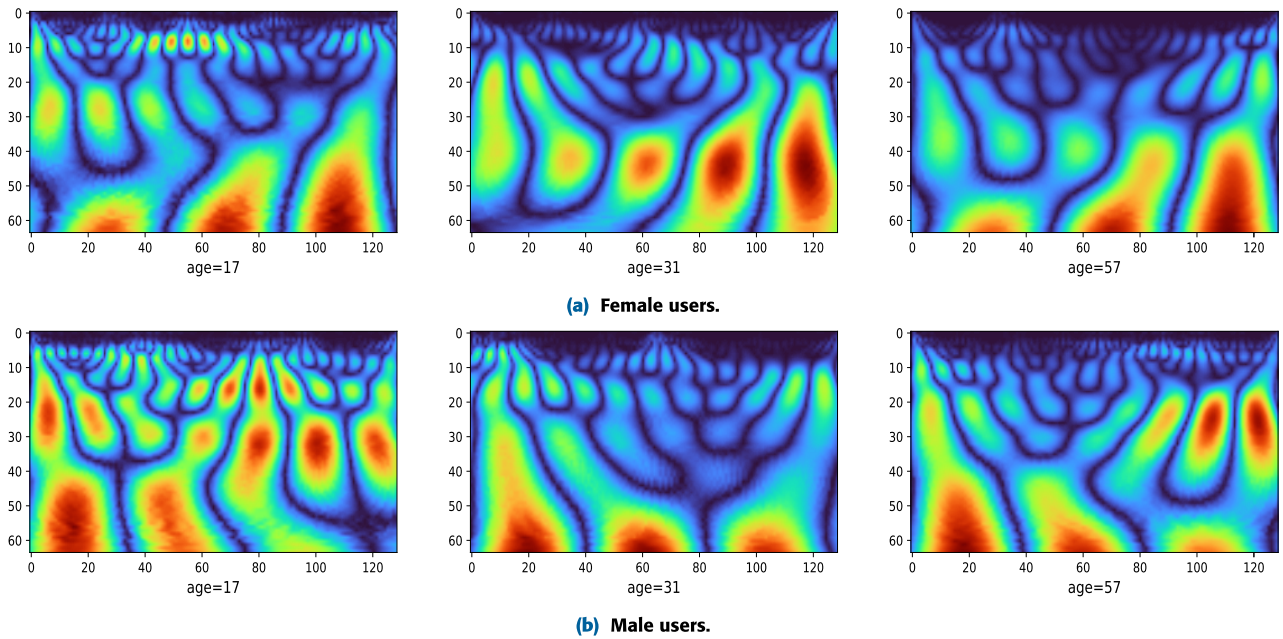
FIGURE 10. Example scalograms of female and male users with different ages.
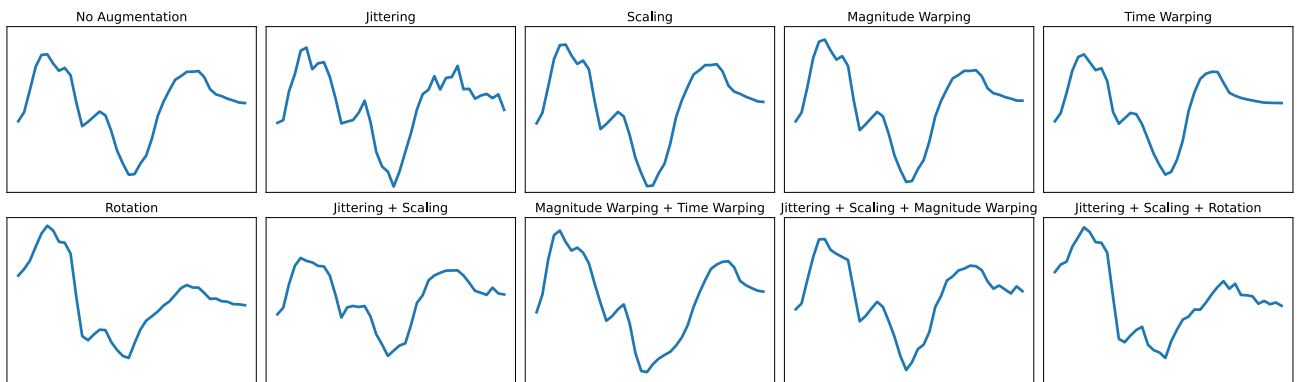

FIGURE 11. Examples of different traditional augmentation techniques.

Upon analyzing the results, we observe that the AC-GAN model outperforms the other two primary methods. The reasons for this superiority can be explained as follows. Traditional methods typically involve certain transformations to augment the training data without improving the data distribution determined by high-level features. Therefore, a technique is needed to estimate the data distribution and generate new data, not solely to augment the training set. Whereas SMOTE and its variations can be employed as oversampling techniques to produce new data samples close to existing instances in the feature space, they may not be as successful in high-dimensional datasets where extracting features may be challenging [61]. In this regard, the AC-GAN model excels at mimicking the distribution of the current training dataset, thereby offering better performance improvements.

When analyzing data distributions from different genders and age groups, we notice that some groups have fewer data samples than others. This imbalance could lead to insufficient

learning of the characteristics of age groups with limited data during the classification. To address this problem, we apply synthetic data generation to balance the number of data samples across different age groups as described in **Algorithm 1**. For this purpose, we divide the users into seven distinct age groups commonly used in the literature (age: <12, 12-17, 18-24, 25-34, 35-44, 45-54, >54). Then, we identify the age group with the most data and balance the data sizes of other age groups to match it. Fig. 12 and Fig. 13 display the original and generated synthetic data distributions for an example training dataset from BOUN Sensor and OU-ISIR Center IMUZ datasets, respectively.

For instance, the age range of the sample training dataset from the BOUN Sensor dataset in Fig. 12 is 18-57 and therefore includes five different age groups. This training set from the BOUN Sensor dataset initially contains 11.750 data samples, with the highest data belonging to the 25-34 age range female subclass, consisting of 2.097 samples. Subsequently, 9.220 synthetic data examples are generated

TABLE 3. The effects of different data generation techniques.

Data Generation Method	Accuracy (%)
No Augmentation	91.37
Jittering	91.43
Scaling	90.88
Magnitude Warping	87.06
Time Warping	86.70
Rotation	91.42
Jittering + Scaling	91.49
Magnitude Warping + Time Warping	86.38
Jittering + Scaling + Magnitude Warping	89.17
Jittering + Scaling + Rotation	91.87
SMOTE	90.68
SMOTESVM	91.14
ADASYN	91.45
AC-GAN	**92.33**
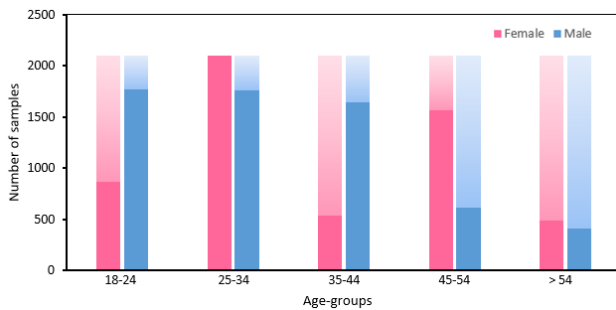
FIGURE 12. Balancing the data distribution for different age groups in the BOUN Sensor dataset. Solid bars represent the original data, whereas translucent bars represent the generated synthetic data.
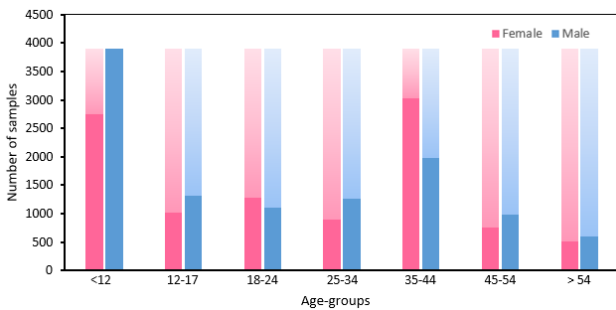
FIGURE 13. Balancing the data distribution for different age groups in the OU-ISIR Center IMUZ dataset. Solid bars represent the original data, whereas translucent bars represent the generated synthetic data.

to balance the data distribution among subclasses. Through a verification step, approximately 11.9% of the produced synthetic data is discarded, and new data is generated to replace them.

In addition, we employ a verification process on the generated synthetic samples. For verification, we utilize a pre-trained 2-Layer CNN model as a classifier to predict the class label of the generated samples. We then ignore the generated samples if the predicted class does not match the class assigned to the generated data. The impact of balancing age group distribution and using verification on the overall

TABLE 4. The effects of the additional steps in data generation.

Data Generation Step	Accuracy (%)
Generating synthetic data	92.33
Balancing age group distribution	93.17
Balancing age group distribution + Verification	**93.84**

TABLE 5. The effects of the different deep learning models in classification.

Deep Learning Model	Accuracy (%)
2-Layer CNN	93.84
CNN-LSTM	90.75
ConvLSTM	94.13
3-Layer CNN + LSTM	**94.83**

TABLE 6. The layers and parameters of 3-Layer CNN + LSTM architecture.

Layer Type	Output shape	# of Parameters
Conv2D	$64 \times 128 \times 64$	6464
MaxPooling	$32 \times 64 \times 64$	0
Conv2D	$32 \times 64 \times 64$	102464
MaxPooling	$16 \times 32 \times 64$	0
Conv2D	$16 \times 32 \times 32$	51232
MaxPooling	$8 \times 16 \times 32$	0
Flatten	4096	0
Reshape	1×4096	0
LSTM	64	1065216
Dropout	64	0
Fully-connected	1000	65000
Fully-connected	2	2002

performance is provided in Table 4. Balancing the age group distribution by representing all age groups equally in the dataset enhances the learning process during classification, resulting in improved performance. Furthermore, removing poor-quality synthetic samples through the verification process contributes to performance improvement by mitigating adverse effects on the training process.

In the classification, in addition to the 2-Layer CNN model, we explore other DNN models such as CNN-LSTM, ConvLSTM, and 3-Layer CNN + LSTM. Table 5 presents the performance results of these models. We observe that incorporating LSTM layers with CNN improves the performance. This enhancement can be attributed to LSTM's sequence prediction and temporal feature extraction capabilities. A typical gait encompasses sequential phases such as stance and swing, so utilizing LSTM enhances the learning procedure. Whereas the single-layer CNN and LSTM did not achieve satisfactory results, increasing the number of CNN layers improved the performance. The ConvLSTM architecture also outperformed the 2-Layer CNN model, but the highest success rate of 94.83% is achieved with the 3-Layer CNN + LSTM model. The layers and corresponding parameter numbers of the utilized 3-Layer CNN + LSTM model are provided in Table 6.

TABLE 7. Comparison of accuracy (%) results for gender detection frameworks using different datasets.

Data Generation	Feature Extraction	Classification	BOUN Sensor	OU-ISIR Android	OU-ISIR Center IMUZ
-	Time-domain features	SVC in [50]	83.33	65.93	67.18
-	Raw sensor data	1D CNN in [50]	88.33	75.21	78.73
-	CWT	2-Layer CNN	91.37	80.37	82.58
Generating synthetic data with AC-GAN	CWT	2-Layer CNN	92.33	82.72	84.23
Balancing age groups with AC-GAN	CWT	2-Layer CNN	93.17	83.82	85.63
Balancing age groups with AC-GAN + Verification	CWT	2-Layer CNN	93.84	84.80	86.16
Balancing age groups with AC-GAN + Verification	**CWT**	**3-Layer CNN + LSTM**	**94.83**	**86.27**	**88.15**

TABLE 8. Classification results for gender detection frameworks with all metrics.

Dataset	Model	Accuracy (%)	Precision Female (%)	Precision Male (%)	Recall Female (%)	Recall Male (%)	F1-score Female (%)	F1-score Male (%)
BOUN Sensor	SVC in [50]	83.33	81.75	85.09	85.53	80.83	83.74	82.91
	1D CNN in [50]	88.33	89.93	86.86	86.33	90.33	88.10	88.56
	Our Proposed Model	**94.83**	**97.69**	**92.15**	**91.67**	**97.83**	**94.58**	**94.91**
OU-ISIR Android	SVC in [50]	65.93	68.81	63.81	58.25	73.69	63.09	68.35
	1D CNN in [50]	75.21	74.73	75.72	76.19	74.23	75.45	74.97
	Our Proposed Model	**86.27**	**88.75**	**84.09**	**83.07**	**89.47**	**85.82**	**86.70**
OU-ISIR IMUZ	SVC in [50]	67.18	68.45	66.07	63.74	70.62	66.01	68.27
	1D CNN in [50]	78.73	78.71	78.74	78.75	78.70	78.73	78.72
	Our Proposed Model	**88.15**	**88.61**	**87.69**	**87.54**	**88.75**	**88.07**	**88.22**

TABLE 9. Classification results of gender detection frameworks for the age range 17-57.

Dataset	Model	Accuracy (%)	Precision Female (%)	Precision Male (%)	Recall Female (%)	Recall Male (%)	F1-score Female (%)	F1-score Male (%)
BOUN Sensor	SVC in [50]	83.33	81.75	85.09	85.83	80.83	83.74	82.91
	1D CNN in [50]	88.33	89.93	86.86	86.33	90.33	88.10	88.56
	Our Proposed Model	**94.83**	**97.69**	**92.15**	**91.67**	**97.83**	**94.58**	**94.91**
OU-ISIR Android	SVC in [50]	71.83	69.78	74.32	76.97	66.67	73.20	70.29
	1D CNN in [50]	80.97	78.20	84.32	85.86	76.06	81.85	79.98
	Our Proposed Model	**91.94**	**89.53**	**94.73**	**95.05**	**88.89**	**92.21**	**91.71**
OU-ISIR IMUZ	SVC in [50]	72.62	70.40	74.86	77.55	67.21	73.80	70.83
	1D CNN in [50]	84.71	80.09	90.97	92.35	77.05	85.79	83.43
	Our Proposed Model	**93.87**	**91.52**	**96.52**	**96.72**	**91.04**	**94.05**	**93.70**

When we evaluate the results obtained so far, we achieve a success rate of 94.83%, with a significant improvement of 6.5% with our proposed method, compared to the highest success rate of 88.33% achieved with the 1D CNN model in the study [50].

In addition, we test our model and its subcomponents on the OU-ISIR datasets. For comparison, we also implement support vector classifier (SVC) and 1D CNN approaches used in the study [50] for the OU-ISIR datasets. Table 7 summarizes the results for all three datasets. Upon analyzing the results, we observe that our proposed model performs successfully on the other two datasets. The steps such as generating synthetic data, applying CWT, and using the 3-Layer CNN + LSTM model for classification similarly improve the results on these two datasets.

Furthermore, Table 8 provides an overview of the results obtained from our final model, encompassing all relevant metrics. The table illustrates that our proposed model consistently performs well across all three datasets. Specifically, when examining recall, which signifies the correct classification of subjects within a positive class, our model exhibits slightly higher recall values for male users. This observation suggests our model makes slightly more accurate predictions for male users. Additionally, we achieve high F1-scores across all datasets, indicating the robust predictive capability of our model.

When comparing our proposed method to the 1D CNN model used in the work [50], we observe an improvement of approximately 11% on the OU-ISIR Android dataset and around 9.5% on the OU-ISIR Center IMUZ dataset. When comparing the results of these two datasets, we notice that the OU-ISIR Center IMUZ dataset, which includes both accelerometer and gyroscope sensor data, performs slightly better than the OU-ISIR Android dataset. On the other

TABLE 10. A comparison of the existing gender detection methods using the OU-ISIR Center IMUZ dataset.

Study	Method	Accuracy (%)	Main Remarks
Mostafa *et al.* [62]	Autocorrelation function as a feature extractor + CNN	70.4	(i) transfer learning and cross-testing are also employed and discussed
Gillani *et al.* [33]	Statistical features + Logistic Regression	68.2	(i) the best performance scores: 68.2% and 65% for different walking sequences (ii) a dataset is divided as 70% training set, 30% test set
Ahad *et al.* [16]	Features from gait dynamic image + temporal convolution network	75.8	(i) a separate test set not published in the OU-ISIR dataset is used for performance evaluation (ii) various deep learning-based solutions employed and the best prediction error of 24.23% is obtained using a temporal convolution network
Sun *et al.* [34]	Sliding window technique on raw sensor data + CNN	82.8	(i) 640 users selected out of 744 (ii) inter-subject monte-carlo validation is employed for performance evaluation
Khabir *et al.* [63]	Time-domain features + SVC	84.8	(i) both IMUZ Sensor and Android datasets are used (ii) datasets are divided as 85% training set, 15% test set
Pathan *et al.* [8]	Both time and frequency-domain features + bagging classifier approach	87.9	(i) CV is used for performance evaluation
Our Proposed Model	**Synthetic data generation + CWT as a feature extractor + 3-Layer CNN + LSTM hybrid model**	**88.1**	**(i) LOGOCV method is employed to ensure reliable performance evaluation**

hand, the relatively higher results obtained from the BOUN Sensor dataset can be attributed to its somewhat narrower age range and lower number of subjects compared to the OU-ISIR datasets. Since the OU-ISIR datasets cover a much more comprehensive age range, from 2 to 78, extracting distinguishing features for gender using our proposed method becomes more challenging.

At this point, we aim to eliminate the influence of different age groups and test our proposed model on the three datasets using the same age range. To achieve this, we remove users aged 0-16 and 58 years and older from the OU-ISIR datasets, aligning them with the age range of the BOUN Sensor dataset (17-57 years). Table 9 presents the results of this identical age range across all metrics. We achieve 91.94% and 93.87% accuracy on the OU-ISIR Android and OU-ISIR Center IMUZ datasets, respectively. Comparing these results to the previous ones, we observe an improvement in accuracy of approximately 6% for both datasets. These outcomes, which closely align with the 94.83% accuracy on the BOUN Sensor dataset, underscore the generalizability of our findings.

Additionally, we compare the performance of our proposed method with other works that utilized the OU-ISIR Center IMUZ dataset. Table 10 summarizes these exemplary studies, their methods, and their performances. We observe that deep learning models outperformed traditional ML methods when analyzing the results. For instance, in [33], using statistical features and LR classifier, they achieved an accuracy rate of 68.2%, consistent with our SVC results. An accuracy rate of 70.4% was obtained using the autocorrelation function and CNN in [62].

On the other hand, works [16] and [34] obtain higher accuracy rates by employing CNN. In [16], a completely different set of users is used as the test set, whereas in [34], the inter-subject monte-carlo validation method is applied. The inter-subject monte-carlo validation ensures that data from the same user are prevented from appearing in the

training and test sets, leading to a more reliable performance evaluation. In contrast, other works [8], [63] achieved high accuracy rates as well, but their evaluation methods may need to be more accurate due to the possibility of the same user's data appearing in both the training and test sets. Considering all these factors, our proposed method outperforms previous approaches applied to the same dataset.

As a result, we can summarize the analysis and findings of our experimental results as follows:

- Applying CWT as feature extractor and utilizing 2D CNN model as classifier yields superior performance compared to using raw sensor data and 1D CNN architecture. There are two main reasons for this observation. Firstly, CWT provides 2D output, capturing detailed information in both the time and frequency domains. Secondly, 2D DNN models exhibit higher learning capacity in classification tasks than their 1D counterparts.

- Generating synthetic data to expand the training dataset enhances overall performance by addressing overfitting problems and improving the generalization capabilities of classifier models. The AC-GAN model demonstrated superior performance among various techniques for synthetic data generation. This is attributed to its ability to mimic the existing dataset's distribution effectively.

- Additionally, we explore two additional steps further to improve the performance of the data generation process. Balancing the age group distribution ensures that all age groups are equally represented in the dataset, allowing for a more effective learning process during classification. Moreover, implementing the verification step enables removing poor-quality synthetic samples, thus fine-tuning the overall model performance.

- In classification, integrating LSTM with CNN and increasing the number of CNN layers in the hybrid model yields improved results. This can be attributed

to the sequence prediction capability of LSTM, which enhances the learning process.

- Finally, we evaluate our proposed method on three different public datasets using LOGOCV. The results show that our proposed method consistently outperforms the baseline models on all datasets. Additionally, our approach demonstrates superior performance compared to other studies that used the same datasets.

VI. CONCLUSION

In this paper, we demonstrate the possibility of gender detection based on analyzing motion sensor data from smart devices. Specifically, we propose a novel gender detection scheme based on DNN architecture, employing synthetic data generation and CWT. The proposed method analyzes the gait characteristics of users by processing sensor data, enabling accurate gender detection. The scheme comprises three main modules: *data generation*, *feature extraction*, and *gender detection*. In the *data generation* module, synthetic data are generated using various techniques to expand the existing training dataset. Subsequently, CWT is applied for extracting 2D feature matrices from time-series data in the *feature extraction* module. In classification, a hybrid DNN architecture combining CNN and LSTM layers is employed to accurately classify the gender of users as either female or male. The proposed method is evaluated on various public datasets and compared to similar works' performance. The experimental results show that our proposed model achieves high detection rates, outperforming the performance of previous methods.

Besides gender detection, this developed model can be applied in multiple domains relying on sensor data from smart devices, including activity recognition, age group estimation, user identification, and authentication. In future works, we will focus on applying this model to different purposes and deploying it for real-time detection on smart devices.

CONFLICT OF INTEREST

No potential conflict of interest is declared.

REFERENCES

[1] E. Davarci and E. Anarim, "Hybrid architecture for gender recognition using smartphone motion sensors," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 801–805.

[2] E. Garcia-Ceja, V. Osmani, and O. Mayora, "Automatic stress detection in working environments from smartphones' accelerometer data: A first step," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 4, pp. 1053–1060, Jul. 2016.

[3] D. Ilisei and D. M. Suciu, "Human-activity recognition with smartphone sensors," in *Proc. OTM Confederated Int. Conf. 'Move Meaningful Internet Syst.'*, Oct. 2020, pp. 179–188.

[4] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016.

[5] R.-A. Voicu, C. Dobre, L. Bajenaru, and R.-I. Ciobanu, "Human physical activity recognition using smartphone sensors," *Sensors*, vol. 19, no. 3, Jan. 2019, Art. no. 458.

[6] N. Al-Naffakh, N. Clarke, and F. Li, "Continuous user authentication using smartwatch motion sensor data," in *Proc. 12th Int. Conf. Trust Manage. (IFIPTM)*, Jul. 2018, pp. 15–28.

[7] E. Davarci and E. Anarim, "User identification on smartphones with motion sensors and touching behaviors," in *Proc. 30th Signal Process. Commun. Appl. Conf. (SIU)*, May 2022, pp. 1–4.

[8] R. K. Pathan, M. A. Uddin, N. Nahar, F. Ara, M. S. Hossain, and K. Andersson, "Gender classification from inertial sensor-based gait dataset," in *Proc. Int. Conf. Intell. Comput. Opt. (ICO)*, Dec. 2020, pp. 583–596.

[9] C. Prakash, R. Kumar, and N. Mittal, "Recent developments in human gait research: Parameters, approaches, applications, machine learning techniques, datasets and challenges," *Artif. Intell. Rev.*, vol. 49, no. 1, pp. 1–40, Jan. 2018.

[10] Y. Makihara, M. S. Nixon, and Y. Yagi, "Gait recognition: Databases, representations, and applications," in *Computer Vision*. Cham, Switzerland: Springer, 2020, pp. 1–13.

[11] D. Jarchi, J. Pope, T. K. M. Lee, L. Tamjidi, A. Mirzaei, and S. Sanei, "A review on accelerometry-based gait analysis and emerging clinical applications," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 177–194, 2018.

[12] C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang, "SilentSense: Silent user identification via touch and movement behavioral biometrics," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, Oct. 2013, pp. 187–190.

[13] S. A. Khan, M. Ahmad, M. Nazir, and N. Riaz, "A comparative analysis of gender classification techniques," *Int. J. Bio-Sci. Bio-Technol.*, vol. 5, no. 4, pp. 223–244, Aug. 2013.

[14] T. C. Ling, T. Alam, and A. A. Hussin, "Investigation of the role of individual's gender in the design of graphical user interface for mobile learning device," *J. Phys., Conf. Ser.*, vol. 1529, no. 3, Apr. 2020, Art. no. 032021.

[15] F. A. Batarseh and J. Pithadia, "Context-aware user interfaces for intelligent emergency applications," in *Proc. 10th Int. Conf. Model. Using Context*, Jun. 2017, pp. 359–369.

[16] M. A. R. Ahad, T. T. Ngo, A. D. Antar, M. Ahmed, T. Hossain, D. Muramatsu, Y. Makihara, S. Inoue, and Y. Yagi, "Wearable sensor-based gait analysis for age and gender estimation," *Sensors*, vol. 20, no. 8, Apr. 2020, Art. no. 2424.

[17] S. Z. S. Idrus, E. Cherrier, C. Rosenberger, S. Mondal, and P. Bours, "Keystroke dynamics performance enhancement with soft biometrics," in *Proc. IEEE Int. Conf. Identity, Secur. Behav. Anal. (ISBA)*, Mar. 2015, pp. 1–7.

[18] T. Chai, S. Prasad, and S. Wang, "Boosting palmprint identification with gender information using DeepNet," *Future Gener. Comput. Syst.*, vol. 99, pp. 41–53, Oct. 2019.

[19] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019.

[20] A. Bhardwaj, W. Di, and J. Wei, *Deep Learning Essentials: Your Hands-On Guide to the Fundamentals of Deep Learning and Neural Network Modeling*. Birmingham, U.K.: Packt Publishing, 2018.

[21] M.-J. Chung and M.-J.-J. Wang, "The change of gait parameters during walking at different percentage of preferred walking speed for healthy adults aged 20–60 years," *Gait Posture*, vol. 31, no. 1, pp. 131–135, Jan. 2010.

[22] P. Scaglioni-Solano and L. F. Aragón-Vargas, "Gait characteristics and sensory abilities of older adults are modulated by gender," *Gait Posture*, vol. 42, no. 1, pp. 54–59, Jun. 2015.

[23] B. Ayotte, M. K. Banavar, D. Hou, and S. Schuckers, "Group leakage overestimates performance: A case study in keystroke dynamics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 1410–1417.

[24] A. Agneessens, I. Bisio, F. Lavagetto, and M. Marchese, "Design and implementation of smartphone applications for speaker count and gender recognition," in *The Internet of Things*. New York, NY, USA: Springer, 2010, pp. 187–194.

[25] Z. Qawaqneh, A. A. Mallouh, and B. D. Barkana, "Age and gender classification from speech and face images by jointly fine-tuned deep neural networks," *Expert Syst. Appl.*, vol. 85, pp. 76–86, Nov. 2017.

[26] K. Zhang, N. Liu, X. Yuan, X. Guo, C. Gao, and Z. Zhao, "Fine-grained age group classification in the wild," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 788–793.

[27] K. Zhang, C. Gao, L. Guo, M. Sun, X. Yuan, T. X. Han, Z. Zhao, and B. Li, "Age group and gender estimation in the wild with deep RoR architecture," *IEEE Access*, vol. 5, pp. 22492–22503, 2017.

[28] P. Barra, C. Bisogni, M. Nappi, D. Freire-Obregón, and M. Castrillón-Santana, "Gait analysis for gender classification in forensics," in *Proc. 5th Int. Conf. Depend. Sens. Cloud Big Data Syst. Appl. (DependSys)*, Nov. 2019, pp. 180–190.

[29] Q. Riaz, M. Z. U. H. Hashmi, M. A. Hashmi, M. Shahzad, H. Errami, and A. Weber, "Move your body: Age estimation based on chest movement during normal walk," *IEEE Access*, vol. 7, pp. 28510–28524, 2019.

[30] A. T. Sabir, H. S. Maghdid, S. M. Asaad, M. H. Ahmed, and A. T. Asaad, "Gait-based gender classification using smartphone accelerometer sensor," in *Proc. 5th Int. Conf. Frontiers Signal Process. (ICFSP)*, Sep. 2019, pp. 12–20.

[31] M. De Marsico and A. Mecca, "Gait recognition: The wearable solution," in *Human Recognition in Unconstrained Environments*. New York, NY, USA: Academic Press, 2017, pp. 177–195.

[32] A. Jain and V. Kanhangad, "Investigating gender recognition in smartphones using accelerometer and gyroscope sensor readings," in *Proc. Int. Conf. Comput. Techn. Inf. Commun. Technol. (ICCTICT)*, Mar. 2016, pp. 597–602.

[33] S. I. Gillani, M. A. Azam, and M. Ehatisham-ul-Haq, "Age estimation and gender classification based on human gait analysis," in *Proc. Int. Conf. Emerg. Trends Smart Technol. (ICETST)*, Mar. 2020, pp. 1–6.

[34] Y. Sun, F. P.-W. Lo, and B. Lo, "A deep learning approach on gender and age recognition using a single inertial sensor," in *Proc. IEEE 16th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, May 2019, pp. 1–4.

[35] Q. Riaz, A. Vögele, B. Krüger, and A. Weber, "One small step for a man: Estimation of gender, age and height from recordings of one step by a single inertial sensor," *Sensors*, vol. 15, no. 12, pp. 31999–32019, Dec. 2015.

[36] A. Sharshar, A. Fayez, Y. Ashraf, and W. Gomaa, "Activity with gender recognition using accelerometer and gyroscope," in *Proc. 15th Int. Conf. Ubiquitous Inf. Manage. Commun. (IMCOM)*, Jan. 2021, pp. 1–8.

[37] T. Meena and K. Sarawadekar, "Gender recognition using in-built inertial sensors of smartphone," in *Proc. IEEE Region 10 Conf. (TENCON)*, Nov. 2020, pp. 462–467.

[38] R. Schwesig, S. Leuchte, D. Fischer, R. Ullmann, and A. Kluttig, "Inertial sensor based reference gait data for healthy subjects," *Gait Posture*, vol. 33, no. 4, pp. 673–678, Apr. 2011.

[39] A. Nedorubova, A. Kadyrova, and A. Khlyupin, "Human activity recognition using continuous wavelet transform and convolutional neural networks," 2021, *arXiv:2106.12666*.

[40] G. Fenza, M. Gallo, V. Loia, F. Orciuoli, and E. Herrera-Viedma, "Data set quality in machine learning: Consistency measure based on group decision making," *Appl. Soft Comput.*, vol. 106, Jul. 2021, Art. no. 107366.

[41] G. Iglesias, E. Talavera, Á. González-Prieto, A. Mozo, and S. Gómez-Canaval, "Data augmentation techniques in time series domain: A survey and taxonomy," *Neural Comput. Appl.*, vol. 35, no. 14, pp. 10123–10145, May 2023.

[42] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[43] H. M. Nguyen, E. W. Cooper, and K. Kamei, "Borderline over-sampling for imbalanced data classification," *Int. J. Knowl. Eng. Soft Data Paradigms*, vol. 3, no. 1, pp. 4–21, Apr. 2011.

[44] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 1322–1328.

[45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020.

[46] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2017, pp. 2642–2651.

[47] R. Bracewell and P. B. Kahn, "The Fourier transform and its applications," *Amer. J. Phys.*, vol. 34, no. 8, p. 712, Aug. 1966.

[48] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990.

[49] W. K. Ngui, M. S. Leong, L. M. Hee, and A. M. Abdelrhman, "Wavelet analysis: Mother wavelet selection methods," *Appl. Mech. Mater.*, vol. 393, pp. 953–958, Sep. 2013.

[50] E. Davarci and E. Anarim, "Gender detection with smartphone motion sensors using convolutional neural networks," in *Proc. 29th Signal Process. Commun. Appl. Conf. (SIU)*, Jun. 2021, pp. 1–4.

[51] A. Baldominos, A. Cervantes, Y. Saez, and P. Isasi, "A comparison of machine learning and deep learning techniques for activity recognition using mobile devices," *Sensors*, vol. 19, no. 3, Jan. 2019, Art. no. 521.

[52] X. Xiaolong, C. Wen, and S. Yanfei, "Over-sampling algorithm for imbalanced data classification," *J. Syst. Eng. Electron.*, vol. 30, no. 6, pp. 1182–1191, Dec. 2019.

[53] Y. Wang and P. He, "Comparisons between fast algorithms for the continuous wavelet transform and applications in cosmology: The one-dimensional case," 2023, *arXiv:2302.03909*.

[54] E. Tsironi, P. Barros, C. Weber, and S. Wermter, "An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition," *Neurocomputing*, vol. 268, pp. 76–86, Dec. 2017.

[55] E. Davarci and E. Anarım, "Bogazici university smartphone accelerometer sensor dataset," *Data Brief*, vol. 40, Feb. 2022, Art. no. 107833.

[56] 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.

[57] C. T. Cornelius and D. F. Kotz, "Recognizing whether sensors are on the same body," *Pervas. Mobile Comput.*, vol. 8, no. 6, pp. 822–836, Dec. 2012.

[58] M. Antal and G. Nemes, "Gender recognition from mobile biometric data," in *Proc. IEEE 11th Int. Symp. Appl. Comput. Intell. Informat. (SACI)*, May 2016, pp. 243–248.

[59] A. Dehghani, T. Glatard, and E. Shihab, "Subject cross validation in human activity recognition," 2019, *arXiv:1904.02666*.

[60] C.-Y. Yam, M. S. Nixon, and J. N. Carter, "Extended model-based automatic gait recognition of walking and running," in *Proc. 3rd Int. Conf. Audio Video-Based Biometric Person Authn. (AVBPA)*, Jun. 2001, pp. 278–283.

[61] S. Suh, "Improving classification performance under imbalanced data conditions using generative adversarial networks," Ph.D. dissertation, Dept. Comput. Sci., Technische Universität Kaiserslautern, Kaiserslautern, Germany, 2021.

[62] A. Mostafa, S. Elsagheer, and W. Gomaa, "BioDeep: A deep learning system for IMU-based human biometrics recognition," in *Proc. 18th Int. Conf. Informat. Control, Autom. Robot.*, 2021, pp. 620–629.

[63] K. M. Khabir, Md. S. Siraj, M. Ahmed, and M. U. Ahmed, "Prediction of gender and age from inertial sensor-based gait dataset," in *Proc. Joint 8th Int. Conf. Informat., Electron. Vis. (ICIEV), 3rd Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR)*, May 2019, pp. 371–376.

ERHAN DAVARCI received the B.Sc. and M.Sc. degrees in electrical and electronics engineering from Boğaziçi University, Istanbul, Turkey, in 2012 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Electronics Engineering.

From 2011 to 2020, he was a Network Engineer with the Network Operation and Service Delivery Department, Turkcell, Turkey. He is a Principal Service Automation Developer with Turkcell. His research interests include machine learning, sensors, user biometrics, and smart devices.

EMIN ANARIM received the B.Sc. (Hons.), M.Sc., and Ph.D. degrees in electrical and electronics engineering from Boğaziçi University, Istanbul, Turkey, in 1981, 1983, and 1985, respectively.

He is currently a Professor with the Department of Electrical and Electronics Engineering, Boğaziçi University. He is also an Adjunct Professor with George Washington University. Throughout his career, he has taught industry and government institutions several courses on signal processing, video coding, secure telecommunications, mobile communications, lawful interception techniques in telecommunications, cryptographic techniques, and network security. He has frequently acted as a consultant for various telecommunications companies in Turkey. He has published more than 250 papers carrying his name, appearing in multiple journals and conference proceedings. His current research interests include multimedia wireless communications, information, and network security.

• • •