Journal of Informatics and
Web Engineering

Vol. 2 No. 2 (September 2023) eISSN: 2821-370X

GenReGait: Gender Recognition using Gait Features

Yue Fong Ti[1], Tee Connie[1*], Michael Kah Ong Goh[1]

[1]Faculty of Information Science and Technology, Multimedia University, Malaysia

**corresponding author: (tee.connie@mmu.edu.my, ORCiD: 0000-0002-0901-3831)*

Abstract - Gender recognition based on gait features has gained significant interest due to its wide range of applications in various fields. This paper proposes GenReGait, a robust method for gender recognition utilizing gait features. Gait, the unique walking pattern of individuals, contains distinct gender-specific characteristics, such as stride length, step frequency, and body posture, making it a promising modality for gender estimation. The proposed GenReGait method begins by extracting landmark positions on the human body using a human keypoint estimation technique. These landmarks serve as informative cues for estimating gender based on their spatial and temporal characteristics. However, environmental factors can impact gait patterns and introduce fluctuations in landmark points, affecting the accuracy of gender estimation. To overcome this challenge, GenReGait introduces a robust preprocessing technique known as Weighted Exponential Moving Average to smoothen the gait signals and reduce noise caused by environmental factors. The smoothed signals are then fed into a deep learning network trained to perform gender estimation based on the gait features extracted from the landmark positions. By leveraging deep learning algorithms, the proposed GenReGait method effectively captures complex patterns and relationships within the gait features, enhancing the accuracy and reliability of gender recognition. Experimental evaluations conducted on the Gait in the Wild dataset and a self-collected dataset validate the robustness and effectiveness of the proposed GenReGait approach.

Keywords— gender recognition, gait recognition, human pose estimation, machine learning, deep learning, computer vision

Received: 27 June 2023; Accepted: 09 August 2023; Published: 16 September 2023

I. INTRODUCTION

Gender recognition, the process of identifying an individual's gender based on their characteristics, has become an area of significant interest due to its wide range of applications. The ability to accurately estimate gender can assist businesses in tailoring their products and services to target customers, leading to improved marketing strategies and enhanced customer satisfaction. Moreover, gender recognition systems have crucial implications in the field of surveillance, enabling the restriction of access to certain areas or facilities based on the opposite sex to ensure privacy and security.

Human gait, the unique manner in which individuals walk, has emerged as a promising source of information for gender estimation. Gait features encompass the distinct patterns and movements observed during walking, such as stride length, step frequency, and body posture. These features exhibit inherent gender-specific characteristics that can be leveraged to estimate an individual's gender. For example, women typically exhibit greater pelvic rotation during walking compared to men. This is partly due to the wider pelvic structure and is associated with maintaining balance

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2023.2.2.10

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

and stability. Besides, women often have a narrower foot placement during walking, with the feet closer together. This can be attributed to the wider hips and differences in body mechanics.

In this study, we propose a robust method, called GenReGait, for gender recognition using gait features. We begin by employing the human keypoint estimation technique to extract landmark positions on the human body. These landmarks serve as informative cues for estimating gender based on their spatial and temporal characteristics. However, accurate gender estimation can be challenging due to fluctuations in landmark points caused by environmental factors affecting an individual's walking pattern.

To address this challenge, we propose a robust preprocessing technique coined as Weighted Exponential Moving Average to smooth the gait signals and reduce the noise caused by environmental factors. The smoothed signals are then fed into a deep learning network, which is trained to perform gender estimation based on the gait features extracted from the landmark positions. By utilizing deep learning algorithms [1], the proposed method effectively captures complex patterns and relationships within the gait features, resulting in enhanced accuracy and reliability of gender recognition.

To evaluate the performance of our proposed approach, we conduct experiments on two datasets: the Gait in the Wild dataset and a self-collected dataset specifically designed for this study. The experimental results validate the robustness of our approach in accurately performing gender recognition. The contribution of this paper is the introduction of the GenReGait method, which combines landmark-based gait feature extraction method with a robust preprocessing technique. This method addresses the challenge of environmental factors affecting gait patterns and achieves enhanced accuracy and reliability in gender estimation.

II. LITERATURE REVIEW

A. Conventional Methods

Lee and Grimson [2] proposed a study of using simple method to do gender classification using gait features. They were using moments extracted from orthogonal view vide silhouettes of human walking motion which is enough to perform well in gender classification. They apply gait average appearance features to the task of gender classification. They show the best 6 features from 57 features based on the p-value of ANOVA for gender classification. After that, they used SVM to train and test the features under two conditions, random person, and random sequence. The random sequence condition can reach to highest 94% accuracy using linear kernel compared to gaussian and polynomial kernels. However, the appearance of a walking subject can affect the performance of result. Face recognition can be included as a feature with gait because face is less vulnerable to appearance alteration.

Makihara et al. [3] used large-scale multi-view gait database for gender and age classification. They process frequency-domain feature extraction comprises to do gait features extraction. After that, they analyzed differences between adult males and females. Males have wider shoulders and females have more rounder bodies from frontal. Differences of breast and hair area can be observed from side view. The walking stride of a male seem narrower than female as males is taller on average. They used KNN algorithm with parameter set to $k = 3$ to do the experiment. The result of classification shows highest 95.97% for adult males and adult female. From the experiment, age can be a factor because the gait of subjects will change with their age.

Wang and Yu [4] used KNN classifier for gender classification. They used IPIP laboratory's gait database which contains equal gender ratio and large number of pedestrians to make sure the result is accurate. Given the different angle of the gait data, it is easier to differentiate between different genders based on gait. They used Locally Linear Embedding (LLE) algorithm which is proposed by Sam T. Roweis and Lawrence K. Saul to do the gait features extraction. The algorithm is a fresh approach to dimension reduction for nonlinear data that preserves the original topology of the data after reduction. The information they extracted is spatial and temporal feature for increasing the gait recognition system's stability. They classify the data with temporal GPCI feature, spatial GPCI feature and both. The highest classification rate is 98.25% with using both features.

Miyamoto and Aoki [5] proposed a study for gender prediction based on the time series variation in the joint positions during walking. They used the modified program named "Skeleton Basics-D2D" provided by Microsoft. They used Kinect v2 device to get the coordinates of joint positions. During feature extraction, they used time series variation in the joint positions and two-dimensional coordinates on a two-dimensional plane corresponding to the angle. The classification results showed that the accuracy is 99.12% when feature extraction using three-dimensional coordinates directly and 99.12% if two-dimensional features are used in the best case.

A study performed by Guffanti, Brunete and Hernando [6] demonstrate the viability of non-invasive and cost-effective depth camera-based systems in capturing relevant human gait patterns. The primary objective was to classify gait by gender using these depth cameras. A total of 81 participants (40 female and 41 male) walked across a 4.8-meter walkway while gait data was recorded by multiple depth sensors. The analysis involved examining various gait features in both the time and spectral domains, such as joint excursions, range of movement, spatial variables, center of mass (COM) position, principal frequency, magnitude, and phase shift during walking. Significant gender differences in these features were used to train a support vector machine (SVM) classifier, achieving an impressive 96.7% accuracy.

Besides, Hema et al. [7] proposed a new descriptor called Gait energy image projection model (GPM) as the first step for gender perception. In the second step, the proposed descriptors were combined with existing ones like GEI, Gait Entropy Images, Dynamic Gait Energy Image (D-GEI), and Gradient Gait Energy Image (GGEI) to improve performance. The SVM classifier was employed for classification, distinguishing between genders. A summary of the conventional methods is provided in Table 1.

Table 1. A Summary of Conventional Methods for Gender Recognition.

Author	Model	Method	Recognition Rate	Pros
Lee and Grimson [2]	Support vector machine (SVM)	Used moments extracted from orthogonal view vide silhouettes Select feature using ANOVA	94%	Can extend to accommodate multiple appearance model
Makihara, Mannami, Hidetoshi & Yasushi [3]	K-nearest neighbors algorithm (k-NN)	Frequency-domain feature extraction comprises	95.97%	Get the information clearly in different age group
Wang and Yu [4]	K-nearest neighbors algorithm (k-NN)	Locally Linear Embedding (LLE) for feature extraction	98.25%	Get information from different angle
Miyamoto & Aoki [5]	Linear Support Vector Machine	Feature extraction using the time series variation in the joint positions	99.12%	High accuracy when 3D and 2D features are used
Guffanti, Brunete and Hernando [6]	Support vector machine (SVM)	Gait features acquired using depth camera	96.7%	Center of mass position shown to provide great discrimination between male and female
Hema et al. [7]	Support vector machine (SVM)	Gait energy image projection model	98%	Fusion of different descriptors yields good accuracy

B. Deep Learning Methods

Ismail et al. [8] used artificial neural network classifier to do gender recognition using gait analysis when the human gait data model-based gender classification is still in infancy. The dataset they used contains 20 subjects of female and male from CASIA gait database. Analysis of variance (ANOVA) and multiple comparison (MCP) will be used as feature selection to assess the efficacy of feature extracts. The statistical significance between a set of independent variables is determined by using ANOVA and identification of difference means between another within a set of means is determined by using MCP. After that, they input different feature groups to ANN and verification. Scaled conjugate gradient (SCG) and Levenberg-Marquardt (LM) algorithms are used to perform weight adjustment in NN layer. SCG is used as the training algorithm for ANN to achieve the best performance with 95% accuracy for male and 90% for female in feature group 2 which contain 20 optimized features.

Wazzeh et al. [9] applied convolutional neural network to do gender detection. Due to the efficiency in computational cost and storage, they decided to use a light version of CNN. They applied a maxout activation call max-feature-map into each CNN layer. They collect the dataset by capturing humans wearing black jumpsuit with markers taped on. In the experiment, they implement light CNN framework with MFM units as network prototype. This technique is appropriate to cast off noisy labels and increase detection rate. The performance shows 92.7% gender detection rate. In this study, female subjects were observed that have some unique movements which enable to discover them compare with male subjects and potentially can be new features of female to show the better results. They also found out that extract features of invariant to variations in walking speed will be more efficient in gender detection.

Besides, Nithyakani et al. [10] had proposed another deep convolutional neural network system by training architecture of neural network with gait energy image. They used TUM gait database which contains 1000 subjects with different views and general condition such as shoes, clothing etc. They proceed silhouette extraction and feature extraction. Silhouette extraction is to prevent the problems faced due to different colour and texture of apparels. After that, they train the data with CNN architectures. Deep convolutional neural network (DCNN) showed 94.7% average accuracy and CNN showed 90.6% average accuracy. DCNN has demonstrated the higher accuracy of gait recognition against the trained data samples through this experiment comparing it to convolutional neural network.

Apart from that, Thomas and Pushpalatha [11] introduces a gender recognition system based on deep learning and texture analysis. It focuses on utilizing the gait energy image, generated by incorporating silhouettes from a complete gait cycle, as a prominent feature for gait-based classification. In addition to the gait energy image, the study explores additional texture features like histogram of oriented gradient (HOG) and entropy for gender identification. The research compares the accuracy of gender classification using whole body images, upper body images, and lower body images. The findings demonstrate that combining multiple texture features improves accuracy compared to analyzing each feature individually. Moreover, the study reveals that full body gait images yield more precise results compared to partial body gait images.

In addition, a deep learning approach was introduced by Deng et al. [12] to divide the GEI based on the human structure model and motion information. The GEI was segmented into head, torso, upper limbs, and lower limbs, and then CNN features are extracted from each part and fused together for gait analysis, with normalization using spatial pyramid pooling. Experimental results on the CASIA B dataset demonstrated that combining the fused CNN features from separated GEI parts yielded better performance compared to using CNN alone or a single GEI. Table 2 summarizes the deep learning methods for gender recognition.

It can be observed that the results reported for some conventional machine learning methods are indeed higher than that of deep learning approaches. This is affected by certain factors which include the type of datasets used and also the sample size involved. Some conventional methods reported higher results as a dataset with simpler setting/configuration is used. Besides, the method will also perform better if a smaller number of samples were used.

In addition, the number of samples selected also affect the accuracy of the methods. In some situations, when dealing with a smaller number of samples, conventional methods may exhibit better performance due to their ability to handle limited data more effectively. Deep learning approaches, on the other hand, may require more extensive datasets to fully harness their potential, and performance gains might become evident with larger and more diverse datasets.

Table 2. A Summary of Deep Learning Methods for Gender Recognition.

Author	Model	Method	Recognition Rate	Pros
Ismail, Tahir and Hussain [8]	Artificial Neural Networks (ANN)	Select feature using Analysis of variance (ANOVA) and multiple comparison (MCP)	95%	Effective with statistical analysis as feature selection
Wazzeh, Birdal and Sertbaş [9]	Light Convolutional Neural Networks (CNN)	Use max-feature-map into CNN layer	92.7%	Save storage spaces and cost
Nithyakani, Shanthini and Ponsam [10]	Deep Convolutional Neural Network (DCNN)	Procced silhouette extraction and feature extraction	94.7%	Designed for convolution, pooling and normalization
Thomas and Pushpalatha [11]	Convolutional Neural Networks (CNN)	Gait energy image and HOG	97.4%	Texture features combination yielded optimal performance
Deng et al. [12]	Fused CNN	Segmented GEI and fusion using spatial pyramid pooling	95%	Segmentation of GEI feature can reduce impact of view changes

III. PROPOSED METHOD: GENREGAIT

In the proposed method, walking patterns of the subjects are acquired from CCTV footages. The body keypoint positions are estimated from the walking subjects using human pose tracking. To enhance the accuracy of the gait signals, a pre-processing step is employed to smooth out any fluctuations in the pattern. The smoothed signals are subsequently input into a deep learning network, which performs feature extraction and classification. This process ultimately determines whether the subject is male or female. A block diagram of the proposed method is depicted in Figure 1.

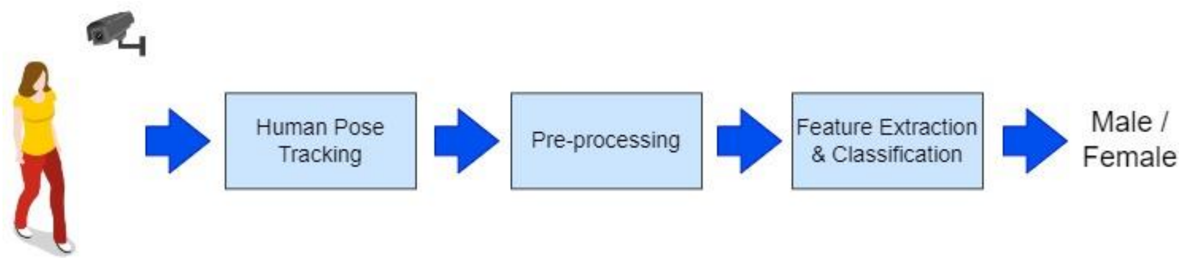

Figure 1. Block diagram of the proposed method.

A. Human Keypoints Tracking

Firstly, the landmark keypoints of the human body are estimated using AlphaPose [13]. AlphaPose is a state-of-the-art method for human pose estimation that leverages a deep learning framework, specifically a convolutional neural network (CNN), to accurately predict the positions of human keypoints in images or videos.

Let X be the input data, and $F(X)$ represent the learned features as in Equation (1):

$$F(X) = \varphi(X) \tag{1}$$

where φ refers to the transformation of the input data X through the layers of the CNN. The extracted features $F(X)$ are then used to predict the coordinates or heatmaps representing the estimated keypoints. These heatmaps, denoted as H, indicate the likelihood of each keypoint's presence and location in the image. The predicted keypoints, P, are obtained by identifying the peaks or maximum values in the heatmaps, as in Equation (2):

$$P = Argmax(H) \tag{2}$$

The overall output of AlphaPose is given by $O = \{(x_i, y_i, c_i)\}$ for $i = 1, \dots, N$, where $P = \{(x_i, y_i)\}$ and the associated confidence scores c_i, while N signifies the number of keypoints. In this study, a total of 17 keypoints are derived. Figure 2 illustrates a sample of keypoints estimated using AlphaPose.

Figure 2. Sample output from human keypoints estimation.

B. Pre-processing using Weighted Exponential Moving Average

Exponential Moving Average (EMA) is a statistical method used to analyse time series data. Unlike the Simple Moving Average (SMA), which gives equal weight to all data points within a specified period, the EMA assigns exponentially decreasing weights to the data points, as in Equation (3);

$$EMA(t) = (\alpha * x(t)) + \big((1 - \alpha) * EMA(t - 1)\big) \tag{3}$$

where $EMA(t)$ is the EMA value at time *t*, $x(t)$ is the current data point at time *t*, $EMA(t - 1)$ is the EMA value at the previous time step $(t - 1)$. The smoothing factor, α, is calculated using the formula, $\alpha = 2/(N + 1)$, where N is the number of periods or time steps.

In certain scenarios, specific body positions, such as the upper body torso and arms, may not significantly affect gait signatures compared to positions on the legs and knees. Therefore, it is beneficial to assign higher weights to the more important body parts and lower weights to the less significant parts. Hence we propose a weighted EMA (WEMA) approach to assign different weights on the different body parts. The formula for WEMA is given in Equation (4);

$$WEMA(t) = (\alpha * w(t) * x(t)) + \big((1 - \alpha) * w(t) * WEMA(t - 1)\big) \tag{4}$$

where $w(t)$ is the weight assigned to the current data point at time *t*. The weights are assigned in the range of 0 to 1 for different body parts based on domain knowledge. For example, if certain body parts, like legs and knees, are considered more critical in gait analysis, a higher weight might be assigned to them (closer to 1). Conversely, less significant body parts, like the upper body torso and arms, might be given lower weights (closer to 0). This way, the WEMA approach would give more importance to the more relevant body parts in the gait analysis process.

The smoothing factor, α, determines the rate at which weights decrease, and it is a value between 0 and 1. Adjusting the smoothing factor allows control over the responsiveness of the WEMA to the different importance of the keypoint positions on the human body. The weighted approach captures the gait signatures more accurately by giving appropriate importance to different body parts.

C. Feature Extraction and Classification

In this study, two types of deep learning architectures, namely bi-directional long-short term memory (BiLSTM) and Gated Recurrent Unit (GRU), are investigated for feature extraction and classification. The reason to evaluate the two architectures is because we want to compare which model performs better for the gender recognition task. GRU is well-known for having a lightweight architecture and its computational efficiency. On the other hand, BiLSTM has the ability to process input sequences in both forward and backward directions simultaneously, which can potentially lead to improved performance. Therefore, we aim to assess a deep learning model that is suitable in accurately recognizing and classifying gender-related features.

Bi-Directional Long-Short Term Memory (BiLSTM)

BiLSTM is a type of recurrent neural network that allows any neural network to store sequence information in both directions, forward and backward [14]. The forward LSTM processes the sequence from start to end, while the backward LSTM processes it in reverse. The input runs in two directions in bidirectional, distinguishing a BiLSTM from a conventional LSTM. Using bi-directional, it can make the input flow in both directions, preserving both the future and the past. An illustration of BiLSTM is provided in Figure 3. The variables *x* and *y* in the figure represent the input and output of the network, respectively.

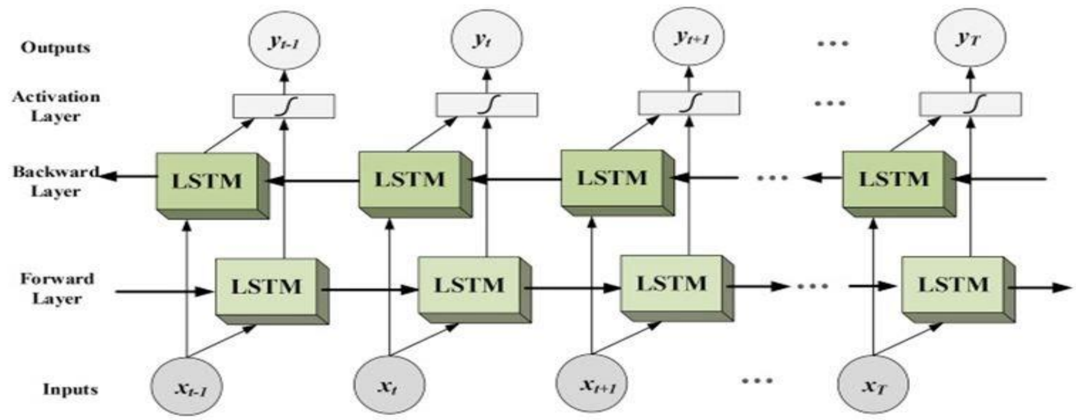

Figure 3. BiLSTM architecture [14].

Gated Recurrent Unit (GRU)

GRU is an improved version of the RNN architecture that addresses the limitations of traditional RNNs by introducing update and reset gates that control the flow of information [15]. Figure 4 presents the architecture of a GRU. The update gate, *z*, determines how much of the previous hidden state to retain and how much new information to add, while the reset gate, *r*, decides how much of the previous hidden state, h_{t-1}, to forget. The current hidden state, h_t, is computed by blending the previous hidden state and the current input using the gates. GRU can capture both short-term and long-term dependencies effectively, making it useful for tasks involving sequential data.

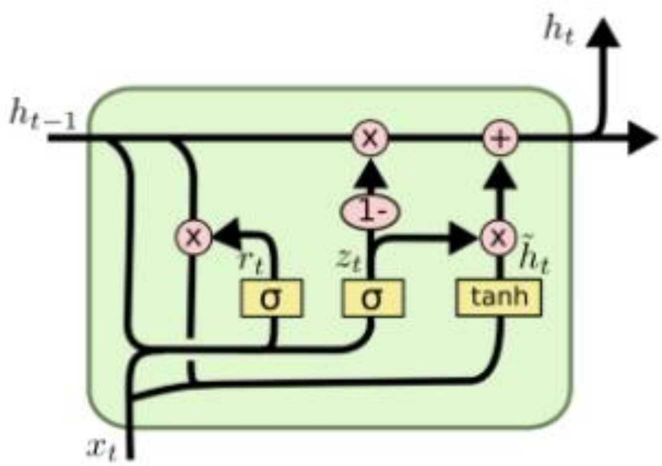

Figure 4. GRU architecture [15].

IV. EXPERIMENTAL RESULTS

A. Dataset

Gait in the Wild Dataset

The Gait in the Wild dataset was created by curating videos from various sources on the Internet. This dataset consists of 470 videos, with 227 videos featuring females and 243 videos featuring males. To ensure comprehensive gait capture, the dataset includes pedestrian movement from four different viewpoints: front, back, left, and right.

AlphaPose was used to obtain thirty body keypoints from each video. However, the keypoints for the head and face were excluded, as they do not significantly contribute to the analysis of a person's gait. The keypoints are stored in a

JSON file where each row corresponds to a frame in a video, and each column represents a specific body keypoint at a particular frame in the video. There is a total of 193,851 rows in the dataset, where 93,651 entries are for females and 100,200 entries for males. The label for each video is represented as 0 for male and 1 for female.

<u>Self-Collected Dataset</u>

Apart from collecting a set of videos from the Internet sources, a Self-Collected Dataset is also established in this study. The Self-Collected Dataset was meticulously constructed following specific filming instructions. A total of 2, 583 subjects, aged between 1 to 75 years old, were invited to participate in the study. They were asked to record a video clip of about one to two minutes containing walking sequences in four different directions: front, back, left and right. Each video was then manually segmented into 4 video segments corresponding to the different directions.

This dataset combines each frame into video-based data. Each row in the Self-Collected Dataset represents a video sequence. The keypoints data is recorded in an AlphaPose JSON file, excluding head and face keypoints. The dataset consists of 2,583 rows of videos. Additionally, each row includes participant information such as filename, age group, and gender. The dataset contains 1,325 males (represented by '0' in the 'gender' column) and 1,258 females (represented by '1' in the 'gender' column). An extract from the Self-Collected Dataset is depicted in Figure 5.

	filename	age_group	gender	keypoints_1	keypoints_2	keypoints_3	keypoints_4	keypoints_5	keypoints_6	keypoints_7	...	keypo
0	Subject_106_F_55_Clip_1_Back	senior	0	956.356567	174.383011	925.851257	144.540390	973.235840	153.011185	897.877441	...	95
1	Subject_106_F_55_Clip_1_Front	senior	0	855.333984	180.870255	866.631226	170.228287	844.578369	169.403976	880.075012	...	87
2	Subject_106_F_55_Clip_1_Left	senior	0	1498.788086	293.602631	1513.815308	279.522186	1498.457886	277.839813	1558.705688	...	14
3	Subject_106_F_55_Clip_1_Right	senior	0	114.032707	257.070465	121.089012	243.775009	95.683739	246.745834	132.296982	...	10
4	Subject_107_M_22_Clip_1_Back	adult	1	997.917358	172.884506	989.499390	158.549652	988.800964	154.812180	879.492554	...	97
...	...	...	...	...	...	...	...	...	...	...	...	
2578	Subject_9_M_34_Clip_1_Right	adult	1	203.681503	62.524403	204.311066	56.249832	200.503738	57.605007	185.984650	...	22
2579	Subject_9_M_34_Clip_2_Left	adult	1	1780.605957	324.680847	1791.310425	312.410278	1778.837036	312.866150	1831.564209	...	158
2580	Subject_9_M_34_Clip_2_Right	adult	1	105.050491	304.773468	104.551872	292.238708	91.365433	293.280670	63.318897	...	19
2581	Subject_9_M_34_Clip_3_Back	adult	1	925.572205	386.831543	894.036133	369.927429	956.350830	378.392670	888.792664	...	93
2582	Subject_9_M_34_Clip_3_Front	adult	1	980.323792	323.064301	990.725403	311.203339	969.772827	311.913116	1000.976807	...	99

2583 rows × 3723 columns

Figure 5. Sample data contained in the Self-Collected Dataset.

B. Assessing the Performance of GenReGait

Experiments have been conducted to evaluate the performance of the proposed method. Both the BiLSTM and GRU models were assessed and the results are presented in Table 3. The BiLSTM model attained an accuracy score of 85.8%, accompanied by a train loss of 31%. On the other hand, the GRU model achieved a slightly lower accuracy score of 84.8%, with a train loss of 32.6%. These findings indicate that the BiLSTM model exhibited a marginally superior accuracy performance on this particular dataset.

Similar evaluations were performed on the Self-Collected Dataset, comparing the BiLSTM and GRU models. The BiLSTM model attained an accuracy score of 58.8%, along with a train loss of 66.7%. In contrast, the GRU model achieved a slightly lower accuracy score of 57.2%, accompanied by a train loss of 66.8%. Comparable to the previous dataset, the BiLSTM model outperformed the GRU model in terms of accuracy on the Gait in the Wild Self-Collected Dataset as well.

The confusion matrices for the different models are depicted in Figures 6 and 7 on the Gait in the Wild and Self-Collected Datasets, respectively. We observe that the GRU model demonstrated higher accuracy in recognizing gender female data, while BiLSTM model showed better performance in identifying gender male data.

Overall, experiments conducted on the Gait in the Wild dataset yielded better results as compared to those conducted using the Self-Collected Dataset. This could be attributed to the variations in view angles present in the Self-Collected Dataset. The videos in the Gait in the Wild dataset predominantly feature straightforward walking sequences, making it easier to distinguish gait features. In addition, the experimental results also highlight the effectiveness of the BiLSTM model in both datasets, showcasing its ability to achieve higher accuracy scores compared to the GRU model.

Table 3. Results for the proposed GenReGait method.

Dataset	Model	Accuracy
Gait in the wild	BiLSTM	85.8%
Gait in the wild	GRU	84.8%
Self-collected	BiLSTM	58.8%
Self-collected	GRU	57.2%

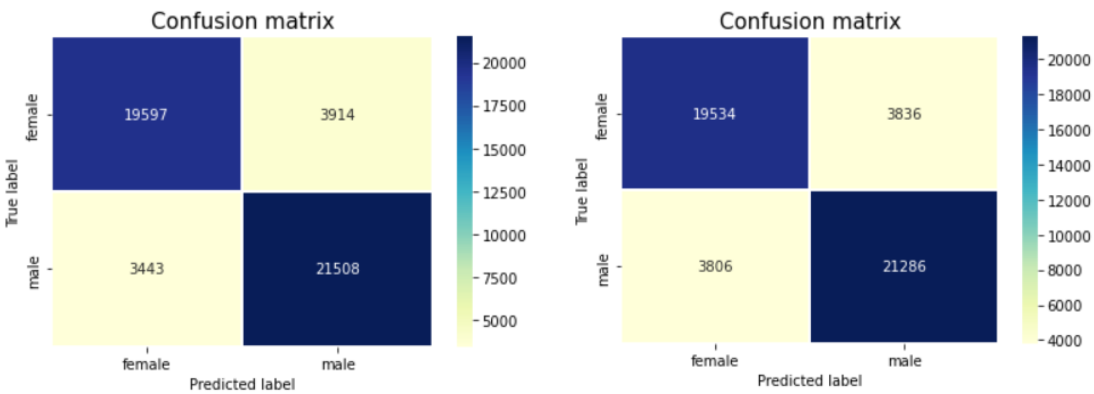

Figure 6. Confusion matrices for BiLSTM (left) and GRU (right) on the Gait in the Wild Dataset.

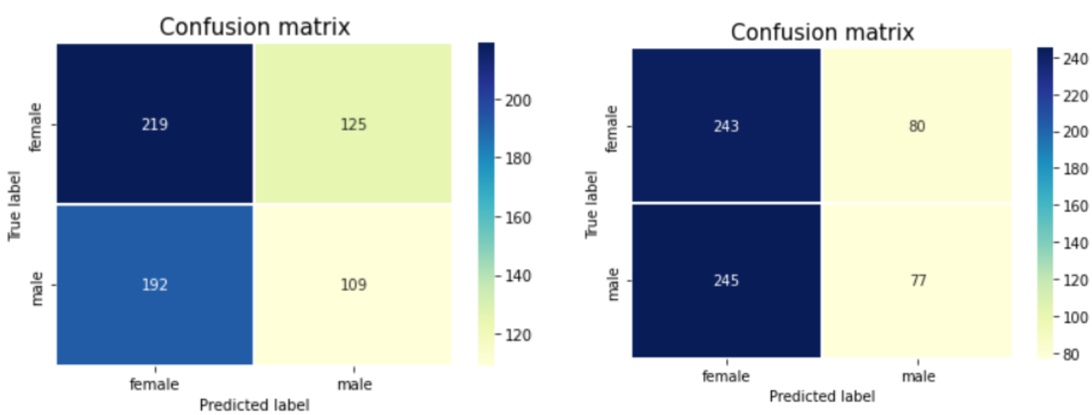

Figure 7. Confusion matrices for BI-LSTM (left) and GRU (right) on the Self-Collected Dataset.

C. Evaluating the Effects of WEMA

In this section, we want to find out the impact of the proposed WEMA technique on the performance of gender recognition. Each dataset consists of keypoints represent the *x* and *y* coordinates. To enhance the relationship between keypoints and eliminate outliers, the *x* and *y* coordinates were separated from the original data and individually filtered.

Table 4 illustrates the performance improvements achieved with the implementation of the proposed WEMA technique. The results of two other filtering techniques namely the Savitzky-Golay filter [16] and Hampel filter [17] are also included for comparison. The Savitzky-Golay filter is based on polynomial fitting and convolution and is useful for reducing high-frequency noise. On the other hand, the Hampel filter is effective at detecting and mitigating the impact of extreme values (outliers) that can distort statistical analysis or modeling. The WEMA filter clearly outperforms the other methods. It effectively identifies and removes outliers from the data. It introduces a smoothing effect by attenuating abrupt changes or fluctuations in the *x* and *y* coordinates. This smoothing effect helps to reduce noise and variability in the data, enabling the model to focus on more consistent patterns and features related to gender recognition. The smoother data aids in capturing the underlying trends and characteristics of gait signals, leading to improved accuracy in gender classification.

Figures 8 and 9 provide visualizations of the *x*- and *y*-coordinates of the gait signals, both before and after the application of WEMA. The original graph lines representing the *x*- and *y*-coordinates display frequent fluctuations, with noticeable ups and downs. This indicates a higher level of variability or volatility in the original gait signals. Since each keypoint represents different parts of the body, the resulting graph appears less smooth due to the inherent differences in movement patterns and body dynamics. However, after applying the WEMA filter, a notable change is observed in the graph lines. The filtered lines appear flatter and smoother, suggesting that the filter has effectively reduced the abrupt changes or fluctuations in the *x* and *y* coordinates. This smoothing effect provides benefit of improved performance for gender recognition.

Table 4. Impact of applying the WEMA technique.

Dataset	Model	EMA	Hampel Filter	Savitzky-Golay Filter	Proposed Method (WEMA)
Gait in the wild	BI-LSTM	82.5%	83.2	84.6	85.8%
Gait in the wild	GRU	81.7%	80.6	82.7	84.8%
Self-collected	BI-LSTM	55.5%	54.4	56.1	58.8%
Self-collected	GRU	55.5%	54.3	55.2	57.2%

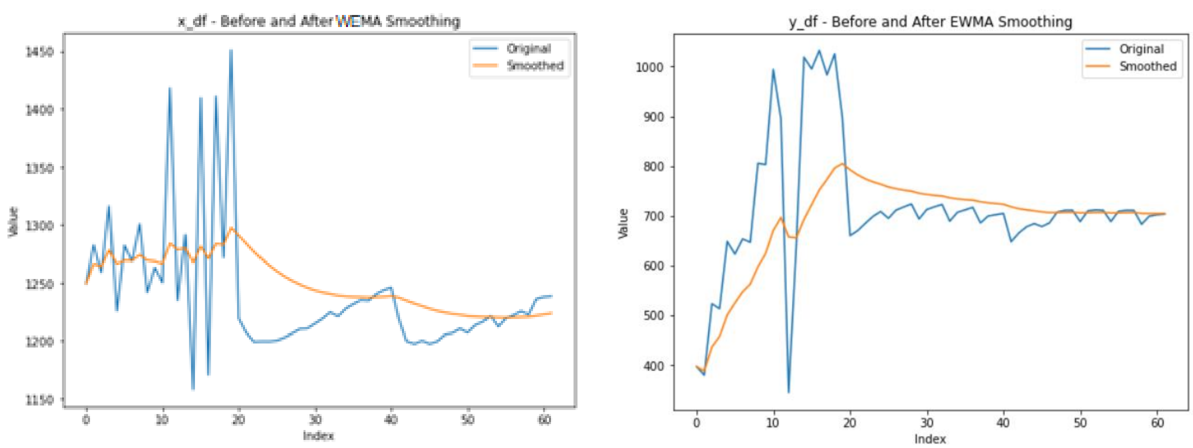

Figure 8. Gait signals before and after applying WEMA on the Gait in the Wild dataset for the *x*-coordinates (left), and *y*-coordinates (right).

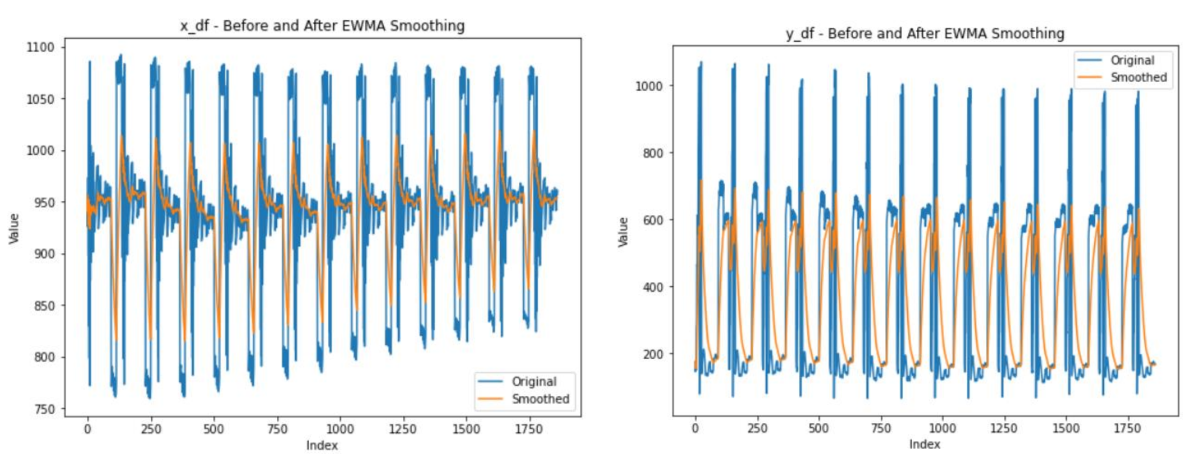

Figure 9. Gait signals before and after applying WEMA on the Self-Collected Dataset for the *x*-coordinates (left), and *y*-coordinates (right).

D. Discussions

Overall, the performance of GenReGait is comparable to the state-of-the-arts. The proposed WEMA technique in particular has outperformed the other well-known filtering methods like Savitzky-Golay filter and Hampel filter. The lower accuracy score observed in the Self-collected dataset, as compared to the Gait in the Wild dataset, can be influenced by various factors. One significant factor to consider is the feature extraction process, particularly in the Self-collected dataset. In the Self-collected dataset, the feature extraction process involves combining 30 frames into one video data segment. While this aggregation technique allows for the representation of a longer temporal sequence, it can introduce challenges in feature extraction. The process of merging frames may result in the loss of certain temporal information or introduce additional noise to the dataset.

By combining multiple frames into a single video segment, there is a possibility that important gait patterns or variations in the individual's walking style may not be adequately captured. The temporal dynamics and subtle nuances in gait, which are vital for accurate recognition, may not be effectively represented in the extracted features. Consequently, this loss of discriminative gait information can negatively impact the model's ability to distinguish between different individuals' gait patterns.

Additionally, the aggregation of frames may introduce noise or irrelevant information into the feature representation. Variations in lighting conditions, camera angles, or environmental factors across the frames can affect the quality and consistency of the extracted features. This noise or inconsistency can further hinder the model's performance in accurately classifying the gender based on gait.

Therefore, the challenges associated with aggregating frames in the Self-collected dataset's feature extraction process can contribute to the lower accuracy score observed. Addressing these challenges and refining the feature extraction technique to better capture the discriminative gait features could potentially improve the accuracy and performance of gender recognition in the Self-collected dataset.

V. CONCLUSION

This paper introduces GenReGait, an approach for gender recognition using gait features, and presents promising results obtained through this method. The experimental findings clearly indicate that the Self-collected dataset performs less accurately compared to the Gait in the Wild dataset. This disparity can be attributed to factors such as increased data variation, limited availability of training examples, and challenges in extracting relevant information. To improve accuracy, several steps will be taken in future research. These include augmenting the training data, enhancing feature extraction techniques, selecting appropriate models, and considering the unique characteristics of gait analysis. Furthermore, the exploration and research of different filtering approaches will be undertaken, aiming to identify filter methods that have a positive impact on the dataset and fulfill the requirements of the gender recognition task.

ACKNOWLEDGEMENT

This research is supported by Fundamental Research Grant Scheme (FRGS/1/2020/ICT02/MMU/02/5) and MMU IR Fund (MMUI/220026).

REFERENCES

[1] M. Xin, L. W. Ang, and S. Palaniappan, “A Data Augmented Method for Plant Disease Leaf Image Recognition based on Enhanced GAN Model Network,” *Journal of Informatics and Web Engineering*, vol. 2, no. 1, Art. no. 1, Mar. 2023, doi: 10.33093/jiwe.2023.2.1.1.

[2] L. Lee and W. E. L. Grimson, “Gait analysis for recognition and classification,” in *Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002. Proceedings*, May 2002, pp. 148–155. doi: 10.1109/AFGR.2002.1004148.

[3] Y. Makihara, H. Mannami, and Y. Yagi, “Gait Analysis of Gender and Age Using a Large-Scale Multi-view Gait Database,” in *Computer Vision – ACCV 2010*, R. Kimmel, R. Klette, and A. Sugimoto, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2011, pp. 440–451. doi: 10.1007/978-3-642-19309-5_34.

[4] Y. Wang and M. Yu, “A Study for Gender Classification Based on Gait via Incorporating Spatial and Temporal Feature Matrix,” in *2013 International Conference on Computational and Information Sciences*, Jun. 2013, pp. 1701–1704. doi: 10.1109/ICCIS.2013.444.

[5] R. Miyamoto and R. Aoki, "Gender Prediction By Gait Analysis Based On Time Series Variation Of Joint Positions," 2015. Accessed: Jun. 19, 2023. [Online]. Available: https://www.semanticscholar.org/paper/GENDER-PREDICTION-BY-GAIT-ANALYSIS-BASED-ON-TIME-OF-Miyamoto-Aoki/9c25e89c80b10919865b9c8c80aed98d223ca0c6

[6] D. Guffanti, A. Brunete, and M. Hernando, "Non-Invasive Multi-Camera Gait Analysis System and its Application to Gender Classification," *IEEE Access*, vol. 8, pp. 95734–95746, 2020, doi: 10.1109/ACCESS.2020.2995474.

[7] H. M. and R. K. Esther, "Gait Energy Image Projections based on Gender Detection using Support Vector Machines," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, Jun. 2020, pp. 1315–1320. doi: 10.1109/ICCES48766.2020.9137900.

[8] A. P. Ismail, N. M. Tahir, and A. Hussain, "Model-based gait analysis for gender recognition," in *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, Mar. 2012, pp. 400–403. doi: 10.1109/CSPA.2012.6194757.

[9] A. Wazzeh, R. G. Birdal, and A. Sertbaş, "Human Gait Based Gender Detection Using Light CNN with Max Feature Map," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, Sep. 2019, pp. 1–4. doi: 10.1109/UBMK.2019.8907174.

[10] P. Nithyakani, A. Shanthini, and G. Ponsam, "Human Gait Recognition using Deep Convolutional Neural Network," in *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)*, Feb. 2019, pp. 208–211. doi: 10.1109/ICCCT2.2019.8824836.

[11] K. T. Thomas and K. P. Pushpalatha, "Deep Learning-based Gender Recognition Using Fusion of Texture Features from Gait Silhouettes," in *Data Science and Security*, S. Shukla, X.-Z. Gao, J. V. Kureethara, and D. Mishra, Eds., in Lecture Notes in Networks and Systems. Singapore: Springer Nature, 2022, pp. 153–165. doi: 10.1007/978-981-19-2211-4_13.

[12] J. Deng, S. Bei, S. Shaojing, T. Xiaopeng, and Z. Zhen, "Gender Recognition via Fused CNN of Separated GEI," in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Jun. 2020, pp. 2032–2037. doi: 10.1109/ITNEC48623.2020.9084792.

[13] H.-S. Fang *et al.*, "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time." arXiv, Nov. 07, 2022. doi: 10.48550/arXiv.2211.03375.

[14] R. Jin, Z. Chen, K. Wu, M. Wu, X. Li, and R. Yan, "Bi-LSTM-Based Two-Stream Network for Machine Remaining Useful Life Prediction," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022, doi: 10.1109/TIM.2022.3167778.

[15] Z. Zeng, J. Sun, Z. Han, and W. Hong, "Radar HRRP Target Recognition Method Based on Multi-Input Convolutional Gated Recurrent Unit With Cascaded Feature Fusion," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3192289.

[16] K. Hasan, M. M. Othman, S. T. Meraj, S. Mekhilef, and A. F. B. Abidin, "Shunt Active Power Filter Based on Savitzky-Golay Filter: Pragmatic Modelling and Performance Validation," *IEEE Transactions on Power Electronics*, vol. 38, no. 7, pp. 8838–8850, Jul. 2023, doi: 10.1109/TPEL.2023.3258457.

[17] N. D. Nagahawatte, L. K. Cheng, R. Avci, L. R. Bear, and N. Paskaranandavadivel, "A generalized framework for pacing artifact removal using a Hampel filter," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2022, pp. 2009–2012. doi: 10.1109/EMBC48229.2022.9871096.