

Research

GAIT analysis based on GENDER detection using pre-trained models and tune parameters

Charmy Vora¹ · Vijay Katkar¹ · Munindra Lunagaria¹

Received: 16 October 2023 / Accepted: 23 February 2024

Published online: 04 March 2024

© The Author(s) 2024 [OPEN](#)

Abstract

In past several decades, gait biometrics has emerged as a viable alternative to traditional identification methods, offering advancements in surveillance, monitoring, and analysis techniques. However, determining gender based on gait remains a challenge, particularly in computer vision applications. This study proposes a robust and adaptable approach to address this issue by leveraging gait analysis. There is a growing need for datasets tailored to gait analysis and recognition to facilitate the extraction of relevant data. While most existing research relies on image-based gait datasets, this study utilizes the OULP-Age dataset from OU-ISIR, representing gait through gait energy images (GEIs). The methodology involves feature extraction from GEIs using pre-trained models, followed by classification with the XGBoost classifier. Gender prediction is enhanced through parameter fine-tuning of the XGBoost classifier. Comparative analysis of 11 pre-trained models for feature extraction reveals that DenseNet models, combined with optimized XGBoost parameters, demonstrate promising results for gender prediction. This study contributes to advancing gender prediction based on gait analysis and underscores the efficacy of integrating deep learning models with traditional classifiers for improved accuracy and reliability.

Keywords Pre-trained models · Gait analysis · Tune models · Machine learning · DenseNet · Deep learning · Gait Based Gender prediction

1 Introduction

The “Video Surveillance” category covers major events, analyses, and other information essential to security experts that install or use video surveillance systems in corporations, educational institutions, and government agencies [1, 2]. The use of gender recognition in video surveillance is challenging. Research into gender detection is difficult since a gait is a dynamic, complex target [3]. However, the biggest problem is hidden camera angles, bad lighting, and low-resolution security footage. Therefore, gait analysis is the choice method for identifying people.

Since human stride differs from other biometric traits, computer vision researchers are focusing on gait detection. [4]. Furthermore psychophysical research shows that stride can differentiate persons. Human gait is motion patterns. Gait recognition is ideal for video surveillance systems that cannot capture high-resolution facial or biometric data. Gait classification uses model-based and model-free methods [1, 5]. Model-free gait algorithms use silhouette images to compute gait features. Model-based gait methods use body components to model each person’s body shape before

✉ Charmy Vora, Charmyvora07@gmail.com; Vijay Katkar, katkarvijayd@gmail.com; Munindra Lunagaria, munindra.lunagaria@marwadieducation.edu.in | ¹Department of Computer Engineering, Marwadi University, Rajkot, India.



measuring features. Gait identification has various uses, including gender, age, diagnosis, action recognition, and security footage monitoring.

Investigations using gait analysis typically aim to achieve their primary objectives first and foremost, which are person identification and authenticity. On the other hand, gender detection are examples of types of soft biometric identification that have a great application potential [6]. Furthermore, some application domains may benefit from the ability to use an individual's stride length to estimate their gender. Currently, Most of the available research on gait-based gender recognition uses the gait feature that is taken from silhouette sequence of gait cycle. This gait feature is known as GEI [7, 8].

GEI has been used in the vast majority of the research conducted on gender detection. The GEI algorithm merges the silhouettes from a single walking cycle in its entirety. Each image object's brightness in GEI indicates the entire motion pattern cycle of the gait [9]. We are able to tackle the problem of gait analysis based on gender detection by employing machine and deep learning. GEIs used to extract features with the assistance of pre-trained models, and classifiers are put to use to classify gender. Moreover, adjusting the classifier's parameters can further improve the proposed system's efficiency.

2 Analysis on biometric gait and feature extraction

Biometric identification uses a person's unique physical or behavioral features. Biometrics including fingerprints, face features, iris patterns, voice, and gait are utilized for secure and reliable identification. Bio-metric gait recognizes people by their walking motion. It measures stride, stance, and swing gait cycles.

2.1 Analysis on biometric gait

A walking motion of humans is referred to as their "gait" [1, 4, 5]. The term "gait cycle" refers to the period of time or the sequence of events that takes place between the impact of one foot on the ground by a human and the subsequent impact of the same leg the floor. The following three things are the most important aspects that contribute to gait such as stride, stance stage and swing stage [5, 8]. The term "stride" describes the distance among two consecutive foot contacts made by the same foot. There are four components to the stance stage, which are as follows: (1) InitialContact, (2) LoadingResponse, (3) MidStance, and (4) TerminalStance. During the stance stage, 60% of the leg is hit on floor. Whereas swing stage, 40% of that same leg is not in contact with the ground that includes the following components: (1) pre-swing, (2) beginning of the swing, (3) middle of the swing, and (4) swing-termination.

2.2 Motivation

Following are some of the reasons why gait analysis is useful for recognizing and identifying individuals:

- Gait recognition provides remote gait monitoring. And it can be carried out using small and cheap equipment. Simple devices detect gait. Cameras, smart phones, and floor sensors can measure human stride.
- Low-resolution gait analysis is feasible. Low-quality video may affect face recognition. Gait recognition uses silhouettes and motions. There is no need for participant participation in order to do gait identification.

2.3 Feature extraction

The gait recognition research focuses on silhouette and GEI images for spatial and temporal feature analysis of human walking images.

2.3.1 Silhouette images

The process of extracting features was carried out on the video in order to make it possible to object recognition with a higher degree of precision. As a consequence, the visual sequence was presented in the form of silhouettes. Reconstructing a body's silhouette requires, to perform thresholding and removed background, and after that, a 3×3 median filter operator was applied in order to eliminate any isolated pixels that may have been present [9, 10].

2.3.2 Gait energy images (GEI)

The weighted average method is utilized by GEI in order to depict the gait sequence of gait cycle [7, 11] from the images captured during the Gait Cycle ($BT(x, y)$), we can derive the GEI:

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N Bt(x, y) \quad (1)$$

Time t that was referred to as (x, y) . Each image B has its own set of coordinates in space, denoted by x and y , and number of frames took from gait cycle is denoted by N .

2.4 GEI based application for gait analysis

Following are the GEI base application for gait analysis that can be useful for recognizing and identifying individuals:

- The GEI image lets us assess the person's stance and swing frequency to predict their ideal stride frequency. We can also predict and order angles. As a result, it will help to adopt a more precise individual motion pattern from their arms, head and legs.
- When it comes to creating system based on gender inquiries. The entry and restrictions requirements for colleges and hostel depending on gender, are a few examples of the kind of applications that fall within this category.

3 Related work

Kwon and [12]. Lee investigated JSE for three-dimensional detection and recognition. JSE kinematic gait measures body joint distance from anatomical planes during model skeleton walking. Instead of dynamic motion, static positions are used to create anatomical planes. Walking skeleton model gait sequences generate median, transverse, and frontal body-centered coordinates in the initial stage. The second stage extracts the three planes' JSEs. Feature extractions merge JSEs. Labelled JSE data completes the classification model. The trained classification model classifies gender-unlabeled JSE data. Four datasets use them: A, B used Kinect v1 and v2, C, and D used UPCVgaitK1 and K2. They explored different machine learning gender classification methods. In dataset B, JSE-SVM scored 98.08%.

Upadhyay and Gonsalves [13] addressed the angles, which prevented the camera from seeing the individual's movement of their body parts and gait-based features for gender detection. After DCT feature vector application, the XGBoost-classifier classifies gender. The proposed system's performance is assessed using the OU-MVLP dataset. The experiment uses fourteen viewing angles to determine gender with a CCR of 95.33%.

Khabir et al. [14], they overcame these challenges and showed how to classify gender from gait dataset based on inertial sensor from the massive OU-ISIR dataset. They found that SVM had the maximum accuracy for gender classification at 84.76%.

Xu et al. [15], they examined an individual frame to determine probability of gender identification. Rather than employing a sequence of gait data for real-world applications. They utilized OL-MVLP, which included 10,307 subjects. The SVM algorithm's CCR for gender classification is 94.27% for a single image.

Bei et al. [10], they investigated optical camera sensor-based gender recognition. SubGEI from the gait cycle extracts temporal movement cues on fusion body appearance for more accurate gait analysis in tough settings. Two-stream CNN reflects GEI and also handles motion information of subGEI better than CNN models. CASIAB gait dataset. The investigations used a CNN with four layers, Inception-V3, and VGG16 in the first-stream and 3 sub-GEIs, TL1, TL2, and TL3, reflecting sub-GEIs having four, six, and eight picture frames in the second-stream. The experiment analyses 11 viewing angles, with the 90° angle yielding the best gender categorization accuracy of 95% when applying the InceptionV3 algorithm on TL-2 sub-GEI.

Gillani et al. [16], they examined based on the attributes that were collected from the data for gender detection. For the purpose of obtaining the inertial signals, Inertial Measurement Units sensors were utilized. These sensors had triangle gyroscopes and accelerometers. They employed the OU-ISIR dataset; they divided the dataset into sequences, and the participants were instructed to walk on a predetermined path two times: once when reaching the route for the first time (seq 0) and once while return to the same route (seq 1). The LogisticRegressor classifier has the highest CCR for gender classification, 68.2% for sequence 0 and 65% for sequence 1.

Chen et al. [17], They developed a customized dataset in which they took into consideration Using insole pressure mats that determined the Center of pressure for feature extraction, where researchers captured 960 steps from each of the 24 younger and older subjects in the study. For the purpose of applying SVM, they split the 30 features in four-stages, such as the initial contact stage, forefoot contact stage, foot flat stage, and the fore foot push off stage. They used 13 different steps, one for each participant, to obtain an accuracy of 95% for the gender classification.

4 Proposed method

In comparison to previous research efforts in gait recognition, which have employed various methodologies and datasets, this study introduces a novel approach centered on the OULP-Age dataset and gait energy images (GEIs) for gender prediction. By utilizing pre-trained models for feature extraction from GEIs and fine-tuning the parameters of the XGBoost classifier, this research presents a distinctive perspective on gait analysis and gender prediction.

Unlike prior studies, which focused on different datasets and methodologies such as Joint Symmetry Estimation (JSE) or Discrete Cosine Transform (DCT), our approach offers a unique dataset and methodology for gender prediction, showcasing the effectiveness of DenseNet pre-trained models and optimized XGBoost parameters. This contribution adds value to the field by addressing the need for improved gender prediction based on gait analysis and provides insights into optimizing feature extraction and classification techniques for gender prediction in gait analysis.

The sections 4.1 to 4.4 provide explanations of the proposed method for predicting a person's gender.

4.1 Flow of proposed method

The purpose of the gait analysis method is to recognize and identify individuals' genders using their unique gait patterns. Figure 1 depicts the flow of the proposed method. In the beginning, analysis of gait is performed using the OULP-Age dataset. After that, the dataset was separated for training and testing phases, and a range of pre-trained models were used to extract features for gait analysis. Once the features have been extracted, the classifier should be used to classify the features. Moreover, tuning of the classifier model's parameters is also performed to improve the accuracy of the results. Based on this, compared the performance of the classifier with the tuning parameters of the classifier to analyze the results for predicting the person's gender based on gait analysis.

In Training and Testing phase, the decision to split the dataset evenly into 50% for training and 50% for testing is grounded in fundamental principles of machine learning and data science. This balanced allocation ensures that the model is exposed to an adequate amount of data during training, facilitating effective learning of patterns and

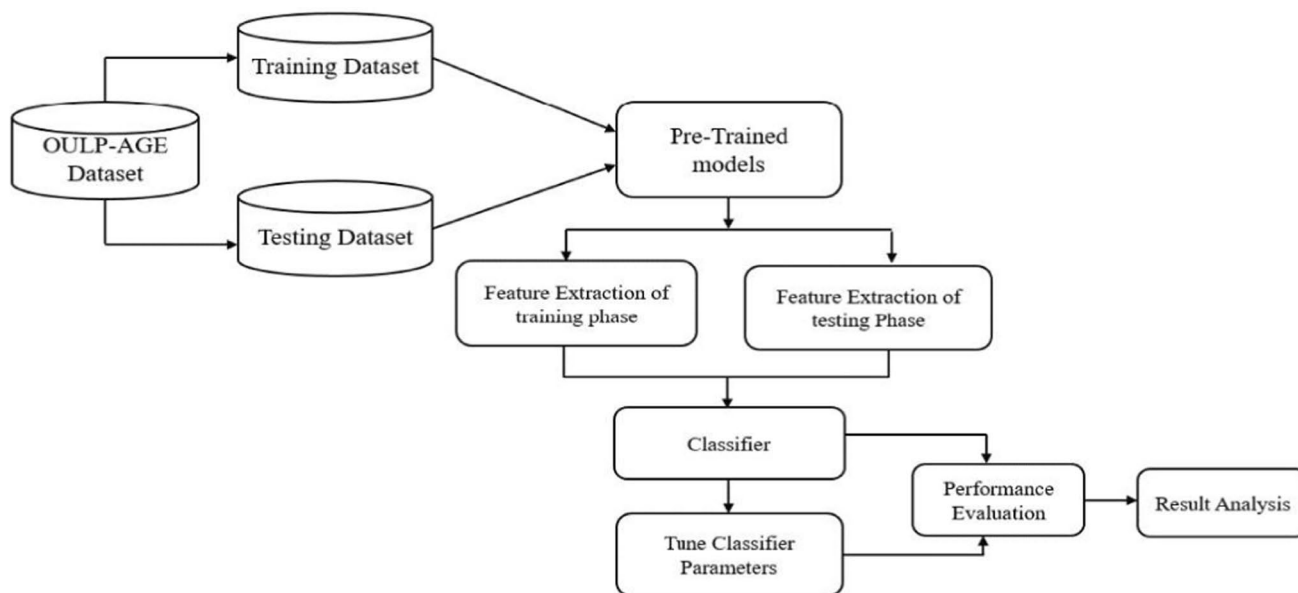


Fig. 1 Flow of the proposed method

relationships within the dataset. Simultaneously, an equal distribution for testing enables a rigorous evaluation of the model's performance on unseen data, essential for assessing its ability to generalize to new examples. This approach mitigates the risks of overfitting or underfitting by providing sufficient training data while allowing for a comprehensive assessment of the model's performance. Moreover, the 50–50 split simplifies the evaluation process, enabling a clear comparison between predicted outcomes and actual observations. Overall, this methodology underscores a commitment to balance, effective evaluation, and robust generalization to new data, enhancing the reliability and efficacy of the machine learning model.

4.2 Pre-trained models

"Pre-trained" term is used in machine learning to describe when parts of a previously trained model are included into new models or used to tackle a different problem [18–20]. This method of machine learning model construction requires less time, money, and labeled data to produce accurate results. With the help of the pre-trained model, we achieved satisfying results with the DenseNet model, the details of which will be shown in the following section.

4.2.1 DenseNet

In our main proposed method, we relied on DenseNet. It uses TensorFlow 2.0 (TF 2.0) and Keras and comes in a variety of versions, like DenseNet (121, 169, and 201). Besides the standard convolutional and pooling layers, DenseNet has two other crucial blocks. DenseNet different version shares the same convolution block, pooling layer, transition layer, and classification layer architectures. However, DenseNet each version have their own unique set of four DenseBlock with different repetition times [21].

The first convolutional block is constructed of 64 filters that are 7×7 in size and have a stride of 2. After this, a MaxPooling layer will take its place, employing a 2 stride with 3×3 max pooling configuration. The Batch Normalization, ReLu activation, and Conv2D layers follow the input layer for each convolutional block. Each dense block performs a pair of convolutions using kernel sizes of 1×1 and 3×3 .

- DenseNet-121 dense Block1, dense Block2, dense Block 3 and dense Block 4 is repeated 6 times, 12 times, 24 times, and 16 times respectively.
- DenseNet-169 dense Block1, dense Block2, dense Block 3 and dense Block 4 is repeated 6 times, 12 times, 32 times, and 32 times respectively.
- DenseNet-201 dense Block1, dense Block2, dense Block 3 and dense Block 4 is repeated 6 times, 12 times, 48 times, and 32 times respectively.

Transition layer channels must be half. The function specifies a 1×1 kernel size for the 2×2 average pooling layer and 2 stride for such convolutional. Classifier Layer uses 7×7 global mean pooling and 1000D connected softmax.

4.2.2 VGG16

VGG 16 includes 13 convolutional and three fully interconnected layers. The first two convolutional layers each have $64 \times 3 \times 3$ -sized filters after two more with 128 filters, two with 256 filters, two with 512 filters, and the final with 512 filters. Three fully connected layers with 4096 neurons each follow the convolutional layers, followed by a 1000-neuron classification layer [22].

4.2.3 VGG19

VGG19 includes 19 layers—3 fully linked and 16 convolutional. The first two convolutional layers have $64 \times 3 \times 3$ -pixel filters, followed by two more with 128 filters, four with 256 filters, four with 512 filters, and the last with 512 filters. Three fully connected layers with 4096 neurons each follow the convolutional layers, followed by a 1000neuron classification output layer [23, 24].

4.2.4 ResNet50

It is called ResNet50 because it has 50 layers. ResNet50 accepts $224 \times 224 \times 3$ images. First 7×7 convolutional layer with stride 2 compresses the input image to $112 \times 112 \times 64$. Batch normalisation and ReLU activation follow. Three convolutional layers with different numbers of filters follow: Block 1: 3×3 convolutional layers with 64 filters, repeated 3 times; Block 2: 128 filters, repeated 4 times; Block 3: 256 filters, repeated 6 times; and Block 4: 512 filters, repeated 3 times. ResNet50 uses a 1×1 convolutional layer to map input to block output. After convolutional layers, a global average pooling layer decreases output spatial dimensions to 1×1 . Then comes a fully linked layer with 1000 units and a softmax activation function, representing the 1000 ImageNet classes [24].

4.2.5 NASNetMobile

NASNetMobile has 22 layers grouped into repeating cells. The input layer receives image data, which is processed through a series of layers: a 3×3 convolutional layer with 32 filters and a stride of 2, a batch normalisation layer, a ReLU activation layer, a 3×3 separable convolutional layer with 32 filters and a stride of 1, another normalisation layer, a second ReLU activation layer, and a 3×3 max pooling layer [25].

4.2.6 NASNetLarge

NASNetLarge is an advanced image classification architecture based on NasNetMobile. Its 87-layer network design is cell-based. Image data enters at the first layer. This is followed by a 3×3 convolutional layer with 96 filters and a stride of 2, a batch normalisation layer that normalises the output of the previous layer to speed up training, and a ReLU activation layer that introduces non-linearity. Another normalisation layer, ReLU activation layer, and 5×5 separable convolutional layer with 256 filters and 1 stride follow. A stride 3×3 max pooling layer finishes the cell [26].

4.2.7 Xception

Xception (short for “Extreme Inception”) is a deep convolutional neural network architecture. Xception accepts $299 \times 299 \times 3$ pictures (where 3 is the number of channels for RGB images). Instead of typical convolutions, Xception uses depthwise separable convolutions, which use a depth wise convolution followed by a point wise convolution. After convolutional layers, a global average pooling layer decreases output spatial dimensions to 1×1 [27].

4.2.8 InceptionResNetV2

InceptionResNetV2 combines the Inception and ResNet networks. Skip connections enhance gradient propagation during training in its 164 layers. A classification layer, stem module, numerous Inception-ResNet-A, B, and C modules, and the stem module processes images via convolutional, pooling, and activation layers. Inception-ResNetA, B, and C modules use 1×1 convolutions and max pooling to collect features at different scales. Ultimately, a global average pooling layer and fully linked layer produce class predictions in the classification layer [26].

4.2.9 InceptionV3

The input layer of InceptionV3 allows shape images (299, 299, 3). Two convolutional layers, max pooling, and batch normalisation reduce the input image’s spatial dimensions and extract basic properties. 17 Inception modules have convolutional filters and pooling procedures. The following module receives concatenated module outputs. Inception modules 5 and 11 add two auxiliary classifiers. Each has a classification softmax layer, global average pooling layer, and 2 fully-connected layers with ReLU-activated. The last layer adds a global average pooling layer, fully—connected layer, and classification softmax output layer [26].

4.3 Evolution metrics for purposed method

The accuracy of gender prediction can be determined by using the evolution metrics that are listed below.

Confusion Matrix A useful classifier is the confusion matrix. Both multiclass and binary categorization are effective. It analyzes algorithmic flaws, compares model output to actual output, and provides statistics for True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

TP: the number of correct positive predictions made by the model.

FP: the number of incorrect positive predictions made by the model.

TN: the number of correct negative predictions made by the model.

FN: the number of incorrect negative predictions made by the model.

Also there are also following statistical tools given the gender distribution of the underlying data: true positive rate (TPR), false positive rate (FPR), true negative rate (TNR), and false negative rate (FNR).

TPR: The ratio of correctly classified positive instances to the total actual positive instances.

$$TPR = \frac{TP}{TP + FN} * 100 \quad (2)$$

FPR: The ratio of incorrectly classified positive instances to the total actual negative instances.

$$FPR = \frac{FP}{FP + TN} * 100 \quad (3)$$

TNR: The ratio of correctly classified negative instances to the total actual negative instances.

$$TNR = \frac{TN}{TN + FP} * 100 \quad (4)$$

FNR: The ratio of incorrectly classified negative instances to the total actual positive instances.

$$FNR = \frac{FN}{FN + TP} * 100 \quad (5)$$

Precision: it measures a model's positive predictions. Hence, precision is the fraction of real positive predictions (i.e., accurately predicted positive samples) over the model's total positive predictions.

$$Precision = \frac{TP}{TP + FP} * 100 \quad (6)$$

Recall: it is commonly called as sensitivity or true positive rate, assesses a model's accuracy in identifying actual positives. True positives divided by the sum of true positives and false negatives is the recall.

$$Recall = \frac{TP}{TP + FN} * 100 \quad (7)$$

F1-Score: it combines precision and recall, is used to evaluate binary classification models.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} * 100 \quad (8)$$

Correct Classification Rate (CCR): It estimates the proportion of samples correctly classified. It is determined by dividing correct predictions by total predictions.

$$CCR = \frac{TP + TN}{TP + TN + FP + FN} * 100 \quad (9)$$

4.4 XGBoost classification

"XGBoost" is the abbreviation for the word "extreme gradient boosting." The implementation of gradient-boosted decision trees is done with the help of the XGBoost algorithm. In XGBoost, the application of weights is an essential component.

Each independent variable receives a weight before being included into the decision tree that is used to make predictions about the results [28]. Variables whose outcomes were incorrectly forecasted by the first decision tree are given added weight before moving on to the second tree. After that, these separate classifiers and predictors are combined in order to produce a model that is both reliable and precise.

5 Experiments and results

Our experiment was designed to assess the accuracy of gender prediction using a gait analysis dataset. To achieve this, we applied a variety of pre-trained models and adjusted classifier parameters. We then measured performance using a range of evaluation metrics to obtain comprehensive results.

5.1 Software and setup tools

We utilized Anaconda and Pycharm. Furthermore, for implementation, we adopted TensorFlow 2.9.1 and imported various Python libraries, including pandas, Sklearn, NumPy, Scikit-learn, Keras, XGBoost, and OpenCV-python.

5.2 Experiment dataset

In this study, we utilized the OU-ISIR Gait Database's OULP-Age. It consists of 63,846 gait photos of people moving down a path. The age range of the participants is 2 to 90. GEIs sequences of 88×128 pixels that were gathered from a side angle of each participant's gait. A testing set of 31,923 subjects (16,426 women and 15,407 men) and a training set of 31,923 subjects (16,327 women and 15,596 men) were created from the database based on the predefined protocol [29].

5.3 Training and testing phase

Including both the training and testing sets, first we used pre-trained models (from TensorFlow 2.0 (TF 2.0) and Keras) for extracting the features on GEI images from OULP—Age dataset. After that, XGBClassifier (from XGBoost) used to classify and predict the accuracy for gender detection using different evolution metrics.

5.4 Tuning classifier

By adjusting boosting settings including max depth, learning rate, and n estimators, this research fine-tunes the XGBoost classifier parameters. The learning rate has been adjusted to fall between 0.01 and 1.0, max depth between 4 to 6 and n estimators was up to 1000.

5.5 Result comparison and discussion

In Table 1 highlights prediction of gender classification in training phase where the highest CCR score is achieved by DenseNet201 with a value of 93.88%. This model also has the highest precision and F1-score for both female and male genders, with values of 95.31% and 94.01% for precision, and 94.32% and 93.75% for F1-score, respectively. The recall for both genders is also high, with values of 92.74% and 92.43% for female and male, respectively. On the other hand, the lowest CCR score is achieved by NASNetMobile with a value of 88.80%. This model also has the lowest precision and F1-score for both female and male genders, with values of 89.17% and 88.73% for precision, and 88.57% and 88.57% for F1-score, respectively. The recall for both genders is also relatively low, with values of 89.02% and 88.41% for female and male, respectively.

In Table 2 shows that DenseNet201 model has the greatest TPR of 92.74% and the highest TNR of 95.11%, demonstrating its high degree of accuracy in correctly identifying positive and negative situations. Also, it has the lowest FPR of all models (4.89%), meaning that it predicts less false positives than other models. Also, the NASNetMobile model has the greatest false positive rate (FPR), which indicates that it generates more false positive predictions than other models. Also, it has the greatest FNR (11.14%), which shows that it falsely labels more genuine positive situations as negative. Overall, according on the metrics supplied in the table, DenseNet201 model appears to be the best-performing model, whereas NASNetMobile model looks to be the worst-performing model.

Table 1 Prediction of gender detection in training phase

Model	Precision		Recall		F1-score		CCR (%)
	Female (%)	Male (%)	Female (%)	Male (%)	Female (%)	Male (%)	
VGG16	91.90	90.00	90.94	89.78	91.72	90.74	90.84
VGG19	90.54	88.55	89.53	88.15	90.19	89.16	89.35
NASNetLarge	92.99	92.03	92.50	91.64	92.64	92.14	92.33
NASNetMobile	89.17	88.86	89.02	88.41	88.73	88.57	88.80
Xception	91.92	89.98	90.94	89.60	91.60	90.59	90.77
ResNet50	92.64	91.07	91.85	90.83	92.44	91.63	91.74
InceptionResNetV2	90.51	89.29	89.90	89.01	90.26	89.63	89.77
InceptionV3	93.05	90.93	91.98	90.86	93.00	91.92	91.95
DenseNet121	94.49	92.53	93.50	92.30	94.32	93.30	93.40
DenseNet169	94.65	92.73	93.68	92.64	94.58	93.60	93.64
DenseNet201	95.31	92.74	94.01	92.43	95.11	93.75	93.88%

The bold value indicates good results compare to other models

Table 2 TPR, FPR, TNR and FNR of gender detection in training phase

Model	TPR (%)	FPR (%)	TNR (%)	FNR (%)
VGG16	90.00	8.28	91.72	10.00
VGG19	88.55	9.81	90.19	11.45
NASNetLarge	92.03	7.36	92.64	7.97
NASNetMobile	88.86	11.27	88.73	11.14
Xception	89.98	8.40	91.60	10.02
ResNet50	91.07	7.56	92.44	8.93
InceptionResNetV2	89.29	9.74	90.26	10.71
InceptionV3	90.93	7.00	93.00	9.07
DenseNet121	92.53	5.68	94.32	7.47
DenseNet169	92.73	5.42	94.58	7.27
DenseNet201	92.74	4.89	95.11	7.26

Table 3 Prediction of gender detection in testing phase

Model	Precision		Recall		F1-score		CCR (%)
	Female (%)	Male (%)	Female (%)	Male (%)	Female (%)	Male (%)	
VGG16	92.54	89.32	90.90	88.88	92.22	90.52	90.72
VGG19	90.26	88.19	89.21	87.82	87.82	88.88	89.05
NASNetLarge	93.02	91.50	92.25	91.18	92.74	91.95	92.11
NASNetMobile	88.08	88.76	88.42	88.03	87.31	87.66	88.05
Xception	92.47	90.58	90.58	89.98	91.98	90.97	91.25
ResNet50	92.90	91.28	92.08	90.79	92.50	91.64	91.86
InceptionResNetV2	90.32	87.90	89.09	87.75	90.20	88.96	89.03
InceptionV3	93.24	93.24	91.90	90.50	93.16	91.81	91.86
DenseNet121	93.39	92.99	93.62	92.47	93.82	93.14	93.39
DenseNet169	94.70	93.30	94.00	93.09	94.53	93.81	93.90
DenseNet201	94.17	93.25	93.71	93.08	94.03	93.55	93.63

The bold value indicates good results compare to other models

In Table 3 demonstrates the prediction of gender classification in testing model. The DenseNet169 achieved the highest CCR of 93.90%. It also had the highest precision and recall scores for both genders, with precision values of 94.70% and 93.30% for females and males, respectively, and recall values of 94.00% and 93.09% for females and males, respectively. Its F1-score was also high, with values of 94.53% and 93.81% for females and males, respectively.

Table 4 TPR, FPR, TNR and FNR of gender detection in testing phase

Model	TPR (%)	FPR (%)	TNR (%)	FNR (%)
VGG16	89.32	7.78	92.22	10.68
VGG19	88.19	10.04	89.95	11.81
NASNetLarge	91.50	7.26	92.74	8.50
NASNetMobile	88.76	12.69	87.31	11.24
Xception	90.58	8.02	91.98	9.42
ResNet50	91.28	7.50	92.50	8.72
InceptionResNetV2	87.90	9.80	90.20	12.10
InceptionV3	90.60	6.84	93.16	9.40
DenseNet121	92.99	6.18	93.82	7.01
DenseNet169	93.30	5.47	94.53	94.53
DenseNet201	93.25	5.97	94.03	6.75

Table 5 Prediction of Gender Detection using DenseNet with Tune Parameter of XGBoost classifier

Model	Training (%)	Testing (%)	Tune Classifiers (XGBoost Classifier)			
			Learning rate	Number of estimators	Max depth	CCR (%)
DenseNet-121	93.40	93.39	0.28	917.49	4.10	95.17
DenseNet-169	93.64	93.90	0.38	992.18	5.76	95.13
DenseNet-201	93.88	93.63	0.23	996.04	4.55	95.41

The bold value indicates good results compare to other models

While NASNetMobile model achieved the lowest CCR of 88.05%. Although it had a high precision score for males (88.76%), its precision score for females was relatively low (88.08%). Its recall score was also lower than most of the other models, with values of 88.42% and 88.03% for females and males, respectively. Its F1-score was also relatively low, with a value of 87.31% for females and 87.66% for males.

In Table 4 illustrates that the DenseNet169 has the highest TPR at 93.30%, meaning it correctly identifies a high proportion of positive cases. DenseNet121 has the lowest FPR at 6.18%, meaning it makes the fewest incorrect positive identifications. DenseNet169 also has the highest TNR at 94.53%, meaning it correctly identifies a high proportion of negative cases, while InceptionResNetV2 has the highest FNR at 12.10%, meaning it incorrectly identifies a high proportion of positive cases as negative. Overall, it seems that DenseNet models perform well on this task, while InceptionResNetV2 may be the weakest performer.

The results from Table 5 demonstrate the prediction accuracy of gender detection using DenseNet models with tuned parameters of the XGBoost classifier. These findings highlight significant improvements in accuracy achieved through parameter tuning, shedding light on the enhanced performance of the models after tuning the classifier.

Before parameter tuning, DenseNet-121, DenseNet-169, and DenseNet-201 exhibited comparable levels of accuracy throughout the training and testing phases for gender detection using the XGBoost classifier. However, after fine-tuning the classifier, notable improvements in accuracy were observed across all models. For instance, DenseNet-201, which initially performed well with an accuracy of 93.88% during the training phase, experienced a further enhancement in accuracy, reaching an impressive 95.41% after tuning the classifier. Similarly, DenseNet-169, which achieved the highest accuracy of 93.90% during the testing phase before tuning, saw its accuracy improve to 95.13% after parameter optimization.

Comparing the performance of DenseNet models before and after tuning the classifier underscores the effectiveness of this optimization process in boosting prediction accuracy. The results demonstrate that parameter tuning enhances the capabilities of the XGBoost classifier, leading to more accurate gender detection outcomes across all DenseNet models.

In summary, the findings from Table 5 underscore the importance of parameter tuning in improving the accuracy of gender detection models based on gait analysis. By optimizing the parameters of the XGBoost classifier, significant enhancements in prediction accuracy can be achieved, ultimately contributing to the effectiveness of DenseNet models in gender prediction tasks.

6 Conclusion

The intention of this research is to demonstrate an approach for predicting gender based on gait analysis that is both cost-effective and adaptive method. Following the extraction of the features, the system being presented makes use of GEs for human gait representation. Further, XGBoost is utilized for the classification of gender detection. When it comes to gait recognition, GEI is superior to binary silhouette sequences in terms of its ability to reduce both store space and computing time.

An approach that is based on appearance is utilized for gait analysis by the proposed system while it is performing the detection of gender. On the OULP-Age dataset provided by OU-ISIR, the suggested system is examined to determine how well it performs. The comparison of the results obtained in the experiment with training, testing, and tuned classifiers demonstrates that the suggested system has superior performance using the DenseNet model with XGBoost Classifier. The results of the comparison reveal that using the tuned parameters of XGBoost Classifier leads to superior accuracy. The proposed system achieved 95.41% accuracy for gender detection respectively.

However, the system that has been proposed does not take into consideration other issues that are associated with gait-based analysis. Some of these issues include occlusion of gait as a result of different conditions regarding holding of baggage and the loose fitting attire. In addition to the approach taken for the analysis from a variety of view angles. Future work in this project will address the difficulty with walking due to a blockage, varied viewpoint angles and build approach resilient to limited access to gait.

Acknowledgements In our research, we utilize the OU-ISIR Gait Database, Multi-View Large Population Dataset (OU-MVLP), maintained by the Institute of Scientific and Industrial Research (ISIR) at Osaka University. This dataset aids in cross-view gait recognition algorithm development by providing a comprehensive collection of gait videos and associated data. ISIR holds the copyright to this dataset, and we will duly acknowledge its contribution in our manuscript. As an open-access resource, no additional permissions are required for its use in our study.

Author contributions CV designed, performed and analysis of training and testing data with achieved output results as well as prepared and wrote the manuscript. VK analysis the input data for training and testing and analysis the achieved output and prepared major key points of manuscript. ML commented on necessary requirements for manuscript and wrote the Supplementary Information. All authors discussed the results and implications and commented on the manuscript at all stages.

Data availability Data supporting this study are available from OU-ISIR gait database comprising the large population dataset (<https://doi.org/https://doi.org/10.1109/TIFS.2012.2204253>) [29].

Declarations

Ethics approval and consent to participate This article does not contain any studies with human participants or animals performed by any of the authors.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Singh JP, Jain S, Arora S, Singh UP. Vision-based gait recognition: a survey. IEEE Access. 2018;6:70497–527. <https://doi.org/10.1109/ACCESS.2018.2879896>.
2. Zhao Y. Effective gait feature extraction using temporal fusion and spatial partial school of computer science, school of artificial intelligence, optics and electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China. 2021; 1244–1248.
3. Zafaruddin GM, Fadewar HS. Face recognition using eigenfaces. Singapore: Springer Singapore; 2018. https://doi.org/10.1007/978-981-13-1513-8_87.
4. Zhang, Z., Hu, M., Wang, Y. 2011. A survey of advances in biometric gait recognition. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 7098 LNCS, 150–158. https://doi.org/10.1007/978-3-642-25449-9_19.
5. Wan C, Wang L, Phoha VV. A survey on gait recognition. ACM Comput Surv. 2018. <https://doi.org/10.1145/3230633>.

6. Kale A, Sundaresan A, Rajagopalan AN, Cuntoor NP, Roy-Chowdhury AK, Krüger V, Chellappa R. Identification of humans using gait. *IEEE Trans Image Process*. 2004;13:1163–73. <https://doi.org/10.1109/TIP.2004.832865>.
7. Luo J, Zi C, Zhang J, Liu Y. Gait recognition using GEI and curvelet. *Guangdian Gongcheng/Opto-Electronic Eng*. 2017;44:400–4. <https://doi.org/10.3969/j.issn.1003-501X.2017.04.003>.
8. Yaacob NI, Tahir NM. Feature selection for gait recognition. *SHUSER 2012—2012 IEEE Symp Humanit Sci Eng Res*. 2012. <https://doi.org/10.1109/SHUSER.2012.6268871>.
9. Collins RT, Gross R, Shi J. Silhouette-based human identification from body shape and gait. In: *Proc—5th IEEE Int Conf Autom. Face Gesture Recognition, FGR*, 2002. <https://doi.org/10.1109/AFGR.2002.1004181>.
10. Bei S, Deng J, Zhen Z, Shaojing S. Gender recognition via fused silhouette features based on visual sensors. *IEEE Sens J*. 2019;19:9496–503. <https://doi.org/10.1109/JSEN.2019.2916018>.
11. Hema, M., Esther Rachel, K. Gait energy image projections based on gender detection using support vector machines. *Proc 5th Int Conf Commun Electron Syst ICCES*, 2020. <https://doi.org/10.1109/ICCES48766.2020.09137900>.
12. Kwon B, Lee S. Joint swing energy for skeleton-based gender classification. *IEEE Access*. 2021;9:28334–48. <https://doi.org/10.1109/ACCESS.2021.3058745>.
13. Upadhyay J, Gonsalves T. Robust and lightweight system for gait-based gender classification toward viewing angle variations. *AI*. 2022;3:538–53. <https://doi.org/10.3390/ai3020031>.
14. Khabir KM, Siraj MS, Ahmed M, Ahmed MU. Prediction of gender and age from inertial sensor-based gait dataset. 2019 *Jt 8th Int Conf Informatics, Electron Vision, ICIEV 2019 3rd Int Conf Imaging, Vis Pattern Recognition, icIVPR 2019 with Int Conf Act Behav Comput ABC*, 2019;371–376. <https://doi.org/10.1109/ICIEV.2019.8858521>.
15. Xu C, Makihara Y, Liao R, Niitsuma H, Li X, Yagi Y, Lu J. Real-time gait-based age estimation and gender classification from a single image. *Proc 2021 IEEE Winter Conf Appl Comput Vision WACV*. 2021. <https://doi.org/10.1109/WACV48630.2021.00350>.
16. Gillani SI, Azam MA, Ehatisham-UI-Haq M. Age estimation and gender classification based on human gait analysis. 2020 *Int Conf Emerg Trends Smart Technol ICETST*. 2020. <https://doi.org/10.1109/ICETST49965.2020.9080735>.
17. Chen YJ, Chen LX, Lee YJ. Systematic evaluation of features from pressure sensors and step number in gait for age and gender recognition. *IEEE Sens J*. 2022;22:1956–63. <https://doi.org/10.1109/JSEN.2021.3136162>.
18. Han X, Zhang Z, Ding N, Gu Y, Liu X, Huo Y, Qiu J, Yao Y, Zhang A, Zhang L, Han W, Huang M, Jin Q, Lan Y, Liu Y, Liu Z, Lu Z, Qiu X, Song R, Tang J, Wen JR, Yuan J, Zhao WX, Zhu J. Pretrained models: past, present and future. *AI Open*. 2021;2:225–50. <https://doi.org/10.1016/j.aiopen.2021.08.002>.
19. Liang H, Fu W, Yi F. A survey of recent advances in transfer learning. *Int Conf Commun Technol Proc, ICCT*. 2019. <https://doi.org/10.1109/ICCT46805.2019.8947072>.
20. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q. A comprehensive survey on transfer learning. *Proc IEEE*. 2021;109:43–76. <https://doi.org/10.1109/JPROC.2020.3004555>.
21. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. *Proc 30th IEEE Conf Comput Vis Pattern Recognition, CVPR*, 2017; 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>.
22. Liu, S., Deng, W.: Very deep convolutional neural network based image classification using small training sample size. In: *Proc. 3rd IAPR Asian Conf Pattern Recognition, ACPR*, 2015. 730–734 (2016). <https://doi.org/10.1109/ACPR.2015.7486599>.
23. Sudha V, Ganeshbabu TR. A convolutional neural network classifier VGG19 architecture for lesion detection and grading in diabetic retinopathy based on deep learning. *Comput Mater Contin*. 2021;66:827–42. <https://doi.org/10.32604/cmc.2020.012008>.
24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2016. <https://doi.org/10.1109/CVPR.2016.90>.
25. Saxen F, Werner P, Handrich S, Othman E, Dinges L, Al-Hamadi A. Face attribute detection with mobilenetv2 and nasnet-mobile. *Int Symp Image Signal Process Anal ISPA*. 2019. <https://doi.org/10.1109/ISPA.2019.8868585>.
26. Albahli S, Albattah W. Detection of coronavirus disease from X-ray images using deep learning and transfer learning algorithms. *J Xray Sci Technol*. 2020;28:841–50. <https://doi.org/10.3233/XST-200720>.
27. Shavit H, Jatelnicki F, Mor-Puigventós P, Kowalczyk W. From Xception to NEXception: new design decisions and neural architecture search. 1–12 (2022).
28. Bentéjac C, Csörgő A, Martínez-Muñoz G. 44. (2019). <https://doi.org/10.1007/s10462-020-09896-5>.
29. Iwama H, Okumura M, Makihara Y, Yagi Y. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans Inf Forensics Secur*. 2012;7:1511–21. <https://doi.org/10.1109/TIFS.2012.2204253>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.