



Human Gait Recognition Using Deep Learning

Prepared By

Name: Waleed M Rasheedy

ID No.: 2021204005

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

UAE

Supervisor Name: **Prof. Dr. Yasser Abdelhamid**

Collage of Informatics, Midocean University, Comoros

Email: waleed@rasheedy.com

Acknowledgements

Special thanks to **Prof. Dr. Yasser Abdelhamid** for his unwavering support, invaluable guidance, and generous patience throughout every stage of this journey. His mentorship has been a source of inspiration and strength. I am truly honored to have learned under his supervision—his deep knowledge, high standards, and sincere dedication have left a lasting mark on my academic and personal growth. I take great pride in having been one of his students.

Abstract

Human gait recognition is a biometric method that has become famous in the past few decades due to easy data acquisition techniques. In the past, different methods have been introduced by authors, but face the degradation problem due to different factors such as walking speed, clothes, noise, shadows, carrying conditions, etc. To address these problems, this work introduced new human gait recognition methods based on three models, such as Model 1, Model 2, and the combined model. Both models are based on 20 layers, but with different convolution layers and parameters. After training of both models, a combined model is designed in which the output of these models is fused and classification is performed. The proposed method is based on some steps, such as preprocessing, including resizing and augmentation. The data augmentation is based on some more steps, such as data rotation, translations, shear intensity transformations, zooming, brightness variations, and random horizontal flipping to introduce reflection invariance, and the fill mode is used to fill in pixels that were introduced through transformations. After preprocessing, both models are trained, and the output from the softmax is fused, and final classification is carried out. The method is evaluated with the help of three publicly available datasets, such as CASIA-A, CASIA-B, and CASIA-C. Using 00^0 angle of CASIA-A, 84.70% accuracy is achieved on Model 1, 98.47% on the 45^0 angle on Model 2, and on the 90^0 angles the 99.31% accuracy is attained on Model 2. In the case of CASIA-B 00^0 angle, the 68.71% is obtained on nm using Model 2, 77.49% on bg using Model 2, and 19.47% on the cl using the combined model. In the case of CASIA-C, the maximum accuracy of 55.59% is attained using a combined model. The results suggest that the proposed human gait recognition method offers promising results on CASIA-A and CASIA-B, but the performance is limited on the CASIA-C dataset.

Keywords: Human Gait Recognition, Deep Learning, CNN Model

List of Publications

No publications have been produced from this work at the time of submission.

Table of Contents

Contents

Acknowledgements	1
Abstract	2
List of Publications	3
Table of Contents	4
Contents.....	4
List of Abbreviations.....	6
List of Figures	7
List of Tables.....	8
Chapter 1 Introduction	9
1.1 Problem Definition	9
1.2 Thesis Objectives.....	10
1.3 Thesis Contribution.....	10
1.4 Thesis Organization	11
Chapter 2 Overview	12
Chapter 3 Related Work.....	14
3.1 Introduction	14
3.2 History	14
Chapter 4 Proposed Approach	20
4.1 Proposed Methodology	20
4.1.1 Dataset Preparation.....	20
4.1.2 Image Resizing and Normalization	22
4.1.3 Data Augmentation.....	22

4.1.4	Model Architectures	23
4.1.5	Model Compilation and Training.....	25
4.1.6	Score-Level Fusion Strategy	26
4.1.7	Evaluation Metrics	26
Chapter 5 Experimental Results and Discussion		27
5.1	Results and Discussion	27
5.1.1	Results of Proposed HGR Method on CASIA-A Dataset	27
5.1.2	Results of Proposed HGR Method on CASIA-B Dataset.....	33
5.1.3	Results of Proposed HGR Method on CASIA-C Dataset.....	35
5.1.4	Discussion and Analysis.....	37
Chapter 6 Conclusion and Future Work		40
References		41

List of Abbreviations

- **Acc.** : Accuracy
- **ANN** : Artificial Neural Network
- **ANOVA** : Analysis of Variance
- **BiLSTM** : Bidirectional long short-term
- **BPNN** : Back propagation neural network
- **BRNN** : Bilinear recurrent neural network
- **CNN** : Convolution neural networks
- **CRC** : Collaborative Representation Classification
- **CV** : Computer Vision
- **GaitPT** : Gait Pyramid Transformer
- **GSI** : Gait Salient Image
- **HGR.** : Human Gait Recognition
- **ML** : Machine Learning
- **ROC** : Receiver Operating Characteristic
- **ViTs** : Vision Transformers
- **XAI** : Explainable AI

List of Figures

Figure 1: Architecture of Proposed Human Gait Recognition System.....	20
Figure 2: Sample Images for CASIA-A [62], CASIA-B [63], and CASIA-C [64]	22
Figure 3: Architecture of CNN Model 1	24
Figure 4: Architecture of CNN Model 2	25
Figure 5 Confusion Matrix for 0 ⁰ angle of CASIA-A using Combined Model	29
Figure 6 Confusion Matrix for 45 angle of CASIA-A using Combined Model.....	31
Figure 7 Confusion Matrix for 90 angle of CASIA-A using Combined Model.....	32
Figure 8 Confusion Matrix for 00 angle of CASIA-B on Experiment 1 using Combined Model.....	33
Figure 9 Confusion Matrix for 00 angle of CASIA-B on Experiment 2 using Combined Model.....	34
Figure 10 Confusion Matrix for 00 angle of CASIA-B on Experiment 3 using Combined Model.....	35
Figure 11 Confusion Matrix for CASIA-C using Combined Model	36

List of Tables

Table 1 Subject Level Classification Results for 0 ⁰ angle of CASIA-A	28
Table 2 Subject Level Classification Results for 45 angle of CASIA-A	30
Table 3 Subject Level Classification Results for 90 angle of CASIA-A	31
Table 4 Detailed Comparison of the Proposed Models	38

Chapter 1 Introduction

1.1 Problem Definition

Several biometric authentication methods are being used for person identification. These methods can be divided into physical or behavioural categories [1]. The face, fingerprints, palm, and ears lie under the physical biometric technique, while in the behavioural category, voice, signature, and stride pattern are considered [2]. In the case of gait, the walking patterns are used without the consent of the subject under consideration [3]. The gait is considered appealing because it is comprised of 24 distinct components, such as angles and displacement of joints of the human body. The gait method is being widely used to ensure the security [4].

However, the majority of techniques are based on lateral-view data of human silhouettes [5] features extraction, which makes it difficult due to a data collection requirement. This problem occurs in the real-time indoor surveillance environment, where a large sequence of data collection is not feasible [6]. Furthermore, the cameras installed for security and surveillance are mostly installed in corridors, exits, or narrow places in which only a frontal view is captured [7]. Therefore, the frontal view gait analysis is preferred to handle this problem.

The methods introduced by researchers for human gait recognition (HGR) but faced the degradation of performance due to different factors such as clothing variations, view variations, and different carrying conditions. Some techniques perform well on a specific dataset, but their performance is minimized on other datasets, so the generalization is not achieved [4, 8]. Many methods are based on multiple streams and perform well, but their computational cost is high, so they cannot be used in real-world environments [9]. In some methods feature selection approach is employed, but fails due to relevant information loss [3, 10].

Furthermore, the methods that use feature optimization and data augmentation are less flexible [11]. This research aims to design an HGR method that overcomes these problems and offers efficiency, adaptivity, generalizability, and flexibility.

1.2 Thesis Objectives

1. To expand the feature extraction and combination of methods with the help of spatial and temporal features, and also ensure the recognition rate and computational efficiency.
2. To develop a strong HGR method that effectively overcomes the problems of variation in angle, clothing, and background variability, and also ensures generalizability to improve the system performance.
3. To evaluate the efficacy of the single and multi-stream methods by extracting such features that are discriminative, conducting analysis to adapt the trade-offs in terms of performance, flexibility, and computational cost of the HGR.
4. To inspect the impression of feature selection and optimization methods using entropy or skewness-based selection of features on the recognition rate and stability of the HGR method, by eliminating the feature loss and improving results.
5. To carry out a comparison of the HGR method performance utilizing multiple datasets, such as CASIA-A, B, and C, to ensure its generalizability and robustness in real-world environments.

These objectives will offer such an HGR method that is generalized. Optimized, robust, and adaptable.

1.3 Thesis Contribution

- This work introduces two distinct deep learning models, each comprising 20 layers but designed with different convolutional structures and parameters. These models are tailored to learn complementary gait features, improving recognition accuracy.
- A combined model is proposed that fuses the softmax outputs of both individual models at the score level. This ensemble approach leverages the strengths of both models to enhance the final classification performance.
- An extensive augmentation strategy is applied during preprocessing to increase dataset diversity and improve model generalization. Techniques include rotation, translation, shear, zoom, brightness adjustments, and horizontal flipping with fill mode.

- The method is rigorously evaluated on three standard benchmark datasets—CASIA-A, CASIA-B, and CASIA-C, under varying conditions (e.g., angles and scenarios), demonstrating its effectiveness and adaptability.
- The proposed models achieve high recognition accuracy on difficult view angles (e.g., 45 and 90 in CASIA-A), showing robustness against view variations in gait patterns.
- The work also highlights challenges associated with the CASIA-C dataset, providing a valuable reference for future research aimed at improving performance in complex and noisy environments.

1.4 Thesis Organization

In this work, a new HGR method is proposed to address the covariant problem. The work is mainly based on six chapters, such as introduction, overview, related work, proposed methodology, results and analysis, and conclusion & future work. All the chapters are explained in detail, such as Chapter 3, related work offers extensive details about current studies. Moreover, in Chapter 4, a detailed explanation of the proposed HGR method is given. After that, the results and analysis are offered in Chapter 5. Finally, conclusion and future work are given in Chapter 6.

Chapter 2 Overview

Gait recognition has captured much attention in computer vision due to its potential in data capturing using non-invasive methods [12]. Researchers have made a significant effort to develop a practical application of gait recognition in the past few decades. However, the majority of techniques are based on lateral-view data of human silhouettes [5] feature extraction, which makes it difficult due to the data collection requirement. This problem occurs in the real-time indoor surveillance environment, where a large sequence of data collection is not feasible [6]. Furthermore, the cameras installed for security and surveillance are mostly installed in corridors, exits, or narrow places in which only a frontal view is captured [7]. Therefore, the frontal view gait analysis is preferred to handle this problem.

These methods addressed the problem of frontal-view recognition at some level, but there is a problem of low accuracy due to the environmental factor and complicated background [13]. Moreover, an extra layer of computational cost also makes it impractical due to the usage of camera depth and radar sensor.

Despite of importance, robustness, and efficiency of gait recognition, there are still problems that interfere with the authentication and recognition of an individual. These factors include variations of view angles, clothing, load carriage, walking speed, shadow, etc [14]. These problems reduce the recognition rate of a gait recognition system. To address these problems, several gait recognition methods are implemented that can be categorized as model-free and model-based.

The model-based approach aims to obtain the characteristics of human stride, which uses the human body [15]. Mostly, this technique faces the problem of a lack of a reliable approximation of the human body. However, this method needs high-quality images, which makes it a challenging task. The model-based method can further be classified as pose-based and sensor-based. The model-free approach, on the other hand, focuses on the extraction of human silhouettes directly from the human body [16].

The gait recognition tasks can be implemented with the help of two broadly used methods, including handcrafted feature extraction [17] and deep learning-based features [18]. The hand-crafted features are referred to as the classical methods and are inaccurate and time-consuming. The deep learning-based methods are accurate and do not require human effort for feature

extraction and automatically obtain features from the input data. The deep learning methods for gait recognition introduce new approaches to process the data and address the covariant problem, such as clothing, carrying, and walking variation. However, these deep learning models reflect the attributes of a black box nature and may be problematic if not properly fine-tuned and need information about the pre-training conditions and parameters. To overcome these problems, there is a demand for robust, accurate, and efficient gait recognition methods that process the data correctly.

Chapter 3 Related Work

3.1 Introduction

Gait recognition has emerged as a crucial biometric identification method, evolving from traditional handcrafted feature-based approaches to modern deep learning techniques. Early methods relied on manually extracted features, such as silhouette and motion descriptors, which often suffered from variations in clothing, carrying conditions, and occlusions. With advancements in deep learning, models like Convolutional Neural Networks (CNNs), Bidirectional LSTM (BiLSTM), and attention mechanisms have significantly improved feature extraction and recognition accuracy. Additionally, multi-view and multi-modality approaches have enhanced gait recognition in challenging environments. However, despite these advancements, various challenges remain, including high computational costs, sensitivity to noise, and dependency on large datasets for training.

Many machine learning (ML) and computer vision (CV) approaches have been previously introduced to enhance gait-based recognition. Typically, a conventional HGR system involves steps such as data preparation, segmentation, feature extraction, and final classification. This chapter highlights a variety of contemporary HGR techniques developed in recent research.

3.2 History

ML and CV approaches have been previously introduced to enhance gait-based recognition. Typically, a conventional HGR system involves steps such as data preparation, segmentation, feature extraction, and final classification. This chapter highlights a variety of contemporary HGR techniques developed in recent research.

3.2.1 Data Preprocessing

Data preprocessing is a basic phase performed before any further operation. It involves background subtraction, noise filtering, and contrast enhancement to normalize input data. The quality of preprocessing directly affects the system's performance, as poor data quality can degrade recognition outcomes. In [19], the authors presented an HGR approach using the Quartile Deviation Normal Distribution for feature extraction, followed by Bayesian fusion

and binomial distribution-based optimization. Their pipeline includes preprocessing via color transformation, background subtraction for ROI detection, feature extraction using LBP and HOG, entropy-based optimization, and final feature selection through a binomial distribution method. MSVM was employed for classification, achieving 98.8%, 87.7%, 91.6%, and 100% accuracy on CASIA-A, CASIA-B, CASIA-C, and AVAMVG, respectively.

In [20], an image segmentation-based method is proposed that starts with HSV color enhancement, followed by channel selection and human subject extraction. A codebook-based technique is then used to fuse features, with Euclidean distance used to choose the most relevant ones. MSVM is again used for classification, achieving 98.6%, 93.5%, and 97.3% on CASIA-A, B, and C.

Li et al. [21] proposed a two-module architecture: the first module uses Gait Energy Images (GEIs) with PCA, LDA, and spatial-domain techniques; the second employs collaborative representation classification (CRC), yielding 83.1% and 86.67% accuracy on CASIA-B and CMU-MoBo. Similarly, ANN-based methods [22] transform videos into frames, apply background subtraction, and extract morphological skeletons for tracking, resulting in superior performance on CASIA-A.

Sugandhi et al. [23] used overlapping cycles with pixel-based horizontal line analysis, employing skeletonized frames to track movement phases. In [24], fuzzy set theory and entropy were used to handle uncertain gait features, with view variations addressed via a size normalization strategy, leading to 93.21% accuracy on CASIA-B.

A frontal view gait representation method called LFHEI based on the spatiotemporal features, is implemented [25]. The method uses dynamic parts of the body and offers robust features. The LFHEI feature descriptor is passed to the deep learning model for feature extraction, and gait recognition is performed using nearest neighbours. The CASIA dataset was used to test the method and obtained 95.0% accuracy on CASIA-A and 89.9% on CASIA-B. the method is also compared with the recent method and shows 4% improved results.

3.2.2 Feature Extraction

Feature extraction aims to distill meaningful data while filtering out redundancy. This can involve classical descriptors or deep learning methods. In [26], CNNs are used to extract high-

level gait features from optical flow, achieving 88.9% on TUM-GAID. Alotaibi et al. [27] built an 8-layer CNN to tackle data scarcity and occlusion, reaching 90.43% on CASIA-B.

Li et al. [27] introduced DeepGait, a CNN architecture based on a modified VGG-D and PCA for dimensionality reduction, verified via a Joint Bayesian model, and tested on OULP with 89.3% accuracy. A machine learning approach in [28] addressed high-dimensional feature issues from GEI concatenation, achieving 60.5% and 90% on CASIA-B and OU-ISIR.

In [29], a template approach using static and dynamic silhouettes is evaluated on SOTON and CASIA-B, scoring 88.9% and 84.6%. Htun et al. [30] employed SURF features with GLCM and a supervised meta-sample classifier, attaining 77.3% (nm), 67.2% (bg), and 73.2% (cl) on CASIA-B.

Zhang et al. [16] used RGB input for CNN-based pose and appearance features, classified via LSTM with results of 81.8% (CASIA-B), 99.5% (USF), and 87.8% (FVG). The method in [31] applied deterministic learning theory on Kinect data, using spatial-temporal kinematic parameters, and reached 97.7%.

A new deep-learning model based on the attention mechanism is introduced to minimize the variability in the gait data [32]. Initially, the data is obtained in the form of gait energy images based on multi-recurrent features. Secondly, the deep learning model is used to obtain the relevant features from input data. they used attention-based blocks to highlight the important gait features for improved performance. The temporal relation of frames is obtained using the Bi-GRU layer. The method testing is carried out with the help of CASIA-B and OULP gait, achieving an accuracy of 93.4% and 97.4%.

To assess the contribution of important gait features in pattern recognition, a new method is presented for gait recognition that utilizes metaheuristic optimization for feature selection [33]. Initially, the features are extracted from the input data, and optimization is performed afterward with the help of the metaheuristic method. The optimized features are passed to various classification models for recognition. The method was tested with the help of 800 individual data points and obtained 99.81% accuracy. The comparison of the method is also performed with the recent feature selection method, and this method offered improved results.

To obtain spatiotemporal information from the contour of the human body, a gait recognition method used on the GaitSet model is presented [34]. This method aims to improve the recognition rate of video data based on the multimodality of silhouette and gait energy images. The method is

tested with the help of CASIA-B and attained 95.8% on NM, 89.3% on BG, and 72.5% on CL, which is improved in contrast to recent gait recognition methods.

A new gait recognition algorithm is designed for indoor data called GaitTS, which is based on spatiotemporal features of different scales [35]. The method aims to address the problem of lacking attention to the features of data. The GaitTS aggregates temporal features of multiple frames. The spatial features are divided into different maps having several sizes to give attention to different parts of the human body. The method obtained an accuracy of 97.9% on NM, 95.4% on BG, and 85.7% on CL respectively. An average accuracy of 76.50% is obtained on occCASIA-B and 93.0% on CCPG.

By employing a deformable convolution, a gait recognition method called DCNGait is which uses the temporal dimension [36]. The deformable convolution aims to enhance the temporal dynamics representation. The proposed TD3D convolution is capable of obtaining diverse features using temporal patterns. Careful feature extraction is performed by designing spatial and temporal blocks. The method DCNGait obtained 93.70% accuracy on CASIA-B.

3.2.3 Feature Selection

Feature selection refines the extracted data to reduce computation and improve accuracy by focusing only on informative features. In [37], view-invariant gait recognition is achieved using AESI with PCA and MDA, reporting 91.47% (CASIA-B), 72.7% (OU ISIR), and 72.53% (USF).

Wang et al. [38] employed Gabor wavelets with 2D PCA and SVM, obtaining better results on CASIA datasets. In [39], statistical features were extracted with modified LDA and Maximum Margin Criterion to reduce dimensionality, achieving 70.3% on CMU-MoCap.

The Gait Salient Image (GSI) template in [40] aggregates energy-encoded gait cycles, processed via PCA+LDA, with Rank-1 and Rank-5 accuracies of 58.44% and 76.60% (USF). In [41], Haralick features from GEIs are optimized via the RELIEF algorithm and classified using KNN, with 80% accuracy on CASIA-B.

A multi-layer autoencoder approach [42] was used to address variation, scoring 63.90% and 97% (with and without variation) on CASIA-B and SZU RGB-D. Pose-based data was used

by Liao et al. [43] in a PTSN model, achieving 96.92% (nm), 85.78% (bg), and 68.11% (cl) on CASIA-B.

The method in [44] extracted trajectory and motion features using HOG and MBH, achieving high accuracy across CMU MoBo (98.5%), CASIA-B (96.6%), NLPR (100%), CASIA-C (99.8%), and TUM (96.5%). A wavelet and entropy-based method [45] scored 79.32% (OU-ISIR) and 78.44% (CASIA-B). A 3D skeleton-based model [46] incorporating Kinect data yielded 92.3% accuracy.

Arshad et al. [47] combined VGG19 and AlexNet for feature extraction, with further refinement using Fuzzy Entropy Controlled Skewness and ML classifiers, resulting in 99.8% (AVAMVG), 99.7% (CASIA-A), 93.3% (CASIA-B), and 92.2% (CASIA-C).

A gait recognition method to acquire cooperative motion of the human exoskeleton is designed based on the feature extraction of the deep learning model [48]. Initially, data is acquired with the help of a lightweight device using three sensors. Different conditions are considered to capture the data, such as walking on stairs, sloping, and being still. A Bi-LSTM model with an attention mechanism is used to obtain features from input data. The method was tested on 27 classes and achieved improved results.

3.2.4 Classification

The final step involves passing the selected feature vectors to classification models. Ben et al. [49] applied tensorial structure analysis with Gabor-based features and nearest neighbor classification, achieving up to 99% accuracy on CASIA-B and 100% on OU ISIR.

In [50], an ensemble learning-based CVGR-EL method combined segmentation, AAD-based feature extraction, HMM training, and ensemble classification, obtaining 96.1% (CASIA) and 97.1% (OU ISIR). Li et al. [51] utilized Joint Bayesian modeling with GEI input and achieved 94.9% on OULP.

Track motion descriptors were used in [52] to extract dense features, resulting in 86.7% (CASIA-B), 95.2% (CASIA-C), 99.8% (TUM-GAID), and more. Wavelet-based signal processing in [53] achieved 93.44% on OU-ISIR.

A CNN trained on optical flow [54] showed Rank-1 and Rank-5 scores of 72.06% and 84.07% (CASIA-B) and 96.45% and 99.24% (TUM-GAID). Codebook and Fisher vector techniques [55] reached 100% and 97.74% (CASIA-A, TUM-GAID).

Silhouette-based age classification [56] assessed attributes like posture and stride, scoring 74.47% on OU-ISIR. Footwear-based gait changes [57] were evaluated using signal dynamics and ground force analysis, reaching 99% accuracy. A posture-based system [58] used Center-of-Body and frame displacement to achieve more than 90% accuracy on CASIA-B.

Mahfouf et al. [59] proposed a tracking and ML classification method with 95% accuracy on CASIA-B, while addressing multi-person tracking in a frame using latent structure SVM, achieving 91.53%.

A new gait recognition technique called Stacked-Gait [60] is designed that is based on several steps, including image resizing, image conversion to grayscale, training, and stacking of autoencoders, and classification using machine learning algorithms. The Stacked-Gait was tested using CASIA-B and attained an average accuracy of 98.18% using CASIA-B.

The effectiveness of deep learning models, including CNN, MLP, SOMs, and transfer learning with EfficientNet, is used for gait recognition [61]. These models are utilized for robust feature extraction from the image data. They used 92,596 images of a 64×64 size obtained from 10 individual sequences. Furthermore, a modified deep learning model based on a dense classification layer is also implemented, which obtained 97.12% on the CASIA-B dataset.

Chapter 4 Proposed Approach

4.1 Proposed Methodology

This section presents a detailed description of the proposed methodology employed to build a deep learning-based multi-class classification system. The system is designed to recognize a person using the gait features with the help of two custom-designed CNN architectures. To enhance robustness and accuracy, a score-level fusion strategy is applied that combines predictions from both models. The implementation was carried out using TensorFlow and Keras libraries in Python. The architecture of the proposed system is shown in **Figure 1**.

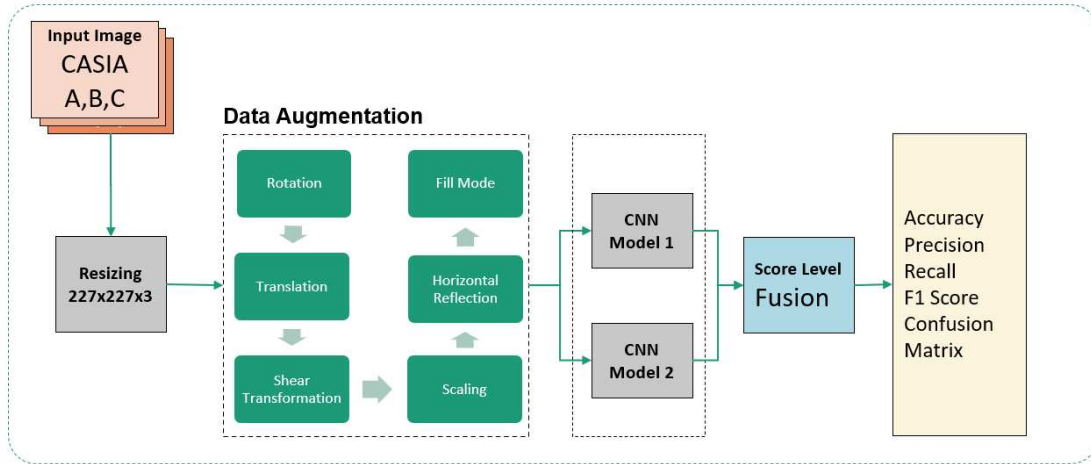


Figure 1: Architecture of Proposed Human Gait Recognition System

4.1.1 Dataset Preparation

This study utilizes three publicly available datasets, CASIA-A [62], CASIA-B [63], and CASIA-C [64]. CASIA-A is one of the gait datasets containing gait sequences of **20 individuals**, captured from **three different angles, such as 0°, 45°, and 90°**. In this dataset, the images are extracted from videos. CASIA-A was used to evaluate the model's ability to recognize gait patterns under a smaller sample size and **limited-view conditions**. All images were resized to 227×227 and normalized. Moreover, augmentations were applied similarly to enhance robustness.

The CASIA-B is a multi-view gait dataset containing gait sequences of 124 subjects recorded from 11 different view angles (0° to 180°) under three conditions: normal walking (nm), walking with a bag (bg), and walking with a coat (cl). The CASIA-B is the primary dataset for training and evaluation due to its large class size and diverse variations in view and appearance. In the training set, the normal walk sequences (e.g., nm-01, nm-02) are included. While in the test set, walk sequences nm-03 to nm-04, bg-01 to bg-02, and cl-01 to cl-02 are included. The images extracted from each video sequence were arranged in directories per subject. In this study, the frontal view analysis is performed to assess the model performance for variations in clothing and carried objects, making this dataset ideal to evaluate the generalization capability of CNNs.

The CASIA-C contains **infrared gait sequences** of 153 individuals recorded at night with four variations such as normal walk, slow walk, walking with a bag, and fast walk. Unlike CASIA-A and CASIA-B, which use visible spectrum images, CASIA-C utilizes **thermal imaging**, adding an **infrared modality** dimension. This dataset is used to validate the **cross-modal generalization** of the trained CNN models. Since the modality is different, grayscale IR images were normalized and resized, and passed through the same preprocessing pipeline as other datasets. The sample images for CASIA-A, B, and C datasets are given in **Figure 2**.

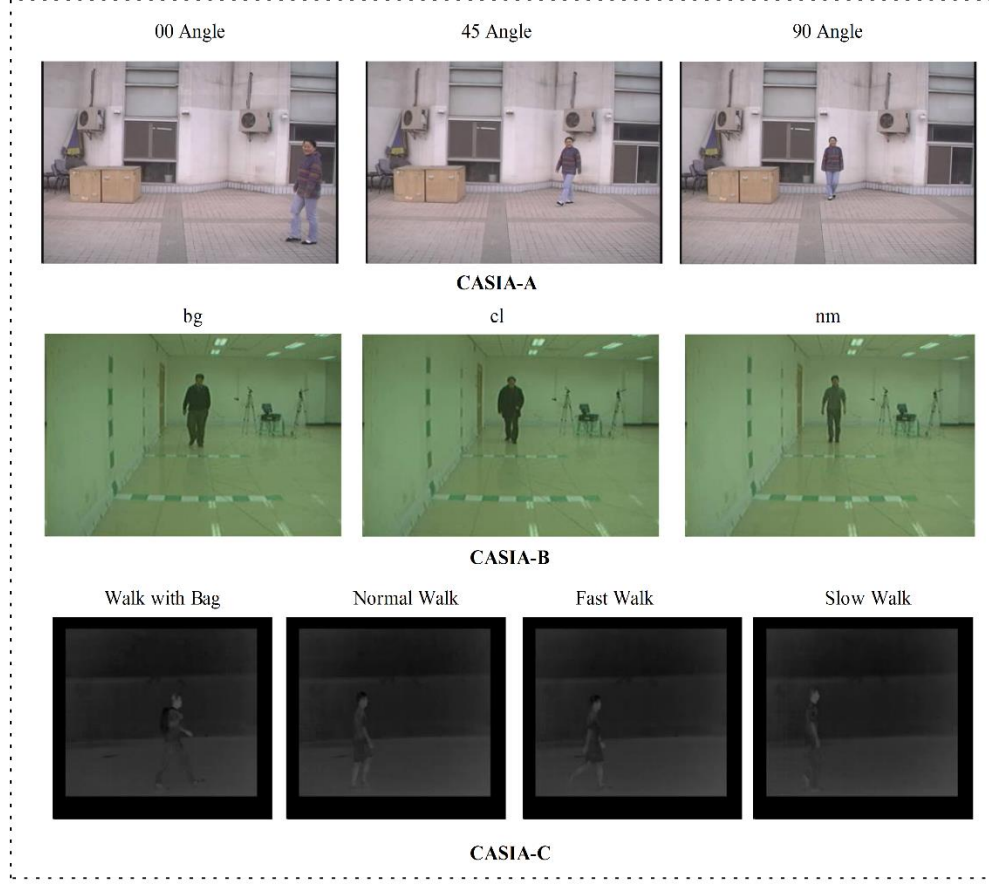


Figure 2: Sample Images for CASIA-A [62], CASIA-B [63], and CASIA-C [64]

4.1.2 Image Resizing and Normalization

All input images were resized to a uniform shape of 227×227 pixels with three RGB channels to ensure compatibility with the input layer of the CNNs. Furthermore, images were normalized by scaling pixel intensities to the range $[0, 1]$. This preprocessing step helps accelerate convergence during training and prevents issues related to numerical instability.

4.1.3 Data Augmentation

To mitigate the effects of overfitting and enhance model generalization, a robust data augmentation pipeline was applied using ImageDataGenerator. The augmentations applied include: data Rotation Range ± 10 degrees to simulate variations in orientation, width and height Shifts up to 10% to introduce minor translations, shear intensity transformations up to 10 degrees to simulate geometric

distortions, zooming between 85 -115% to simulate scale variations, brightness range between 0.8 and 1.2 to simulate lighting conditions, random horizontal flipping to introduce reflection invariance and fill mode used to fill in pixels that were introduced through transformations. These augmentation strategies help the network learn from a more diverse set of examples, which enhances its ability to generalize to unseen data.

This proposed preprocessing pipeline ensures that the model is exposed to a wide range of simulated real-world distortions. These shifts thereby enhance its **invariance to spatial shifts, viewpoint changes, lighting variations, and geometric distortions**. These transformations make the model focus on essential gait patterns rather than overfitting to background noise or specific image conditions.

4.1.4 Model Architectures

Two CNN architectures were designed and trained independently. Both models use similar components (convolutional, batch normalization, pooling, dropout) but differ in depth and number of filters. The Model 1 is a lightweight CNN that consists of three convolutional blocks with increasing complexity. The architectural configuration includes an input layer of $227 \times 227 \times 3$, Conv2D (32 filters, 5×5 kernel) + L2 regularization, BatchNormalization + ReLU Activation, MaxPooling2D (2×2) + Dropout (25%), Conv2D (64 filters, 3×3) + same sequence, Conv2D (128 filters, 3×3) + Dropout (50%), Flatten Layer, and Dense Layer (Softmax) as shown in Figure 3. The size of the Dense layer depends on the number of classes in the dataset. Moreover, L2 regularization was added to each convolutional layer with a factor of 0.001 to penalize complex models and avoid overfitting.

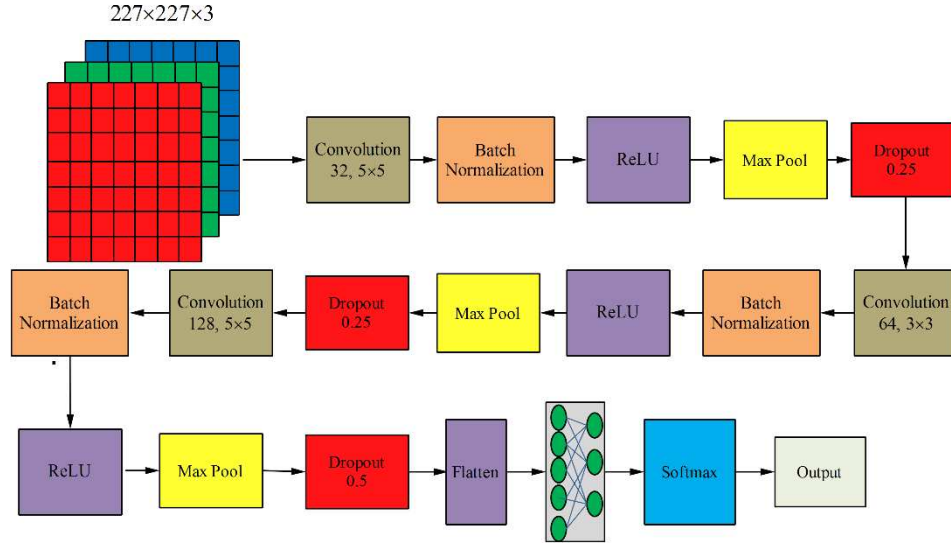


Figure 3: Architecture of CNN Model 1

The Model 2 is a deeper CNN model as compared to Model 2, which consists of three convolutional blocks with increasing complexity. The architectural configuration includes an input layer of $227 \times 227 \times 3$, Conv2D (64 filters, 5×5 kernel) + L2 regularization, BatchNormalization + ReLU Activation, MaxPooling2D (2×2) + Dropout (25%), Conv2D (128 filters, 3×3) + same sequence, Conv2D (1256 filters, 3×3) + Dropout (50%), Flatten Layer, and Dense Layer (Softmax) as shown in **Figure 4**. The Dense layer of Model 2 also depends on the number of classes in the dataset. Moreover, L2 regularization was added to each convolutional layer with a factor of 0.001 to

improve the training of complex models and avoid overfitting. Model 2 is expected to extract more complex and abstract features due to its increased depth and capacity.

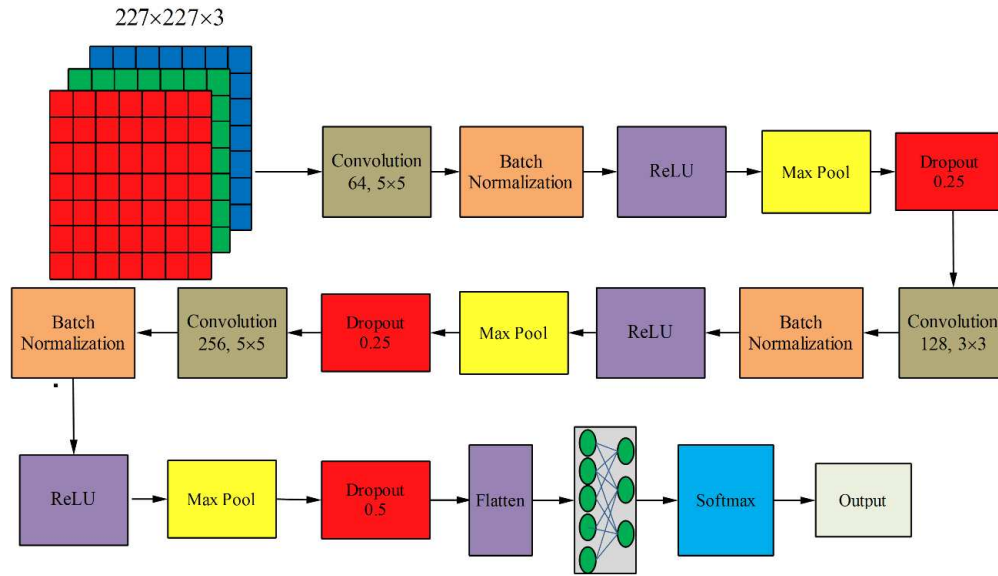


Figure 4: Architecture of CNN Model 2

4.1.5 Model Compilation and Training

Both models were compiled using the Adam optimizer with a learning rate of 0.0003, which provides adaptive learning rates and works well for sparse gradients. The loss function used was categorical cross-entropy, as this is a multi-class classification problem with softmax output. Moreover, to improve training efficiency and prevent overfitting, the following callbacks were used:

- **EarlyStopping**: Monitors training loss and halts training if no improvement is observed for 10 consecutive epochs.
- **ModelCheckpoint**: Saves the best model weights based on the lowest training loss (model1_best.h5, model2_best.h5).

Each model was trained for up to 100 epochs on the training dataset using a batch size of 64. Due to limited validation data, evaluation metrics were later computed using fused predictions on training images.

4.1.6 Score-Level Fusion Strategy

A score-level fusion strategy was employed to enhance classification performance and reduce reliance on a single model's prediction. The idea is to combine the softmax outputs (probability distributions) of both models. Where 0.7 assigned as weight for Model 1 due to its better convergence and 0.3 for Model 2 due to lower performance as compared to Model 1. Where the final label is obtained using `argmax(fused_score)` by passing the resultant fused score. This fusion allows leveraging the strengths of both models, where one model may correct errors made by the other.

4.1.7 Evaluation Metrics

After fusion, predictions were evaluated using the following metrics, including the confusion matrix, which shows true vs predicted class distributions, accuracy denotes the Percentage of correct predictions, the Classification Report includes precision, recall, and F1-score for all classes, and a visual representation of the confusion matrix using Seaborn for pattern analysis. The evaluation revealed how the fusion strategy improved classification accuracy by reducing errors from individual models.

Chapter 5 Experimental Results and Discussion

5.1 Results and Discussion

This section offers the detailed results obtained using the proposed HGR method on three publicly available datasets, CASIA-A, CASIA-B, and CASIA-C. Both CNN models are trained by setting hyperparameters, including the learning rate 0.0003, batch size of 64, Adam optimizer, and 100 epochs. Moreover, during training, cross-entropy loss is computed for both models. The experiments are conducted on Jupyter Notebook using a personal computer with an intel Core Ultra 9 275HX CPU, 64 GB of RAM, 2 TB of SSD, and an NVIDIA RTX 5070 Ti 12 GB of GPU.. The detailed results are given in the following sections.

5.1.1 Results of Proposed HGR Method on CASIA-A Dataset

The proposed HGR method is evaluated on CASIA-A for all three angles, such as 00^0 , 45^0 , and 90^0 . Each angle is employed to conduct three types of experiments: Model 1, Model 2, and the combined model. The results of the combined model are based on the outcomes of Model 1 and Model 2.

For CASIA-A 00^0 , the Model 1 reached the training accuracy of 99%, training loss of 0.0021, test accuracy of 84.70%, and test loss of 1.3358. For the Model 2, on the same parameters, the training accuracy of 98.80%, the training loss of 0.0715, the test accuracy of 77.14%, and the test loss of 1.8350 are attained. For the combined model, various evaluation parameters are computed, including precision, recall, F1-Score, and accuracy, and attained average values are 91.35%, 91.0%, 88.30%, and 85.0% respectively. A detailed subject-level recognition is also given in the **Table 1**.

Table 1 Subject Level Classification Results for 0° angle of CASIA-A

Class	Precision	Recall	F1-Score
fyc	1.00	0.59	0.74
hy	1.00	1.00	1.00
ljg	1.00	0.26	0.41
lqf	0.91	1.00	0.95
lsl	0.47	1.00	0.64
ml	1.00	1.00	1.00
nhz	1.00	0.67	0.80
rj	1.00	1.00	1.00
syj	1.00	1.00	1.00
wl	1.00	1.00	1.00
wq	0.92	1.00	0.96
wyc	1.00	0.19	0.33
xch	0.51	1.00	0.68
xxj	1.00	0.49	0.66
yjf	0.76	1.00	0.86
zc	1.00	1.00	1.00
zdx	1.00	1.00	1.00
zjg	1.00	1.00	1.00
zl	1.00	1.00	1.00
zyf	0.70	1.00	0.83

Moreover, the confusion matrix for the combined model by employing CASIA-A 0° is also computed as shown in the **Figure 5**.

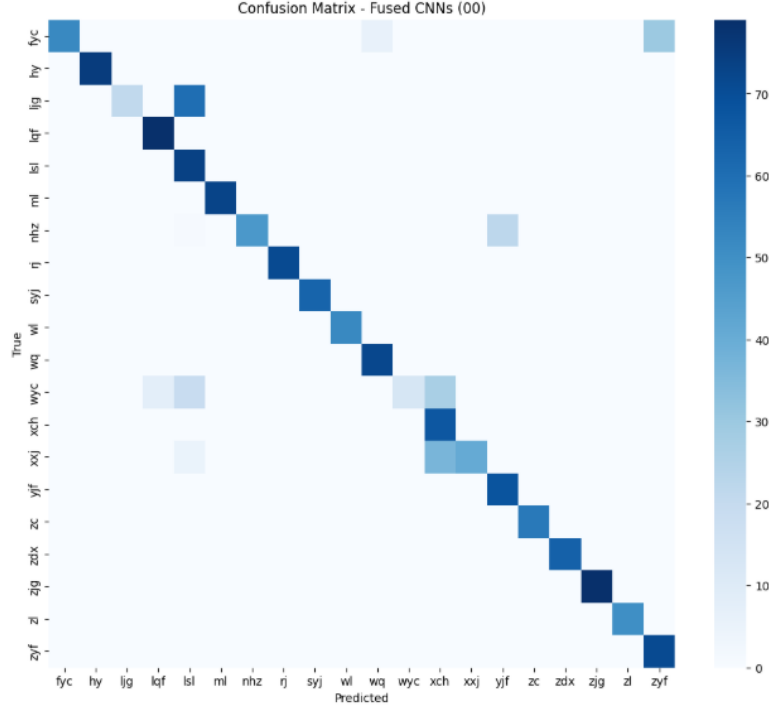


Figure 5 Confusion Matrix for 0^0 angle of CASIA-A using Combined Model

For 45^0 of the CASIA-A dataset, the same experimental protocols are used on Model 1, Model 2, and the combined model. For Model 1 on 45^0 of the CASIA-A, the attained results are training accuracy of 99.45%, training loss of 0.0471, test accuracy of 11.32%, and test loss of 252.4321 for Model 1. For Model 2, the attained values for training accuracy, training loss, test accuracy, and test loss are 99.93%, 0.0013, 98.47%, and 0.0462, respectively. While using the combined model, the average achieved values for precision, recall, F1-Score, and accuracy are 45.05%, 29.60%, 28.75%, and 29%. The detailed subject-level classification results are given in the **Table 2**.

Table 2 Subject Level Classification Results for 45 angle of CASIA-A

Class	Precision	Recall	F1-Score
fyc	1.00	0.37	0.54
hy	0.07	1.00	0.13
ljg	0.00	0.00	0.00
lqf	0.00	0.00	0.00
lsl	1.00	0.22	0.36
ml	1.00	0.15	0.26
nhz	1.00	0.02	0.05
rj	0.00	0.00	0.00
syj	1.00	0.39	0.56
wl	1.00	1.00	1.00
wq	0.00	0.00	0.00
wyc	0.00	0.00	0.00
xch	0.00	0.00	0.00
xxj	0.00	0.00	0.00
yjf	0.00	0.00	0.00
zc	0.94	1.00	0.97
zdx	0.00	0.00	0.00
zjg	1.00	0.85	0.92
zl	1.00	0.92	0.96
zyf	0.00	0.00	0.00

Furthermore, the confusion matrix for the combined model by employing CASIA-A 45⁰ is also computed as shown in the **Figure 6**.

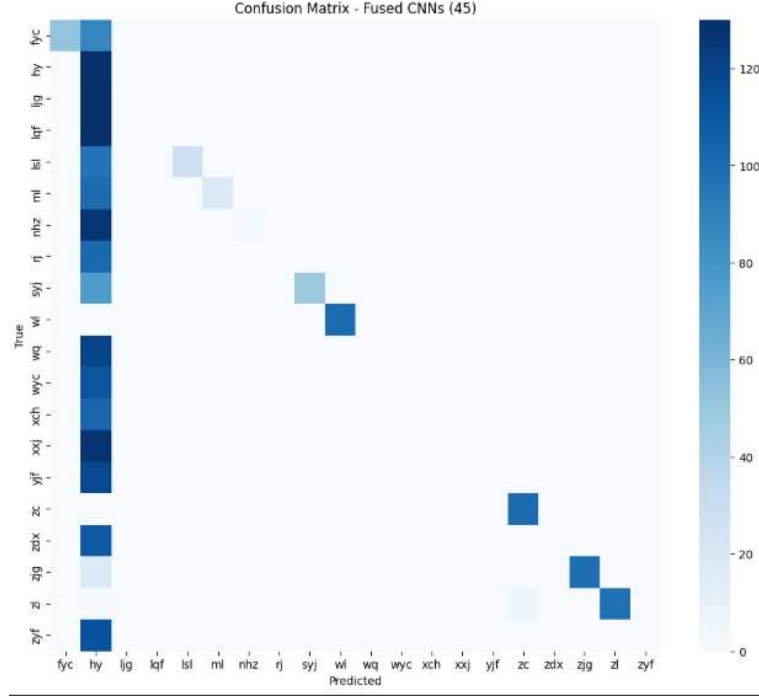


Figure 6 Confusion Matrix for 45 angle of CASIA-A using Combined Model

The same procedure is adopted for experiments on 90^0 of the CASIA-A using Model 1, Model 2, and the combined model. The achieved value for training accuracy is 99.81%, training loss is 0.0180, test accuracy is 70.52%, and test loss is 11.7668, respectively, using Model 1. In the case of Model 2, the attained values are 99.87%, 0.0056, 0.0286, and 99.31% for training accuracy, training loss, test accuracy, and test loss, respectively. The results are also computed for the combined model by employing 90^0 angle of the CASIA-A dataset and an average precision of 96.60%, recall of 93.40%, F1-score of 92.75% and accuracy of 89% is attained. A detailed classification results based on subject-level are given in the **Table 3**.

Table 3 Subject Level Classification Results for 90 angle of CASIA-A

Class	Precision	Recall	F1-Score
fyc	1.00	1.00	1.00
hy	1.00	1.00	1.00
lfg	1.00	1.00	1.00
lqf	0.32	1.00	0.48
lsl	1.00	0.78	0.87
ml	1.00	1.00	1.00
nhz	1.00	1.00	1.00

rj	1.00	1.00	1.00
syj	1.00	1.00	1.00
wl	1.00	1.00	1.00
wq	1.00	0.68	0.81
wyc	1.00	0.06	0.11
xch	1.00	1.00	1.00
xxj	1.00	0.16	0.28
yjf	1.00	1.00	1.00
zc	1.00	1.00	1.00
zdx	1.00	1.00	1.00
zjg	1.00	1.00	1.00
zl	1.00	1.00	1.00
zyf	1.00	1.00	1.00

The confusion matrix for the combined model is also calculated by employing CASIA-A 90⁰ as shown in the **Figure 7**.

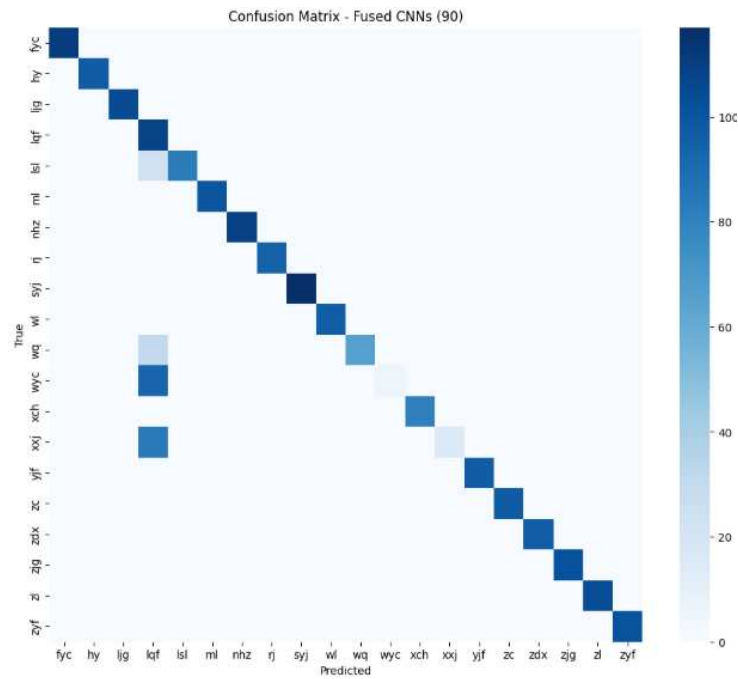


Figure 7 Confusion Matrix for 90 angle of CASIA-A using Combined Model

5.1.2 Results of Proposed HGR Method on CASIA-B Dataset

Using the CASIA-B dataset, the frontal angle (00^0) is used to assess the proposed HGR method's performance. The frontal analysis is conducted using three experiments, such as setting the training set as (e.g., nm-01, nm-02), are included. While in the test set, walk sequences nm-03 to nm-04, bg-01 to bg-02, and cl-01 to cl-02 are included. Model 1, Model 2, and the combined model are evaluated on these three sets, and training accuracy, training loss, test accuracy, and test loss are computed, respectively.

In the first experiment, nm-01, nm-02 are considered as the training set, and nm-03 to nm-04 as the test set. All three models, such as Model 1, Model 2, and the combined model, are used to test the performance. By employing Model 1, the attained values are 96.50%, 0.1508, 21.96%, and 0.1508 for training accuracy, training loss, test accuracy, and test loss, respectively. While using the Model 2, the obtained values for training accuracy, training loss, test accuracy, and test loss are 97.17%, 0.0888, 68.71%, and 1.845, respectively. In the case of the combined model, the obtained accuracy is 47.31%, average precision of 68.40%, recall of 44.50%, and F1-score of 48.80%, respectively. The results are also verified by computing a confusion matrix as shown in the **Figure 8**.

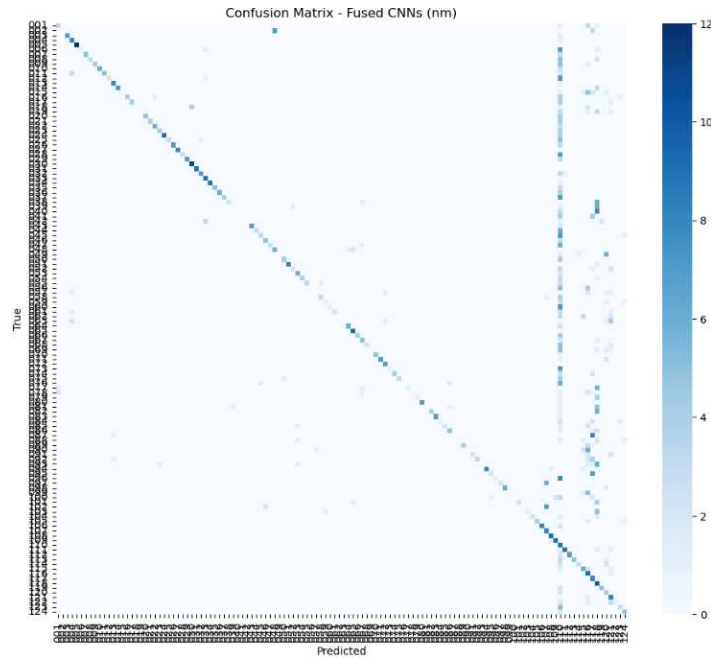


Figure 8 Confusion Matrix for 00 angle of CASIA-B on Experiment 1 using Combined Model

In the second experiment, using the gallery set as nm-01, nm-02, and the test as bg-01 to bg-02, Model 1 attained the training accuracy of 97.17%, training loss of 0.1011, test accuracy is 10.16%, and test loss is 23.7638, respectively. Using Model 1, the achieved values of training accuracy, training loss, test accuracy, and test loss are 97.65%, 0.1025, 77.49%, and 1.1108, respectively. While on the combined model, the achieved accuracy of 57.37%, average precision of 70.80%, average recall of 57.40%, and average F1-score of 58.70%, respectively.

The results are also verified by computing a confusion matrix as shown in the **Figure 9**.

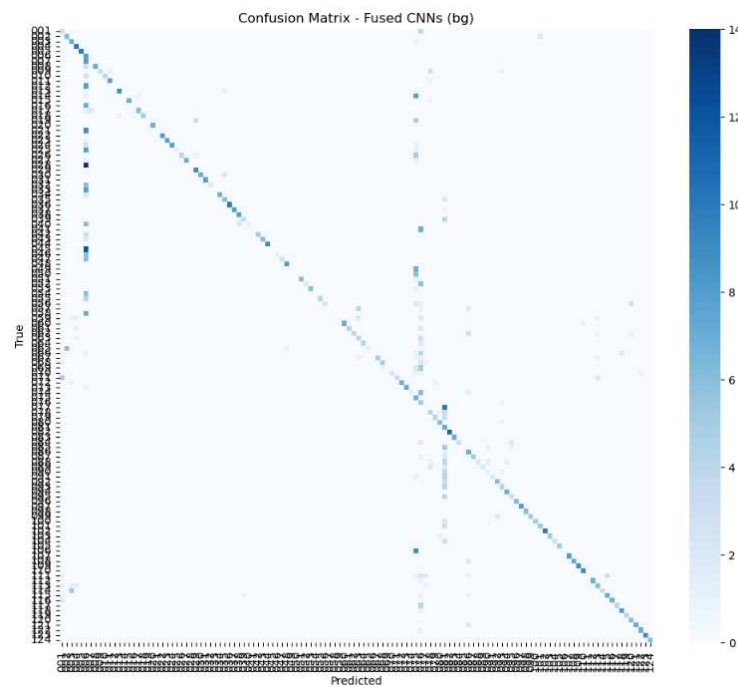


Figure 9 Confusion Matrix for 00 angle of CASIA-B on Experiment 2 using Combined Model

In the third experiment, the training set is the same as the previous one, but the test set is based on cl-01 to cl-02. The same evaluation parameters are used to assess the performance of Model 1, Model 2, and the combined model. In the case of Model 1, the obtained values are 98.32%, 0.0598, 13.83%, and 19.8603 for training accuracy, training loss, test accuracy, and test loss, respectively. With the help of Model 2, the obtained value of training accuracy is 97.03%, training loss is 0.0703, test accuracy is 19.27%, and test loss is 16.5450, respectively. By employing the combined model,

the achieved value of accuracy is 19.47%, the average of precision is 31.0%, the average recall of 26.20%, and the average F1-score of 24.70%, respectively. The results are also verified by computing a confusion matrix as shown in the **Figure 10**.

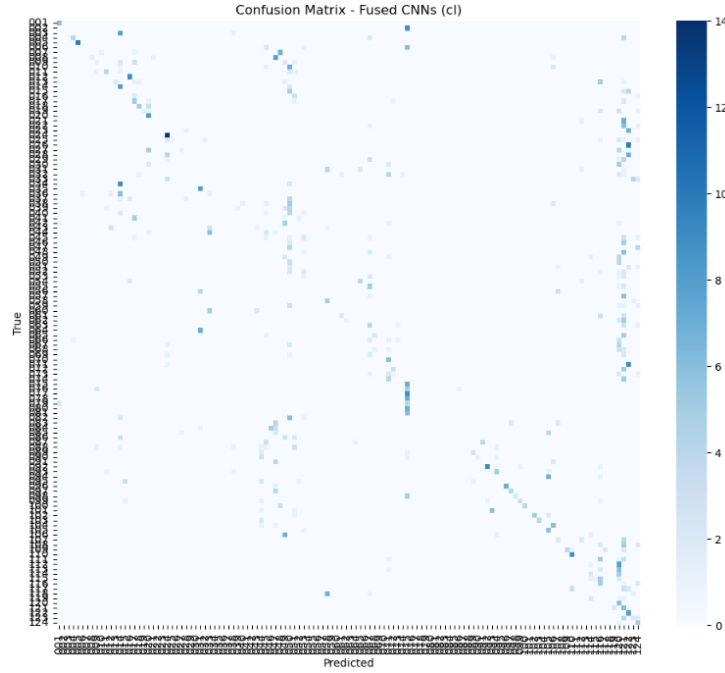


Figure 10 Confusion Matrix for 00 angle of CASIA-B on Experiment 3 using Combined Model

5.1.3 Results of Proposed HGR Method on CASIA-C Dataset

To assess the performance of the proposed HGR method, CASIA-C is also used using Model 1, Model 2, and the combined model. The same parameters and evaluation parameters, such as training accuracy, training loss, test accuracy, and test loss, are used to test the performance. Using Model 1, the attained value is 56.25% for training accuracy, 0.9755 for training loss, 55.37% for test accuracy, and 0.9976 for test loss. In the case of Model 2, the obtained values are 54.61%, 1.0117, 53.20%, and 1.0316 for training accuracy, training loss, test accuracy, and test loss, respectively. The performance of the combined model is also assessed using accuracy, average precision, average recall, and average F1-score. The achieved value for accuracy is 55.59%, 60.0% for average precision, 48.0% for average recall, and 48.50% for average F1-score, respectively.

The confusion matrix is also computed using a combined model for CASIA-C as shown in the **Figure 11**. The obtained true positive values are 1267, 4009, 260, and 1073, for CASIA-C using combined model.

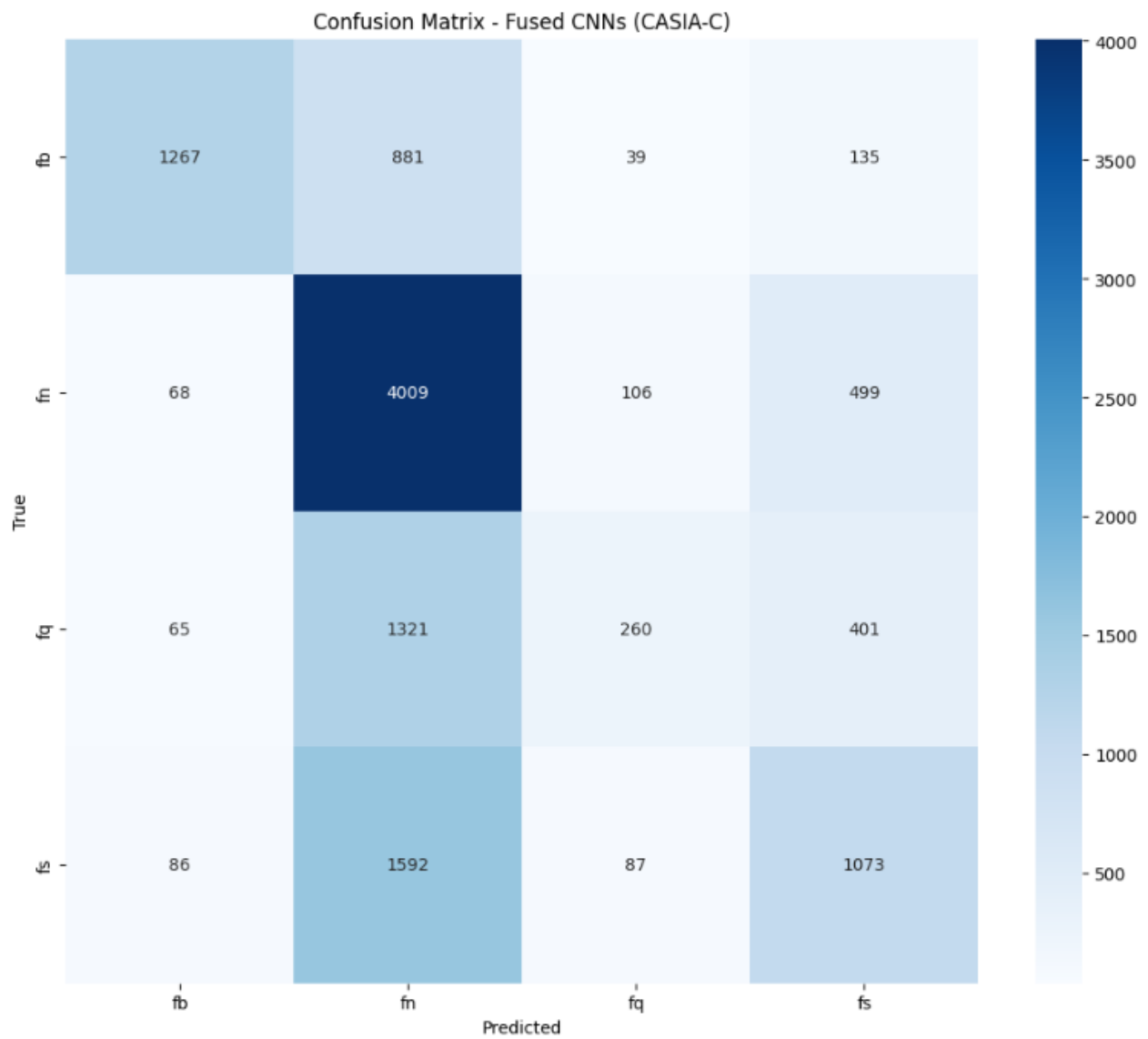


Figure 11 Confusion Matrix for CASIA-C using Combined Model

5.1.4 Discussion and Analysis

The experimental evaluation of the proposed HGR method across multiple datasets, CASIA-A, CASIA-B, and CASIA-C, demonstrates its robustness and adaptability across different viewpoints and walking conditions. The method incorporates three configurations: Model 1, Model 2, and a combined model that fuses the outputs of the individual models for enhanced performance.

On the CASIA-A dataset, three angles (00° , 45° , and 90°) were analyzed. For the 00° angle, both Model 1 and Model 2 achieved high training accuracies (99.00% and 98.80%, respectively) and low training losses, indicating effective learning during training. The combined model outperformed the individual models in test accuracy and yielded high precision (91.35%), recall (91.00%), F1-score (88.30%), and test accuracy (85.00%), showing the benefit of fusion.

At the 45° angle, although Model 1 failed to generalize well (only 11.32% test accuracy), Model 2 achieved excellent test performance (98.47% accuracy), indicating that it captured viewpoint-specific features more effectively. The combined model, however, suffered due to the discrepancy in individual model performances, with test accuracy dropping to 29% and F1-score to 28.75%, revealing that score-level fusion may not always be beneficial if one model performs poorly.

For the 90° angle, both individual models performed strongly, with Model 2 achieving near-perfect test accuracy (99.31%). The combined model maintained high generalization capabilities with a precision of 96.60%, a recall of 93.40%, an F1-score of 92.75%, and an accuracy of 89.00%, demonstrating effective synergy between the models when both perform well.

Experiments using the frontal angle (00°) of CASIA-B explored performance under different walking conditions (normal, carrying a bag, and wearing a coat). In the normal walking scenario (nm-01/02 $\rightarrow$ nm-03/04), Model 2 significantly outperformed Model 1 in test accuracy (68.71% vs. 21.96%), indicating stronger generalization. The combined model offered moderate performance (47.31% accuracy) with improved F1-score (48.80%), suggesting fusion helped balance the strengths and weaknesses of the models.

In the carrying bag scenario (nm-01/02 $\rightarrow$ bg-01/02), Model 2 again showed robust performance (77.49% test accuracy) compared to Model 1 (10.16%), and the combined model achieved a reasonably high F1-score of 58.70%. This scenario reflects the system's capacity to handle occlusions, and appearance changes due to carried objects.

Under clothing variation (nm-01/02 $\rightarrow$ cl-01/02), all models struggled, with Model 2 slightly outperforming Model 1 (19.27% vs. 13.83% test accuracy). The combined model also showed low accuracy (19.47%) and F1-score (24.70%), indicating clothing has a significant negative effect on recognition performance. These results suggest that more robust feature extraction methods may be required to overcome such variations.

The CASIA-C dataset, recorded using infrared imaging, presented a different modality challenge. Both Model 1 and Model 2 yielded relatively lower training and test accuracies ($\sim 55\%$), with moderately high training losses. The combined model marginally improved overall performance, achieving 55.59% accuracy and a slightly better F1-score of 48.50%. These results indicate that while the HGR method has some capacity for cross-modality generalization, further improvements are needed for thermal imagery-based gait recognition.

Overall, the combined model frequently outperforms the individual models when both contribute positively, confirming the effectiveness of the fusion strategy. However, the results also highlight that:

- Model 2 tends to generalize better in more complex scenarios.
- The fusion approach is sensitive to the individual models' balance in performance.
- Performance under clothing variation and thermal conditions remains a challenge and opens avenues for further research using more robust or adaptive representations.

The detailed comparison of Model 1, Model 2, and the combined model is given in the **Table 4**.

Table 4 Detailed Comparison of the Proposed Models

Dataset	Angle / Experiment	Model	Train Acc (%)	Train Loss	Test Acc (%)	Test Loss	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
CASIA-A	00	Model 1	99.00	0.0021	84.70	1.3358	-	-	-	-
		Model 2	98.80	0.0715	77.14	1.8350	-	-	-	-
		Combined	-	-	-	-	91.35	91.00	88.30	85.00
CASIA-A	45	Model 1	99.45	0.0471	11.32	252.4321	-	-	-	-
		Model 2	99.93	0.0013	98.47	0.0462	-	-	-	-
		Combined	-	-	-	-	45.05	29.60	28.75	29.00
CASIA-A	90	Model 1	99.81	0.0180	70.52	11.7668	-	-	-	-
		Model 2	99.87	0.0056	99.31	0.0286	-	-	-	-
		Combined	-	-	-	-	96.60	93.40	92.75	89.00
CASIA-B	nm (01-02 $\rightarrow$ 03-04)	Model 1	96.50	0.1508	21.96	0.1508	-	-	-	-
		Model 2	97.17	0.0888	68.71	1.8450	-	-	-	-
		Combined	-	-	47.31	-	68.40	44.50	48.80	47.31
CASIA-B	nm (01-02 $\rightarrow$ bg-01-02)	Model 1	97.17	0.1011	10.16	23.7638	-	-	-	-
		Model 2	97.65	0.1025	77.49	1.1108	-	-	-	-
		Combined	-	-	57.37	-	70.80	57.40	58.70	57.37
		Model 1	98.32	0.0598	13.83	19.8603	-	-	-	-

CASIA-B	nm (01-02 → cl-01-02)	Model 2	97.03	0.0703	19.27	16.5450	-	-	-	-
		Combined	-	-	19.47	-	31.00	26.20	24.70	19.47
CASIA-C	Whole	Model 1	56.25	0.9755	55.37	0.9976	-	-	-	-
		Model 2	54.61	1.0117	53.20	1.0316	-	-	-	-
		Combined	-	-	55.59	-	60.00	48.00	48.50	55.59

Chapter 6 Conclusion and Future Work

This study offers a new HGR method to address the common challenges, like changes in walking speed, clothing, shadows, and noise. The proposed method is based on an approach of two deep learning models, each with 20 layers, but designed with different parameters. After that, a combined model is designed that fuses their outputs for final classification. We also applied various data augmentation techniques, such as rotation, translation, zoom, and brightness adjustments, to improve the performance of the model. The proposed method was tested on three publicly available datasets, CASIA-A, CASIA-B, and CASIA-C. The results showed strong performance on CASIA-A, achieving up to 99.31% accuracy, and competitive results on CASIA-B, especially under normal and background conditions. However, the performance on CASIA-C was more limited, likely due to the complex and noisy nature of the data. Overall, the results demonstrate that combining models and using thoughtful data preparation can significantly boost gait recognition performance in many scenarios.

The current results are promising, but there is still room for improvement and exploration. In the future, better data augmentation and synthetic data generation can be explored to make the model more robust to variations in lighting, clothing, and environment. Moreover, new and deeper model architectures, like residual networks or attention-based models, can be used to capture the subtle patterns in human gait. The experiment with different fusion techniques, not just at the output level, but also by combining features during earlier stages of the model to improve accuracy and consistency, can be carried out. Furthermore, to improve robustness to real-world noise, such as background movement, occlusions, and weather conditions, can be improved.

References

- [1] M. I. Sharif *et al.*, "Deep learning and kurtosis-controlled, entropy-based framework for human gait recognition using video sequences," vol. 11, no. 3, p. 334, 2022.
- [2] T. Teepe, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "Towards a deeper understanding of skeleton-based gait recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1569-1577.
- [3] H. Arshad *et al.*, "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition," p. e12541, 2020.
- [4] A. Mehmood, J. Amin, M. Sharif, and S. J. P. R. L. Kadry, "Human Gait Recognition by using Two Stream Neural Network along with Spatial and Temporal Features," vol. 180, pp. 16-25, 2024.
- [5] M. Zhai, X. Xiang, N. Lv, and X. J. P. R. Kong, "Optical flow and scene flow estimation: A survey," vol. 114, p. 107861, 2021.
- [6] M. Munsif, S. U. Khan, N. Khan, A. Hussain, M. J. Kim, and S. W. J. K.-B. S. Baik, "Contextual visual and motion salient fusion framework for action recognition in dark environments," vol. 304, p. 112480, 2024.
- [7] M. Deng, Z. Fan, P. Lin, and X. J. I. T. o. M. Feng, "Human gait recognition based on frontal-view sequences using gait dynamics and deep learning," vol. 26, pp. 117-126, 2023.
- [8] V. Rani, M. J. M. T. Kumar, and Applications, "Transfer learning for human gait recognition using VGG19: CASIA-A dataset," pp. 1-15, 2024.
- [9] R. N. Yousef, A. T. Khalil, A. S. Samra, and M. M. Ata, "Model-based and model-free deep features fusion for high performed human gait recognition," *J Supercomput*, vol. 79, no. 12, pp. 1-38, Mar 19 2023.
- [10] A. Mehmood *et al.*, "Human gait recognition: A deep learning and best feature selection framework," 2021.
- [11] F. Saleem *et al.*, "Human Gait Recognition: A Single Stream Optimal Deep Learning Features Fusion," *Sensors (Basel)*, vol. 21, no. 22, p. 7584, Nov 15 2021.
- [12] K. Soni, D. P. Dogra, A. A. Sekh, S. Kar, H. Choi, and I.-J. J. E. S. w. A. Kim, "Person re-identification in indoor videos by information fusion using Graph Convolutional Networks," vol. 210, p. 118363, 2022.
- [13] M. Deng, Z. Zhong, Y. Zou, Y. Wang, K. Wang, and J. J. N. P. L. Liao, "Human Gait Recognition Based on Frontal-View Walking Sequences Using

- Multi-modal Feature Representations and Learning," vol. 56, no. 2, p. 133, 2024.
- [14] E. J. Harris, I.-H. Khoo, E. J. F. i. R. Demircan, and AI, "A survey of human gait-based artificial intelligence applications," vol. 8, p. 749274, 2022.
 - [15] W. Xu, C. Zhu, and Z. J. P. R. L. Wang, "Multiview max-margin subspace learning for cross-view gait recognition," vol. 107, pp. 75-82, 2018.
 - [16] Z. Zhang *et al.*, "Gait Recognition via Disentangled Representation Learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4705-4714, Long Beach, CA, USA: IEEE.
 - [17] G. Liu, S. Zhong, and T. Li, "Gait recognition method of temporal–spatial HOG features in critical separation of Fourier correction points," *Future Generation Computer Systems*, vol. 94, pp. 11-15, May 1, 2019 2019.
 - [18] A. Mehmood, J. Amin, M. Sharif, S. Kadry, and J. J. P. o. Kim, "Stacked-gait: A human gait recognition scheme based on stacked autoencoders," vol. 19, no. 10, p. e0310887, 2024.
 - [19] H. Arshad, M. A. Khan, M. Sharif, M. Yasmin, and M. Y. Javed, "Multi-level features fusion and selection for human gait recognition: an optimized framework of Bayesian model and binomial distribution," (in en), *International Journal of Machine Learning and Cybernetics*, 2019-03-18 2019.
 - [20] M. Sharif *et al.*, "A Machine Learning Method with Threshold Based Parallel Feature Fusion and Feature Selection for Automated Gait Recognition," vol. 32, no. 2, pp. 67-92, 2020.
 - [21] W. Li, C.-C. J. Kuo, and J. Peng, "Gait recognition via GEI subspace projections and collaborative representation classification," (in en), *Neurocomputing*, vol. 275, pp. 1932-1945, 01/2018 2018.
 - [22] A. S. George, E. Roy, A. Antony, and M. J. I. J. I. A. C. S. Job, "An efficient gait recognition system for human identification using neural networks," vol. 6, no. 5, pp. 2347-8616, 2017.
 - [23] K. Sugandhi, F. F. Wahid, and G. Raju, "Detection of human gait cycle: An overlap based approach," in *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)(ICTUS)*, 2017, pp. 1-3: IEEE.
 - [24] H. Sharma and J. Grover, "Human identification based on gait recognition for multiple view angles," (in en), *International Journal of Intelligent Robotics and Applications*, vol. 2, no. 3, pp. 372-380, 9/2018 2018.
 - [25] M. Deng, Y. Zou, Z. Zeng, Y. Wang, X. Feng, and Y. J. E. A. o. A. I. Liu, "Human gait recognition based on frontal-view sequence using discriminative optical flow feature representations and learning," *Engineering Applications of Artificial Intelligence*, vol. 145, p. 110213, 2025.

- [26] F. M. Castro, M. J. Marín-Jiménez, N. Guil, and N. Perez De La Blanca, "Automatic learning of gait signatures for people identification," in *International work-conference on artificial neural networks*, 2017, pp. 257-270: Springer.
- [27] M. Alotaibi and A. Mahmood, "Improved gait recognition based on specialized deep convolutional neural network," (in en), *Computer Vision and Image Understanding*, vol. 164, pp. 103-110, 11/2017 2017.
- [28] X. Ben, C. Gong, P. Zhang, R. Yan, Q. Wu, and W. Meng, "Coupled Bilinear Discriminant Projection for Cross-view Gait Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1-1, 2019 2019.
- [29] J. Condell, P. Chaurasia, J. Connolly, P. Yogarajah, G. Prasad, and R. Monaghan, "Automatic Gait Recognition and its Potential Role in Counterterrorism," (in en), *Studies in Conflict & Terrorism*, vol. 41, no. 2, pp. 151-168, 2018-02-01 2018.
- [30] K. Z. Htun and S. Maung Maung Zaw, "Human Identification System Based on Statistical Gait Features," in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, 2018, pp. 508-512, Singapore: IEEE.
- [31] M. Deng and C. Wang, "Human Gait Recognition Based on Deterministic Learning and Data Stream of Microsoft Kinect," (in en), *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 12, pp. 3636-3645, 12/2019 2019.
- [32] M. I. Junaid, S. Madarapu, S. J. S. Ari, Image, and V. Processing, "Human gait recognition using attention based convolutional network with sequential learning," vol. 19, no. 1, pp. 1-8, 2025.
- [33] D. Xu *et al.*, "A new method proposed for realizing human gait pattern recognition: Inspirations for the application of sports and clinical gait analysis," vol. 107, pp. 293-305, 2024.
- [34] X. Shi, W. Zhao, H. Pei, H. Zhai, and Y. J. I. A. Gao, "Research on gait recognition based on GaitSet and multimodal fusion," 2025.
- [35] L. Zhang, Z. Men, W. J. S. Xie, Image, and V. Processing, "Gaitts: indoor gait recognition with multi-scale temporal-spatial information aggregation," vol. 19, no. 1, pp. 1-12, 2025.
- [36] J. Kim, B. Kim, and H. J. I. A. Lee, "Temporally deformable convolution for gait recognition," 2025.
- [37] H. Aggarwal and D. K. Vishwakarma, "Covariate Conscious Approach for Gait Recognition Based Upon Zernike Moment Invariants," (in en), *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 2, pp. 397-407, 6/2018 2018.

- [38] X. Wang, J. Wang, and K. Yan, "Gait recognition based on Gabor wavelets and (2D)2PCA," (in en), *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12545-12561, 5/2018 2017.
- [39] M. Balazia and P. Sojka, "Gait Recognition from Motion Capture Data," (in en), *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 14, no. 1s, pp. 1-18, 2018-02-21 2018.
- [40] M. H. Ghaeminia and S. B. Shokouhi, "GSI: efficient spatio-temporal template for human gait recognition," (in en), p. 23.
- [41] A. O. Lishani, L. Boubchir, E. Khalifa, and A. Bouridane, "Human gait recognition based on Haralick features," (in en), *Signal, Image and Video Processing*, vol. 11, no. 6, pp. 1123-1130, 9/2017 2017.
- [42] S. Yu, H. Chen, Q. Wang, L. Shen, and Y. Huang, "Invariant feature extraction for gait recognition using only one uniform model," (in en), *Neurocomputing*, vol. 239, pp. 81-93, 05/2017 2017.
- [43] R. Liao, C. Cao, E. B. Garcia, S. Yu, and Y. Huang, "Pose-Based Temporal-Spatial Network (PTSN) for Gait Recognition with Carrying and Clothing Variations," in *Biometric Recognition*, vol. 10568, J. Zhou *et al.*, Eds. Cham: Springer International Publishing, 2017, pp. 474-483.
- [44] M. H. Khan, M. S. Farid, and M. Grzegorzec, "Spatiotemporal features of human motion for gait recognition," (in en), *Signal, Image and Video Processing*, vol. 13, no. 2, pp. 369-377, 3/2019 2019.
- [45] M. S. Islam, M. R. Islam, M. A. Hossain, A. Ferworn, and M. K. I. Molla, "Subband entropy-based features for clothing invariant human gait recognition," (in en), *Advanced Robotics*, vol. 31, no. 10, pp. 519-530, 2017-05-19 2017.
- [46] J. Sun, Y. Wang, J. Li, W. Wan, D. Cheng, and H. Zhang, "View-invariant gait recognition based on kinect skeleton feature," (in en), *Multimedia Tools and Applications*, vol. 77, no. 19, pp. 24909-24935, 10/2018 2018.
- [47] H. Arshad *et al.*, "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition," vol. 39, no. 7, p. e12541, 2022.
- [48] H. Zhan, J. Kou, Y. Cao, Q. Guo, J. Zhang, and Y. J. M. Shi, "Human Gait phases recognition based on multi-source data fusion and BILSTM attention neural network," vol. 238, p. 115396, 2024.
- [49] X. Ben, P. Zhang, Z. Lai, R. Yan, X. Zhai, and W. Meng, "A general tensor representation framework for cross-view gait recognition," *Pattern Recognition*, vol. 90, pp. 87-98, June 1, 2019 2019.
- [50] X. Wang and W. Q. Yan, "Cross-view gait recognition through ensemble learning," (in en), *Neural Computing and Applications*, vol. 32, no. 11, pp. 7275-7287, 2019-5-16 2019.

- [51] C. Li, S. Sun, X. Chen, and X. Min, "Cross-view gait recognition using joint Bayesian," in *Ninth International Conference on Digital Image Processing (ICDIP 2017)*, 2017, p. 1042009, Hong Kong, China.
- [52] F. M. Castro, M. J. Marín-Jiménez, N. G. Mata, and R. Muñoz-Salinas, "Fisher Motion Descriptor for Multiview Gait Recognition," (in en), *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no. 01, p. 1756002, 01/2017 2017.
- [53] J. Kovač, V. Štruc, and P. Peer, "Frame-based classification for cross-speed gait recognition," (in en), *Multimedia Tools and Applications*, vol. 78, no. 5, pp. 5621-5643, 3/2019 2019.
- [54] A. Sokolova and A. Konushin, "Gait Recognition Based on Convolutional Neural Networks," (in en), *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W4, pp. 207-212, 2017-05-10 2017.
- [55] M. H. Khan, F. Li, M. S. Farid, and M. Grzegorzek, "Gait Recognition Using Motion Trajectory Analysis," 2018, pp. 73-82: Springer International Publishing.
- [56] M. Nabila, A. I. Mohammed, and B. J. Yousra, "Gait-based human age classification using a silhouette model," (in en), *IET Biometrics*, vol. 7, no. 2, pp. 116-124, 2018-03-01 2018.
- [57] D. Marcin, "Human gait recognition based on ground reaction forces in case of sport shoes and high heels," in *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 2017, pp. 247-252, Gdynia, Poland: IEEE.
- [58] N. Khamsemanan, C. Nattee, and N. Jianwattanapaisarn, "Human Identification From Freestyle Walks Using Posture-Based Gait Feature," (in en), *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, pp. 119-128, 1/2018 2018.
- [59] Z. Mahfouf, H. F. Merouani, I. Bouchrika, and N. Harrati, "Investigating the use of motion-based features from optical flow for gait recognition," (in en), *Neurocomputing*, vol. 283, pp. 140-149, 03/2018 2018.
- [60] A. Mehmood, J. Amin, M. Sharif, S. Kadry, and J. J. P. o. Kim, "Stacked-gait: A human gait recognition scheme based on stacked autoencoders," *PLOS ONE*, vol. 19, no. 10, p. e0310887, 2024.
- [61] N. Aman, M. R. Islam, M. F. Ahamed, and M. J. T. Ahsan, "Performance Evaluation of Various Deep Learning Models in Gait Recognition Using the CASIA-B Dataset," vol. 12, no. 12, p. 264, 2024.
- [62] L. Wang, T. Tan, H. Ning, W. J. I. t. o. p. a. Hu, and m. intelligence, "Silhouette analysis-based gait recognition for human identification," vol. 25, no. 12, pp. 1505-1518, 2003.

- [63] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, vol. 4, pp. 441-444: IEEE.
- [64] D. Tan, K. Huang, S. Yu, and T. Tan, "Efficient night gait recognition based on template matching," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, vol. 3, pp. 1000-1003: IEEE.
- [65] C. Shorten and T. M. J. J. o. b. d. Khoshgoftaar, "A survey on image data augmentation for deep learning," vol. 6, no. 1, pp. 1-48, 2019.
- [66] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Icdar*, 2003, vol. 3, no. 2003: Edinburgh.
- [67] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," 2012, pp. 1097–1105.
- [68] H. Zhang, M. Cisse, Y. N. Dauphin, and D. J. a. p. a. Lopez-Paz, "mixup: Beyond empirical risk minimization," 2017.
- [69] L. Perez and J. J. a. p. a. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017.
- [70] F. Chollet, *Deep learning with Python*. simon and schuster, 2021.
- [71] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026-1034.
- [72] J. Kittler, M. Hatef, R. P. Duin, J. J. I. t. o. p. a. Matas, and m. intelligence, "On combining classifiers," vol. 20, no. 3, pp. 226-239, 2002.