Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

HUMAN GAIT RECOGNITION FOR BIOMETRIC IDENTIFICATION USING DEEP NEURAL NETWORK

Aqsa Akram[*1], Naila Batool[2], Dr.Umair Muneer Butt[3], Dr. Imtiaz Hussain[4]

[*1,2,3,4]*Department of Computer Science, University of Management and Technology, Sialkot, Pakistan*

[*1]aqsacheemacs@gmail.com, [2]nailabatool@gmail.com, [3]umair.muneer@skt.umt.edu.pk, [4]imtiaz.hussain@skt.umt.edu.pk

DOI: https://doi.org/10.5281/zenodo.18066212

Keywords
Human Gait Recognition, Biometric Identification, lightweight CNN, Deep Neural Network

Article History
Received: 28 October 2025
Accepted: 13 December 2025
Published: 27 December 2025

Corresponding Author: *
Aqsa Akram

Abstract
Gait recognition has become an increasingly valuable solution for the non-invasive identification of any individual, particularly in the domains of security and healthcare. Despite its advantages, the variations in viewing angles, clothing, and carried objects continue to hinder the robustness of the model in real-world scenarios. Therefore, in this study, we present an automated gait recognition system that addresses these limitations through a series of deep learning-based novel enhancements. Firstly, a sequential contrast enhancement technique is used to improve the visual quality of the image, followed by another technique named, mean-based contrast enhancement. The resultant input channels are then further fused into a unified representation using a streamlined integration method. In addition, the Bayesian hyperparameter optimization technique is employed to optimize the performance of the model. The proposed system also incorporates a Parallel Feature Extraction and Fusion (PFxF) method to enhance the spatial and temporal information, along with an improved Scale and Rotation Invariance Optimization (SRIO) technique for a robust selection of features. Furthermore, we evaluated multiple models to perform classification tasks, including L-SVM, C-SVM, Q-SVM, BG Tree, and MD-NN models on the CASIA-B dataset. The performance of each classifier was assessed using standard metrics such as True Positive Rate (TPR), error rate, precision, recall, accuracy, and Area Under the Curve (AUC). The results show that the BG Tree classifier consistently outperformed its counterparts across all covariate conditions. It recorded the highest TPR of 98.40%, 96.20%, and 95.80% for the BAG, COAT, and NORMAL classes, respectively. Moreover, it achieved the leading precision of 97.57%, recall of 97.57%, accuracy of 97.60%, and AUC of 97.67%, while maintaining the lowest error rate of 2.40%. These outcomes demonstrate the reliability and adaptability of our proposed approach in diverse conditions, making it an effective solution for future gait-based biometric applications.

INTRODUCTION

With visual cameras, gait recognition (GR), a developing biometric modality, allows for remote person identification. Since it is more difficult to discern between real and fake

signals, GR offers a safe and dependable substitute for face and fingerprint authentication. Moreover, gait recognition is adaptable to all kinds of contexts because of its resilience to spoofing. Deep learning has led to consistently better advancements in human gait recognition technology, encouraging outcomes in various settings. With the increasing use of video surveillance, additional challenges emerge, including maintaining consistency in performance assessment across several protocols, maintaining dependable identification in the face of changing illumination, managing variations in walking patterns, and safeguarding privacy [1]. Compared to other biometric identification systems, this research seeks to provide an overview of gait recognition and examine the environmental factors and any issues that may impact it. Also, empirical evaluation of the current deep learning (DL) methods for human GR is the main objective, as it may lead to new research directions [2, 3]. Human gait recognition (HGR) is a biometric technique that uses gait traits to identify people without coming into contact with them directly. Gait is a type of human motion that depends on spatial, static, and temporal characteristics [4]. Due to its superior discriminative abilities over other biometric identification methods like facial recognition, fingerprint analysis, palm print scanning, and iris recognition, it has consequently grown significantly in popularity over the past few decades in the fields of machine learning (ML) and computer vision (CV). Gait recognition has also seen a substantial increase in its practical applications, finding usage in forensics, video surveillance, and crime prevention. Low contrast gait identification methods are also being investigated [4]. Human Gait Recognition (HGR) faces significant challenges due to variations in viewing angles, clothing, and carried items, which critically impact recognition accuracy. In the context of Convolutional Neural Networks (CNNs) [5], the traditional practice of using static hyperparameters includes learning rate, epochs, and momentum. These can limit the model's ability to adapt to these variations during training [6]. Additionally, existing feature selection techniques, including entropy-based, distance-based, and evolutionary algorithms [7], often struggle to consistently retain essential features, leading to suboptimal model performance [8]. To overcome these issues, we introduce a new deep learning-based gait recognition method that improves image quality by applying contrast enhancement methods and combines feature channels with a simplified representation scheme. Our approach also uses Bayesian hyperparameter optimization to adaptively adjust learning parameters, and employs a Parallel Feature Extraction and Fusion (PFxF) mechanism with an improved Scale and Rotation Invariance Optimization (SRIO) for more efficient feature selection. Also, we compared several models to conduct classification tasks such as L-SVM, C-SVM, Q-SVM, BG Tree, and MD-NN models. The performance of the classifiers was evaluated using common metrics like True Positive Rate (TPR), error rate, precision, recall, accuracy, and Area Under the Curve (AUC). On testing on the CASIA-B dataset, our system obtained impressive recognition performance, with the highest performing classifier achieving an accuracy of 97.60%, reflecting its strong performance under different covariate conditions.

Our key contributions include:

1. A sequential contrast enhancement method that segments images into separate channels before applying mean-based contrast enhancement.
2. A simple fusion procedure that integrates enhanced channels into a single frame.
3. The application of Bayesian optimization for dynamic fine-tuning of deep learning model hyperparameters.
4. The development of a Parallel Features Extraction and Fusion

(PFxF) method for robust feature engineering.

5. Implemented five different classifiers and compared results to obtain the optimal most. The rest of the article is organized as related work, proposed methodology, results and discussion are elaborated. Finally, the conclusion and future work are in the last of paper.

Related Work

Human gait recognition has advanced into a robust biometric identification technique due to its non invasive nature and effectiveness under diverse environmental conditions. Early methods relied on handcrafted and geometric models, while recent trends have moved toward deep learning, and multimodal systems. This review provides an in depth overview of foundational and recent contributions in the field. One of the foundational approaches in gait recognition focused on model-based methods, which utilized geometrical and kinematic representations of the human body to analyze walking patterns. Tafazzoli et al. [9] laid early groundwork by modeling human limb motion using elliptic shapes and Fourier descriptors, enabling a mathematical abstraction of gait dynamics. Similarly, Yam et al. [10] enhanced system resilience under occlusion through coupled oscillators that simulated coordinated limb movements. Expanding spatial modeling capabilities, Ryu and Kamata [11] employed stereo vision to generate 3D point clouds, addressing the complexity of frontal view gait recognition. In a practical surveillance context, Condell et al. [12] integrated Kinect-based skeletal data with deterministic learning to capture gait dynamics in real-world environments. In recent years, traditional model-based constraints have been addressed by integrating deep learning. Notably, Naz et al. [13] proposed the E-Gait framework, utilizing Graph Convolutional Networks (GCNs) with spatiotemporal joint attention to capture joint dynamics over time. To handle viewpoint variation, Zou et al. [14] applied a transformer architecture with progressive perspective training, achieving over 97% accuracy across 11 angles. A study by Zhang et al. [15] further improved cross condition robustness by disentangling pose and appearance representations. Additionally, Wang et al. [16] introduced a self adaptive Hidden Markov Model (SAHMM) guided by Local Gait Energy Images (LGEIs) to manage variable gait sequences. Complementing these, Gabor region-based modeling [17] was adopted to enhance silhouette representation using frequency-domain analysis. Sharif et al. [2] proposed a robust feature selection framework by combining high value features and refining them using Euclidean distance and correlation-based techniques, achieving over 94% accuracy. Similarly, Abdullah and Alfy [18] employed statistical analysis of Gabor responses extracted from region-level GEIs, on CASIA-B and OU-ISIR datasets. Collectively, these advancements demonstrate a clear evolution toward more robust, view invariant, and semantically rich model-based gait recognition systems. Further advancements in skeleton-based gait recognition highlight the effectiveness of body joint representations for modeling human motion. Gao et al. [19] introduced a spatiotemporal decomposition technique to minimize redundancy in joint trajectories while improving occlusion resilience. Building on this, Shopon et al. [20] proposed a residual GCNN architecture that preserved low-level kinematic dependencies through residual pathways. In parallel, Zhu et al. [21] advanced 3D body shape modeling by reconstructing full-body contours, enhancing invariance to clothing and viewpoint changes. Additionally, Slemensek et al. [22] found that skeleton-based features yielded high generalizability across classifiers in activity recognition tasks. To further boost robustness under occlusion, Ye et al. [23] proposed multiscale gait optical flow pyramids to capture motion at various resolutions, while Chen et al. [24] enhanced skeleton modeling using adversarial synthesis and spatial channel attention mechanisms.

Parallel to model-based developments, model free approaches gained momentum due to their relative ease of implementation and visual interpretability. Surveys by Shirke et al. [25] and Harris et al. [26] outlined the evolution of these techniques and their application in domains like healthcare and surveillance. Zhu et al. [27] addressed benchmark limitations by introducing the GREW dataset, which offers diverse, real-world gait samples. Among silhouette-based methods, Chao et al. [1] proposed the GaitSet model, which treats silhouettes as unordered sets and achieved state of the art results on CASIA-B. To improve robustness under clothing variation and occlusion, Shi et al. [28] enhanced this approach by fusing gait energy and silhouette images. Template-based methods have significantly contributed to gait recognition by aggregating silhouette sequences into spatiotemporal representations.

Ghaeminia and Shokouhi [29] introduced the Gait Salient Image (GSI), which effectively compresses motion energy distributions over time, while Lishani et al. [30] enhanced discriminability using Haralick features extracted from localized GEI regions. To improve robustness under covariate conditions like clothing changes and partial occlusion, Deng et al. [31] proposed a hybrid framework that fuses local and global entropy-based features into GEIs. Complementary approaches include Isaac et al. [32], who utilized genetic algorithms for optimal template segmentation, and MLOOP descriptors [8], designed to preserve spatial layout under appearance changes. Liu et al. [33] combined Fourier and HOG features, achieving 96.3% accuracy on CASIA-B. Expanding applicability, Sudha and Bhavani [34] combined spatial, temporal, and wavelet features within an SVM framework, demonstrating strong performance on the NLPR dataset. Addressing view variation, Tian et al. [35] introduced a view adaptive mapping approach leveraging walking trajectory fitting, while Xu et al. [36] proposed multiview max margin subspace learning, achieving effective cross view clustering across CASIA-B and OU-ISIR. Addressing broader trends, S¸ ahan et al. [37] surveyed recent appearance-based gait recognition methods, outlining key challenges such as occlusion and view variation, and emphasizing the need for adaptable, AI-driven solutions. With the rise of deep learning, gait recognition has seen remarkable improvements in both feature extraction and classification accuracy. Shehzad et al. [38] and Li et al. [39] independently contributed to spatial-temporal modeling, Shehzad introduced a

two-stream deep network for multi-angle gait recognition on CASIA-B, while Li developed DeepGait, combining a VGG-based network with Joint Bayesian modeling to enhance silhouette-based cross-view identification. Similarly, Sezavar et al. [40] proposed DCapsNet, a hybrid CNN and capsule network with dynamic routing—achieving up to 99.3% accuracy on WISDM and WhuGAIT, demonstrating its viability for mobile deployment. Russakovsky et al. [41] highlighted the foundational role of the ImageNet challenge in advancing deep learning-based visual recognition, which has greatly influenced gait recognition via transfer learning. Reinforcing low resource performance, Mehmood et al. [42] used CNNs with MediaPipe features, reaching 89.58% accuracy in remote surveillance. Optical flow based solutions [43, 44] employed bi-layer neural networks across CASIA-A/B/C, reporting up to 99.6% accuracy. Yu et al. [45] applied stacked autoencoders to generate invariant gait features, generalizing well under view and clothing variation. Sharif et al. [46] and Li et al. [47] advanced fusion strategies—Sharif using threshold-based parallel fusion of shape and geometric features (98.6% on CASIA-A), and Li leveraging GEI subspace projections with collaborative classification for efficiency. Temporal modeling also progressed, Noori and Ali [48] integrated ConvLSTM with AlexNet and ResNet-150, achieving 100% accuracy

on local data. Aman et al. [49] benchmarked CNNs, MLPs, SOMs, and EfficientNet, with CNN yielding 97.12% accuracy on CASIA-B post tuning. Sharma and Grover [50] proposed the GII-BPSF descriptor for improved multi angle normalization, while Castro et al. [51] validated that optical flow cuboids fed to CNNs perform well even on low-res datasets like TUM-GAID. Focusing on optimized fusion, Khan et al. [52] combined EfficientNetB0 and MobileNetV2 with Moth Flame Optimization, achieving 99.7% accuracy at 180° on CASIA-B with reduced computational cost. Likewise, Pundir et al. [53] used MobileNetV1-Xception with SVM, attaining 98%, while Arshad et al. [54] fused AlexNet and VGG19 via entropy-based selection. Kim et al. [3] introduced temporally deformable convolutions to adaptively learn intra cycle gait variations. Despite high performance, most of these methods require substantial datasets and computation, limiting deployment in constrained settings. Bayesian optimization has proven effective for hyperparameter tuning in gait recognition. Khan et al. [55] introduced a hybrid model leveraging Bayesian methods for dynamic parameter adjustment, improving robustness under covariate shifts. Likewise, Bilal et al. [56] developed Gait-Star with attention-based spatiotemporal feature reweighting, and Shi et al. [57] proposed SConvLSTM, a hybrid network capturing waveform and temporal dependencies from inertial data. Peng et al. [58] combined silhouette and skeleton inputs for enriched gait representation. To tackle covariate variations, Li et al. [59] introduced a joint intensity transformer network using contrastive and triplet loss functions, while Arshad et al. [60] fused features using Bayesian modeling, achieving 100% accuracy on AVA-MVG and 98.8% on CASIA-A. An et al. [61] addressed data scarcity by releasing OUMVLP-Pose, a large-scale, multiview pose-based dataset. Although Sezavar et al. [62] improved spatial hierarchies with capsule networks, such models often suffer from high computational cost and complex tuning requirements. Sensor based gait recognition has gained traction due to its suitability for real time and healthcare applications. Hasan et al. [63] developed a lightweight CNN optimized for on device processing, while Huang et al. [64] implemented stride segmentation using time warping. Kinect-based systems by Souza and Stemmer [65] and Deng and Wang [66] demonstrated high recognition rates (99%) by capturing skeletal and kinetic features, with the latter incorporating deterministic learning for surveillance contexts. Shi et al. [67] integrated accelerometer and gyroscope data via deep networks, and Balazia and Sojka [68] extracted robust features from raw motion capture signals. Matovski et al. [69] highlighted clothing variation as a more significant covariate than elapsed time. To enhance recognition accuracy, Xu et al. [70] employed metaheuristic optimization to select biomechanically relevant features, achieving 99.8% accuracy. Moreover, benchmarking under real world conditions was supported by Wang et al. [6], who introduced a dataset focused on clothing changes, while Zheng et al. [71] used robust PCA for effective view transformation in cross view settings. Multimodal gait recognition systems have significantly expanded the potential of biometric identification by integrating complementary modalities. For instance, Ekpo [72] developed a CNN based model that fused facial and gait cues, achieving 92.3% accuracy, while Maity et al. [73] advanced this by merging low resolution facial data with frontal-view gait features, yielding a rank-1 accuracy of 95.9%. To enhance temporal modeling, Mehmood et al. [74] employed spatial–temporal neural networks, whereas Asif et al. [7] tackled view invariant recognition across gallery settings using deep CNNs. Further optimization strategies include Rao et al. [75] firefly algorithm and Hassan et al. [76] analysis of dimensionality reduction techniques. Khan et al. [77] improved robustness by canonical

correlation analysis for efficient CNN feature fusion. Si [4] proposed a hybrid LSTM-DBN model for sequential learning, and Khan et al. [78] utilized point cloud modeling for frontal view recognition, reflecting a shift toward adaptable multimodal architectures. Several studies have focused on optimizing feature selection and fusion strategies to improve gait recognition performance, particularly under high dimensional and covariate rich conditions. Dou et al. [79] introduced the CLASH framework, which integrates dense spatial-temporal descriptors with neural architecture search, achieving a notable 98.8% accuracy, Derlatka et al. [80] further expanded the field by employing heterogeneous ensemble classifiers to enhance model generalization, and Khan et al. [5] proposed a resource efficient deep CNN framework tailored for constrained computational environments. Complementing these innovations,
Mehmood et al. [81] achieved 100% accuracy on CASIA-A by encoding gait features with Fisher vectors, although their approach lacked deep contextual modeling. Hussain et al. [82] addressed feature redundancy and dimensionality challenges by integrating the sine cosine algorithm into Harris Hawks Optimization, demonstrating up to 92% accuracy on high dimensional datasets. Collectively, these approaches signify a shift toward intelligent feature optimization, where accuracy, efficiency, and adaptability are simultaneously prioritized through hybridization and learning-based strategies. From the reviewed literature, it is evident that while substantial progress has been made in gait recognition through deep learning challenges such as covariate variations, spatial distortions, and feature redundancy continue to hinder system robustness and generalization [83]. Many existing models struggle to maintain a balance between accuracy and efficiency, often performing well in controlled settings but failing to adapt to real world complexities. Addressing these limitations, our study proposes a lightweight yet highly effective framework built upon the CNN-55 model. By integrating sequential contrast enhancement, efficient channel fusion, and Bayesian optimization, along with PFxF based spatial-temporal extraction and SRIO based feature refinement, the proposed system not only aligns with current research trends but also demonstrates strong potential for realtime and reliable gait recognition across diverse scenarios.

Proposed Methodology

We propose a deep learning-based approach to address the challenges of reliable gait recognition. Earlier research in gait recognition primarily focused on gait sequence representation, considering it a critical step in building generalized gait recognition systems. As a result, various feature descriptors were introduced, among which the Gait Energy Image (GEI) became the most widely used. GEI is
a single image generated by averaging all frames within a gait cycle or sequence, making it computationally simple and efficient.
In our approach, we aim to capture both the spatial and temporal dynamics of human gait by generating a Spatial-Temporal Dynamic Gait Image (STDGI). This section details the proposed methodology, which leverages a pretrained deep convolutional neural network (CNN) with customized layers, followed by the application of five different classifiers. Each step of the proposed pipeline is described in detail, highlighting the techniques and design choices that contributed to achieving accurate gait recognition results.

Proposed model architecture

As Figure, 1 illustrates. Two streams form the basis of this approach. Before being processed further, the video data is first transformed into picture frames. To extract better characteristics, the picture frames are transformed into grayscale images in the second stage. In the third phase, a CNN55 deep learning model was implemented already which was specifically trained for feature

extraction. Deep feature extraction is done via a two-stream network. The Horn and Schunk (HS) technique is applied in Stream-1 to extract optical flow. Following that, optical flow frames are sent into the proposed CNN-55, where deep features are extracted on the 49th layer, known as the Max Pooling Layer, with the retrieved features vector (FV) being FV1 and the parameter named "maxpool6." Stream-2 involves feeding grayscale pictures straight into a CNN-55, which then extracts spatial features on the 49th layer, known as the Max Pooling Layer, to produce FV2. These two FVs are serially fused to create FV3 in the fourth stage. To optimise the characteristics, GA is used in the sixth phase. The One-Versus-All SVM (OVASVM) classifier receives the final FV in the last stage in order to make the final recognition.

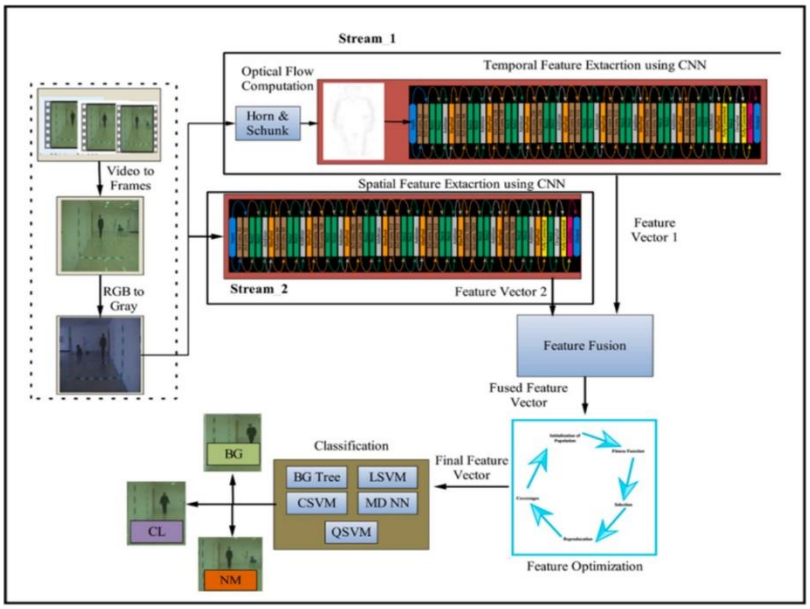

Fig 1. A two-stream architecture for human gait classification

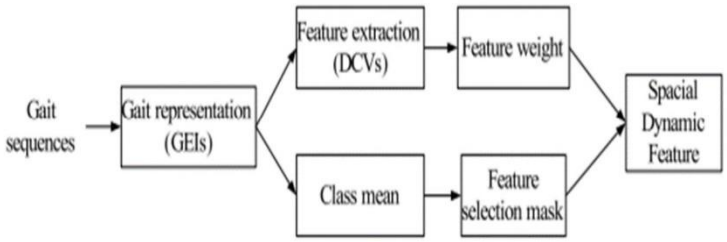

Fig 2. Process diagram of spatial dynamic attribute extraction

GEIs are acquired for every gait sequence as shown in Figure. 2. Two distinct dynamic feature extraction methods are then used based on the GEIs. The dynamic feature selection mask is acquired by preserving the sample pixels' class mean intensity levels threshold, and the dynamic feature weight is derived from the learnt DCVs.

Dataset Description

In this research, we employed the CASIA-B gait database. The preprocessing of sounds was

not relevant because the CASIA database was acquired in an indoor setting with a basic background. It was utilised for feature description comparison and included a variety of walking circumstances. We procured the publicly accessible CASIA-B dataset [67]. The dataset known as CASIA-B is acquired within an enclosed space. It is predicated on 124 actors' data. There are eleven viewing angles in this dataset, ranging from 000 to 1800. Each perspective has three variations: the person walking normally (NORMAL), the person wearing a coat (COAT), and the person carrying a bag (BAG). Ten movies have been recorded for each topic; six of the videos are of NORMAL, two are of COAT, and two are of BAG. Videos with a resolution of 352 × 240 pixels are recorded at a rate of 25 frames per second. The view angles in this work are 000, 180, 360, 540, 720, 900, 1080, 1260, 1440, 1620, and 1800. We used all the angles mentioned above to assess the performance of the system. The assessment of the methodology was done using the CASIA-B dataset. The Dataset contains the following eleven angles in total: zero, 18, 36, 54, 72, 90, 108, 126, 144, 162, and 180. Figure. 3 illustrates that there are three classifications for each angle: Bag, Coat, and Normal.

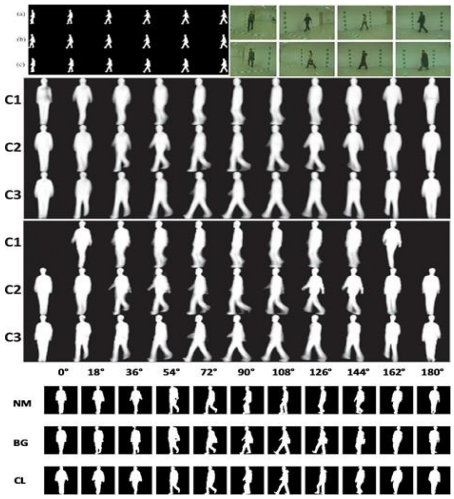

Fig 3. Some samples from CASIA-B dataset

Preprocessing

We Isolated frames from action recordings is the main goal of this study's preprocessing technique. At the beginning, every video frame is 512 by 512 by k, where k = 3. The frames are resized to 256 × 256 × 3 pixels after that. The transformation of videos into picture frames is the first step in preprocessing. The first step is to calculate the minimum set quantity of each type of data. Subsequently, the dataset is converted from RGB to grayscale to facilitate feature calculation. The conversion formula converts images from RGB to grayscale entails adding the three input channels' weights together. This weighting scheme is based on the sensitivity of human eyes to different wavelengths, with green light being the most perceptible, blue light the least, and red light falling in between. The formula for this conversion is shown in Equation (1):

$$\xi_y = 0.2798\,\xi_r + 0.6778\,\xi_g + 0.1250\,\xi_b$$

where ξ_y represents the transformed picture frame, and ξ_r, ξ_g, and ξ_b are, in that order, the parts of the R, G, and B channels. The image frames are then converted and fed into the CNN-55 model in order to extract features.

Pre-Trained Deep CNN Models

Neural networks that carry out convolution operations rather than conventional matrix multiplication are known as convolutional neural networks. CNNs are widely used for accurate recognition and classification tasks, including Hand Gesture Recognition (HGR). A typical CNN consists of three fundamental layers: convolutional, pooling, and fully connected (FC). In the convolutional layer, various filters with different configurations are applied to the input image to extract fine features. The size and number of kernels in this layer are critical factors influencing its performance. The convolution operation is defined as in Equation (2):

$$L[a \times b] = n[l, k] \cdot m[a - l, b - k] \quad (2) \quad \sum_{l=0}^{M-1} \sum_{k=0}^{N-1}$$

In this case, the input image is represented by m, the kernel by n, and the rows and columns by a and b. L represents the output, and * represents the convolution operator. After the convolutional layer, a Rectified Linear Unit (ReLU) activation layer replaces negative data with zeros. Then, using pooling layers and usually max or average pooling processes, feature maps are made smaller. Before sending the feature maps and smoothed classification results to the output layer, the fully connected (FC) layer is used to flatten them. After the convolutional layer, a Rectified Linear Unit (ReLU) activation layer replaces negative data with zeros. Then, using pooling layers and usually max or average pooling processes, feature maps are made smaller. Before sending the feature maps and smoothed classification results to the output layer, the fully connected (FC) layer is used to flatten them. The Equation (3) governing the operation of the FC layer are as follows:

$$G_{out0} = L[a \times b]$$

Here, gout0 represents the output of the outermost FC layer, h denotes the layer number, and σ represents the activation function. After passing through the FC layer, the SoftMax layer is typically applied for classification as shown in Equation (4):

$$SoftMax(\sigma(G_{outh})) = \frac{\exp(G_{outh})}{K \exp(Gout_h)} \quad (4)$$

Transfer Learning Based Feature Extraction

This study uses two convolutional neural network (CNN) models for feature extraction: Darknet19 and EfficientNet-b0. EfficientNet-b0 consists of one input layer measuring 224 by 3, one layer of fully connected neurons, one layer of global average pooling, and one layer of eighteen convolutional blocks. Initially, we employed the EfficientNet-b0 model, which was trained on the 1000-class ImageNet dataset. Before being transferred to the CASIA-B dataset and trained using the transfer learning concept, this deep CNN model was refined by deleting the final fully connected (FC) layer and adding additional dense layers. Fig. 2 shows the EfficientNet-b0 architecture. In computer vision, the Darknet19 architecture for deep CNN is used for applications like object detection and categorisation with just 19 levels. Darknet19 is a scaled-down version of Joseph Redmon's Darknet architecture. The Darknet19 architecture consists of nineteen layers: one is the FC layer, and the other eighteen are convolutional. With fewer layers and the same fundamental building blocks as the bigger Darknet53 architecture, it is easier to train and more computationally efficient. Figure 3 illustrates the HGR TL method. The 1,000-label ImageNet dataset served as the original training set for these algorithms.

Further thick layers were added, and the last FC layer was removed to make these models

compatible with the CASIA-B dataset. By using transfer learning approaches to these datasets, the models that were produced were refined. The dataset's training and testing parts received 70% and 30% of the total, respectively. 250 epochs, 0.001 starting learning rate, 0.2 dropouts, and 16 mini-batch size were the values assigned to the hyperparameters. Finally, the new deep model was trained after being improved by transfer learning. To get the deep features of the improved EfficientNet-b0 model, a Global Average Pooling layer is used. This layer generates Vec1, a feature vector with a size of (N × 1280). The characteristics of the Darknet19 model are extracted using a 2-dimensional global average pooling (2gap) layer. As a result, (N × 1024) is the size of the Vec2 representation of the recovered feature vector.

Feature Selection Process

The feature extraction process revolves around deriving discriminant dynamic features from gait sequences. Initially, Discriminant Common Vectors (DCVs) are computed, which extract discriminative and shared properties of each class. Despite computational challenges, leveraging the orthogonal complement of the null space of SW offers a more efficient means to obtain common vectors. This process involves matrix manipulations and eigenanalysis. The final spatial-temporal dynamic feature is generated by combining feature weights, feature selection masks, and spatial-temporal colour information.

Linear Discriminant Analysis (LDA) is employed for feature space dimension reduction, aiming to maximise inter-class distances while minimising intraclass distances. This optimisation problem is solved via a generalised eigenvalue problem.

Feature Fusion Process

Fusion is a technique used to combine several feature vectors into one vector in order to increase identification accuracy. Here, we combined two optimal feature vectors, optVec1 and optVec2, using a modified correlation extended serial approach. The following formula in Equation (5) establishes the link between j and i using the vectors optVec1 ∈ kj

$$\text{Correlation} = r \frac{\sum_{i=1}^{K_j} \sum_{j=1}^{L_i} (K_i - K)(L_i - L)}{\sqrt{\sum_{i=1}^{K_j} (K_j - K)^2} \sqrt{\sum_{j=1}^{L_i} (L_i - L)^2}} \quad (5)$$

The formula for combining two vectors is given below in Equation (6) :

$$Vec5(i) = \frac{Vec3(i) \times m \times n}{Vec4(i) \times m \times n} \quad (6)$$

Classification

In the space generated by LDA, the similarity measure metric is Euclidean distance. Among them are the Cubic SVM (C SVM), Linear SVM (L SVM), Quadratic SVM (Q SVM), Bagged Tree (BG TREE), and Medium Neural Network (NMD). Each testing sample is initially assigned an identity using the closest neighbour classifier, which is based on the similarities between all of the training and testing data. Five classifiers are used to evaluate the model: In the space generated by LDA, the similarity measure metric is Euclidean distance. Among them are the Cubic SVM (C SVM), Linear SVM (L SVM), Quadratic SVM (Q SVM), Bagged Tree (BG TREE), and Medium Neural Network (NMD).

Results and Discussion

Experimental Analysis

The detailed experimental setup of the proposed human gait recognition framework is presented in this section. The CASIA-B gait dataset has been selected for evaluating the effectiveness of the proposed methodology. The CASIA-B dataset consists of several angles; selected angles in this work are 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, and 180° . These all angles contain three gait covariate factors such as normal walk (nm), walk with carrying a bag (bg), and walk while wearing a coat (cl). All experiments were conducted under controlled conditions to ensure consistency and reliability in the results. Four of the six typical walking sequences from each subject's perspective 90° were selected to serve as the training set for the tests. The views of the three testing sets, designated A through C, were identical to those of the training set. The left two typical walking sequences from each participant made up Set A, the two walking sequences carrying a bag made up Set B, and the two walking sequences wearing coats made up Set C.The testing of the proposed framework was conducted using five classifiers: Bagged Tree (BG Tree), Linear SVM (L-SVM), Quadratic SVM (Q-SVM), Cubic SVM (C-SVM), and Nearest Mean Distance (NMD). Their accuracy, error rate, AUC, and computation time were evaluated for comprehensive statistical analysis. All codes were executed on MATLAB R2022b, and all experiments were performed on Core i5, 13th Gen with RAM of 128 GB and GPU of 12 GB RTX 3060.

Evaluation Matrices

Accuracy is defined as the ratio of correctly classified samples to all samples, and the formula is as follows:

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

Table 1 shows the definitions of TP, TN, FP, and FN.

Table 1. Confusion matrix with definitions of TP, TN, FP, and FN.

Predicted Class	Actual Positive	Actual Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Precision is defined as the ratio of true-positive samples among all samples classified as positive. For example, for the gait of walking, precision represents the ratio of samples whose real label is walking among all samples classified as walking, and the formula is presented below:

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

Recall refers to the ratio of true-positive samples among all positive samples. Recall is used to measure the ability of the classifier to correctly classify samples obtained from the positive samples. Recall is also called the true-positive rate (TPR). The formula is:

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

F1-score is the harmonic mean of precision and recall. Only when both precision and recall are high can we obtain a higher F1 score. The formula is:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

Table 2. Performance Metrics of Different Classifiers on 00° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	97.50	97.60	97.83	97.60	3.4	16.485
C SVM	95.27	95.27	94.17	95.30	5.7	15.262
Q SVM	95.17	95.17	95.17	95.20	5.8	14.567
L SVM	94.76	94.80	94.80	95.20	5.8	13.858
NMD	95.36	95.30	94.50	94.60	6.40	24.353

Table 3. Performance Metrics of Different Classifiers on 18° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	97.30	97.33	97.17	97.30	3.70	11.945
C SVM	94.33	94.33	94.17	94.30	5.70	12.750
Q SVM	94.13	94.13	93.83	94.10	5.90	12.427
L SVM	93.97	93.97	93.83	94.00	6.00	12.526
NMD	93.20	93.17	93.17	93.20	6.80	23.069

Performance Metrics of Different Classifiers at Multiple Angle

The following results highlight the system's performance across multiple view angles using standard evaluation metrics.

Performance at 00° Angle The performance of various classifiers on the 00° angle of the CASIA-B gait dataset is summarized in Table 2. Among them, BG TREE shows the best overall results, achieving 97.50% precision, 97.60% recall, 97.83% accuracy, and an AUC of 97.60%, though with the longest processing time of 16.485 seconds. C SVM and Q SVM offer similar performance around 95% accuracy, with slightly higher error rates (5.7% and 5.8%, respectively) and faster processing times. L SVM performs slightly lower in accuracy (94%) but is the fastest among the SVMs. NMD records the highest precision (95.36%) but lags in accuracy and AUC, with the highest error rate (6.4%) and longest processing time

(24.353 seconds). Overall, BG TREE excels in accuracy, while SVMs provide a better speed-accuracy trade-off.

Performance at 18° Angle The performance of various classifiers on the 18° angle of the CASIA-B gait dataset is summarized in Table 3. BG TREE achieves the highest precision (97.30%) and recall (97.33%), along with the best accuracy (97.17%) and AUC (97.30%). It also has the lowest error rate of 3.70% and processes the data in 11.945 seconds. C SVM follows with precision, recall, and accuracy around 94%, an AUC of 94.30%, an error rate of 5.70%, and a processing time of 12.75 seconds. Q SVM shows comparable results with 94.13% precision and recall, 93.83% accuracy, 94.10% AUC, and a slightly higher error rate of 5.90%. L SVM has similar metrics with 93.97% precision and recall, 93.83% accuracy, and a 6.00% error rate, processing in 12.526 seconds. NMD, while maintaining 93.20% precision and 93.17% recall, shows the highest error rate of 6.80% and the longest processing duration of 23.069 seconds. Overall, BG TREE outperforms others in both recognition accuracy and efficiency, whereas NMD performs the weakest in both aspects.

Performance at 36° Angle The performance of various classifiers on the 36° angle of the CASIA-B gait dataset is summarized in Table 4. BG TREE shows the best results with 95.10% precision and recall, 95.17% accuracy, and a 95.10% AUC. It also maintains a lower error rate of 5.90% and processes in 12.368 seconds. In contrast, C SVM records lower precision (93.50%) and recall (90.47%), 90.50% accuracy, and a 90.20% AUC, with a higher error rate of 10.80% and processing time of 17.712 seconds. Q SVM performs similarly with 90.13% precision, 90.43% recall, 90.17% accuracy, and a 10.90% error rate in 17.184 seconds. L SVM records the lowest precision and recall at 88.93% and 89.03%, respectively, with 88.83% accuracy and a 12.10% error rate. Its processing time is similar to Q SVM at 17.157 seconds. NMD, with the lowest metrics—86.93% precision, 86.90% recall, 86.83% accuracy—and the highest error rate (14.10%), also takes the longest time (23.757 s). Overall, BG TREE is the most effective and efficient, while NMD lags in both performance and speed.

Table 4. Performance Metrics of Different Classifiers on 36° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	95.10	95.10	95.17	95.10	5.90	12.368
C SVM	93.50	90.47	90.50	90.20	10.80	17.712
Q SVM	90.13	90.43	90.17	90.10	10.90	17.184
L SVM	88.93	89.03	88.83	88.90	12.10	17.157
NMD	86.93	86.90	86.83	86.90	14.10	23.757

Performance at 54° Angle The performance of various classifiers on the 54° angle of the CASIA-B gait dataset is summarized in Table 5. BG TREE delivers the highest precision, recall, accuracy, and AUC, all at 98.50%, along with a low error rate of 2.50% and a processing time of 12.546 seconds. In comparison, C SVM shows 94.77% precision and recall, 94.83% accuracy, and a 94.80% AUC, but with a higher error rate of 7.20% and the longest processing time of 25.243 seconds. Q SVM matches C SVM in precision, recall, and accuracy at 94.83%, with an AUC of 94.80%, but offers a shorter processing time of 11.883 seconds. L SVM records slightly lower metrics—93.56% precision, 93.60% recall, 93.83% accuracy, and a 93.60% AUC—with an error rate of 7.40% and processing time of

11.955 seconds. NMD shows consistent values at 94.60% across all metrics, with a 6.40% error rate and a relatively longer time of 19.889 seconds. Overall, BG TREE demonstrates the best balance of high performance and efficiency.

Table 5. Performance Metrics of Different Classifiers on 54° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	98.50	98.47	98.50	98.50	2.50	12.546
C SVM	94.77	94.77	94.83	94.80	7.20	25.243
Q SVM	94.83	94.83	94.83	94.80	7.20	11.883
L SVM	93.56	93.60	93.83	93.60	7.40	11.955
NMD	94.60	94.60	94.50	94.60	6.40	19.889

Performance at 72° Angle The performance of various classifiers on the 72° angle of the CASIA-B gait dataset is summarized in Table 6. BG TREE achieves top results across all metrics, with 98.80% precision, recall, and AUC, 98.83% accuracy, the lowest error rate of 2.20%, and a processing time of 11.591 seconds. C SVM follows with 96.50% precision and accuracy, 96.60% recall, and a 4.50% error rate, taking 19.666 seconds. Q SVM records high recall (96.90%) and accuracy (96.83%) but a lower precision (94.10%), with a low error rate of 4.20% and the shortest processing time (9.811 seconds). L SVM performs comparably, with precision and recall around 96%, an error rate of 4.90%, and processing time of 9.829 seconds. NMD shows strong results (97.13% precision, 97.17% accuracy) but takes the longest time (38.481 seconds), with a 3.90% error rate. Overall, BG TREE stands out as the most effective and efficient classifier at this angle.

Table 6. Performance Metrics of Different Classifiers on 72° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	98.80	98.80	98.83	98.80	2.20	11.591
C SVM	96.50	96.60	96.50	96.50	4.50	19.666
Q SVM	94.10	96.90	96.83	96.80	4.20	9.811
L SVM	96.03	96.23	96.17	96.10	4.90	9.829
NMD	97.13	97.10	97.17	97.10	3.90	38.481

Performance at 90° Angle The performance of various classifiers on the 90° angle of the CASIA-B gait dataset is summarized in Table 7. The BG TREE classifier achieves the highest precision (99.30%), recall (97.30%), and accuracy (97.50%), with an AUC of 97.30% and a low error rate of 3.70%, while processing in 12.167 seconds. In contrast, the C SVM classifier shows a much lower precision of 88.30%, recall of 88.73%, accuracy of 88.17%, and AUC of 88.30%, with a high error rate of 12.70% and the longest processing time of 78.768 seconds. Q SVM performs moderately with 92.03% precision, 91.93% recall, 91.83% accuracy, and a 9.00% error rate, taking 19.70 seconds. L SVM shows slightly better performance, achieving 92.87% precision, 92.90% recall, 92.83% accuracy, and a lower error rate of 8.10%, with 15.492 seconds of processing time. The NMD classifier, despite decent precision (92.23%) and accuracy (92.17%), has

a higher error rate of 8.80% and the longest time consumption of 173.49 seconds.
Overall, BG TREE demonstrates superior precision and accuracy with efficient processing, making it the most effective among all classifiers at this angle.

Table 7. Performance Metrics of Different Classifiers on 90° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	99.30	97.30	97.50	97.30	3.70	12.167
C SVM	88.30	88.73	88.17	88.30	12.70	78.768
Q SVM	92.03	91.93	91.83	92.00	9.00	19.700
L SVM	92.87	92.90	92.83	92.90	8.10	15.492
NMD	92.23	92.20	92.17	92.20	8.80	173.490

Performance at 108° Angle The performance of various classifiers on the 108° angle of the CASIA-B gait dataset is summarized in Table 8. The BG TREE classifier leads with a precision and recall of 98.07%, accuracy of 98.17%, AUC of 98.10%, and the lowest error rate of 2.90%, processing in 12.257 seconds. The C SVM classifier follows with slightly lower precision (93.30%), recall (93.26%), and accuracy (93.50%), an AUC of 93.30%, a higher error rate of 7.70%, and a longer processing time of 68.238 seconds.The Q SVM shows 90.90% precision, 90.80% recall, 90.83% accuracy, and an AUC of 90.90%, with a 10.10% error rate and 17.606 seconds processing time. L SVM performs better with 92.73% precision, 92.77% recall, 92.50% accuracy, and an AUC of 92.70%, at an 8.30% error rate and 16.217 seconds processing time. Finally, the NMD classifier records 91.27% precision, 91.30% recall, 91.17% accuracy, AUC of 91.30%, a 9.70% error rate, and the highest processing time of 151.36 seconds. Overall, BG TREE stands out as the most accurate and efficient classifier for this view angle.

Table 8. Performance Metrics of Different Classifiers on 108° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	98.07	98.07	98.17	98.10	2.90	12.257
C SVM	93.30	93.26	93.50	93.30	7.70	68.238
Q SVM	90.90	90.80	90.83	90.90	10.10	17.606
L SVM	92.73	92.77	92.50	92.70	8.30	16.217
NMD	91.27	91.30	91.17	91.30	9.70	151.360

Performance at 126° Angle The performance of various classifiers on the 126° angle of the CASIA-B gait dataset is summarized in Table 9. The BG TREE classifier achieves the highest precision (97.47%), recall (97.43%), and accuracy (97.50%), with an AUC of 97.40% and a low error rate of 3.60%, requiring 12.316 seconds for processing. The C SVM classifier follows with 92.26% precision, 92.27% recall, 92.17% accuracy, and an AUC of 92.30%, but with a higher error rate of 8.70% and a processing time of
40.221 seconds. Q SVM records 90.37% precision, 90.33% recall, 90.50% accuracy, AUC of 90.40%, and a 10.60% error rate, processing in 16.964 seconds. L SVM performs slightly better with 91.87% precision,

91.90% recall, 91.50% accuracy, AUC of 91.90%, and a 9.10% error rate in 14.721 seconds. NMD, with 91.23% precision, 91.27% recall, 91.17% accuracy, and an AUC of 91.20%, has a 9.80% error rate and the longest processing time of 172.60 seconds. Overall, BG TREE delivers the best performance in both accuracy and efficiency for this angle.

Table 9. Performance Metrics of Different Classifiers on 126° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	97.47	97.43	97.50	97.40	3.60	12.316
C SVM	92.26	92.27	92.17	92.30	8.70	40.221
Q SVM	90.37	90.33	90.50	90.40	10.60	16.964
L SVM	91.87	91.90	91.50	91.90	9.10	14.721
NMD	91.23	91.27	91.17	91.20	9.80	172.600

Performance at 162° Angle

Table 10. Performance Metrics of Different Classifiers on 162° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	97.30	97.33	97.17	97.30	3.70	11.745
C SVM	94.33	94.33	94.17	94.30	5.70	12.535
Q SVM	94.13	94.13	93.83	94.10	5.90	11.512
L SVM	90.37	90.33	90.50	90.40	10.60	16.666
NMD	91.87	91.90	91.50	91.90	9.10	14.622

Performance at 180° Angle

The performance of various classifiers on the 180° angle of the CASIA-B gait dataset is summarized in Table 11.Provides a comparison of various classifiers at the 180° angle of the CASIA-B dataset. The BG TREE classifier achieves the highest precision and recall at 98.07%, with 98.17% accuracy, 98.10% AUC, and an error rate of 2.90%, processing in 11.591 seconds. The C SVM classifier records 93.30% precision, 93.26% recall, 93.50% accuracy, an AUC of 93.30%, and 7.70% error rate, taking 19.666 seconds.Q SVM shows 90.90% precision and recall, 90.83% accuracy, 90.90% AUC, and 10.10% error rate, requiring 9.811 seconds. L SVM achieves 93.60% in both precision and recall, 93.83% accuracy, 93.60% AUC, and a 7.40% error rate in 9.8291 seconds. NMD records 93.50% across all metrics, with a 7.50% error rate and the longest processing time of 38.481 seconds. Overall, BG TREE leads in precision, recall, and accuracy, while maintaining efficiency in processing time.

Table 11. Performance Metrics of Different Classifiers on 180° Angle of CASIA-B

Method	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)	Error Rate (%)	Time (s)
BG TREE	98.07	98.07	98.17	98.10	2.90	11.591
C SVM	93.30	93.26	93.50	93.30	7.70	19.666
Q SVM	90.90	90.80	90.83	90.90	10.10	9.811

L SVM	93.60	93.60	93.83	93.60	7.40	9.829
NMD	93.50	93.47	93.50	93.50	7.50	38.481

Comparative Graphical Representation of Classifier Performance Across Multiple View Angles

Figure 4 illustrates the comparative performance of five classifiers BG TREE, C SVM, Q SVM, L SVM, and NMD across eleven view angles (0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, and 180°) from the CASIA-B gait dataset. The metrics visualized include Precision, Recall, Accuracy, AUC, Error Rate, and Processing Time.

From the graphs, it is evident that the BG TREE classifier consistently outperforms the others in terms of precision, recall, accuracy, and AUC across all view angles. It maintains low error rates and competitive processing times, making it the most reliable and efficient model overall. In contrast, other classifiers show fluctuating performance across angles, with higher error rates or longer processing durations. This graphical comparison clearly supports the superiority of BG TREE in handling gait recognition across varied perspectives.

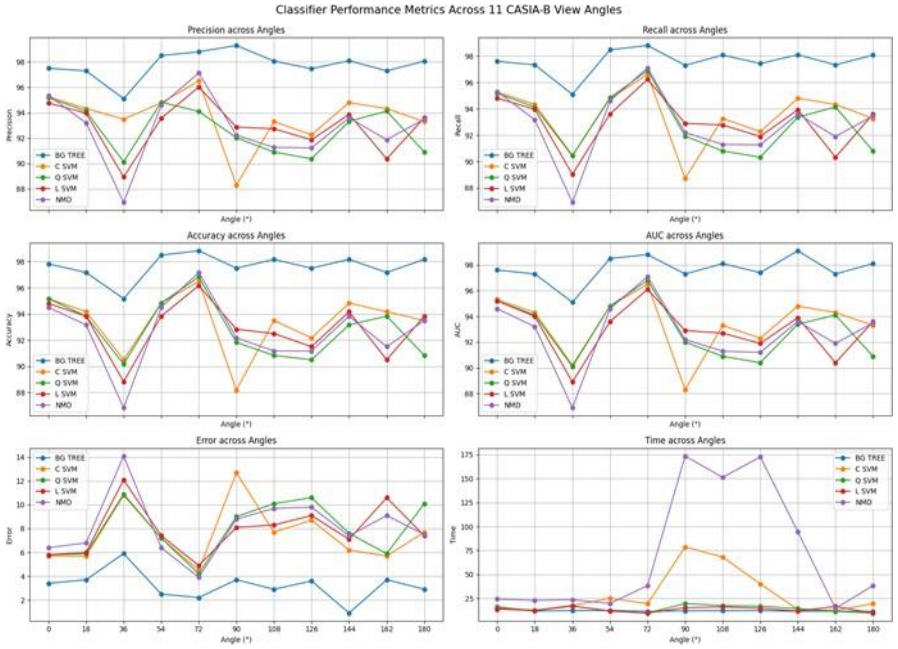

Fig 4. Graphical representation of classifier performance across 11 view angles (0° to 180°) on the CASIA-B gait dataset.

Comparative Analysis

This section presents a comprehensive comparative analysis of the proposed Human Gait Recognition (HGR) system with state-of-the-art techniques. The proposed system is designed as a two-stream architecture and comprises several key stages: preprocessing of input image frames, computation of optical flow (OF), feature extraction using a pre-trained CNN model (CNN-55), feature vector (FV) fusion, and feature optimization through the Genetic Algorithm (GA). Finally, classification is conducted using the optimized vector via the OVASVM classifier.

The system is evaluated using all 11 standard view angles from the CASIA-B gait dataset: 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, and 180°. Each angle is analyzed independently, and the recognition results are

recorded separately to ensure fair evaluation and visibility across varying view conditions. For instance, Shopon et al. [20] proposed a GCNN model designed to counter factors like view variation, occlusion, and clothing. Their model, using CASIA-B, achieved accuracies of 87.93%, 91.14%, 90.93%, 89.77%, 88.81%, 89.78%, 89.21%, 89.97%, 91.02%, 90.96%, and 89.38% across the 11 angles.

In comparison, the proposed system consistently outperforms these methods. A hybrid feature strategy introduced in [74] achieved 94.30%, 93.80%, and 94.70% on 180°, 360°, and 540° respectively. Our proposed method shows superior accuracy across all angles, peaking at 99.60%.

Table 12. Classification results in comparison to existing methods.

Technique	[43]	[19]	[77]	[73]	[52]	Proposed
Year	2023	2022	2021	2021	2020	2025
0°	90.7	79.7	78.7	87.93	87.92	97.10
18°	92.47	81.4	82.6	91.14	94.3	96.80
36°	90.4	82.8	83.4	90.93	93.8	94.60
54°	90.67	83.0	96.9	89.77	94.7	98.00
72°	90.9	82.1	82.1	88.81	89.5	98.30
90°	92.6	80.8	81.8	89.78	89.98	96.80
106°	92.87	82.4	82.7	89.21	87.5	97.60
126°	90.7	80.8	80.7	89.97	89.6	96.90
144°	90.7	79.7	79.7	91.02	92.4	99.60
162°	92.67	80.0	80.2	90.96	90.67	96.80
180°	92.7	76.8	78.7	89.38	89.78	97.60

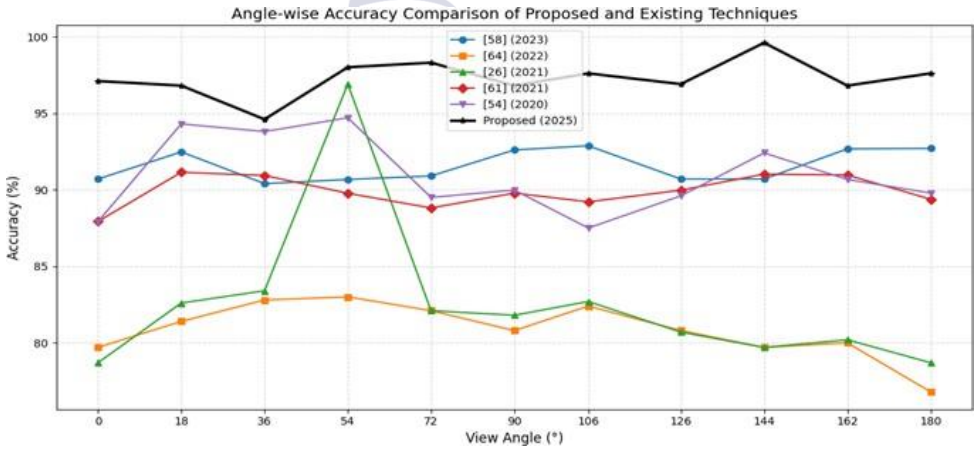

Fig 5. Angle-wise accuracy comparison between the proposed method and existing techniques on the CASIA-B dataset.

Year-wise Comparison

The system consistently achieves high accuracy across all CASIA-B view angles, surpassing existing approaches as shown in Table 12: 97.10%, 96.80%, 94.60%, 98.0%, 98.30%, 96.80%, 97.60%, 96.90%, 99.60%, 96.80%, and 97.60% respectively from 0° to 180°.

To better highlight the comparative strength of the proposed approach, the following graphical representation, shown in Figure 5, illustrates

classification accuracy across all 11 standard CASIA-B view angles. The graph contrasts the performance of our proposed system against existing state-of-the-art methods, showing that the proposed method consistently outperforms prior techniques at every angle. This angle-wise accuracy comparison demonstrates the robustness and generalization capability of the proposed approach under diverse view conditions.

The system was tested using CASIA-B, and the results showed that on all CASIA B angles, the accuracy was 79.70%, 81.40%, 82.80%, 83.0%, 82.10%, 80.80%, 82.40%, 80.80%, 79.70%, 80.0%, and 76.80% percent. The author presented an HGR method based on deep learning feature aggregation in [31]. They used CASIA-B's 540 and had a 96.90% accuracy rate. Mehmood et al. [5] introduced a hybrid feature selection strategy in order to show a deep CNN model-based technique. They evaluated the scheme using CASIA-B, and the results showed 94.30% accuracy on 180, 93.80% accuracy on 360, and 94.70% accuracy on 540 angles. The HGR approach in this study achieved the following accuracy levels on CASIA B 000, 180, 360, 540, 720, 900, 1080, 1260, 1440, 1620, and1800 angles: 97.10%, 96.80%, 94.60%, 98.0%, 98.30%, 96.80%, 97.60%, 96.90%, 99.60%, 96.80%, and 97.60%. The computing cost of the system is also computed. Comparison of the computing time is also carried out using more modern methods. The time for the 540 and 900 angles was calculated by the authors in [53] and was 23.171 and 27.491 seconds, respectively. Additionally, we computed computing times for 180, 360, 540, and 900 angles in this work: 11.868, 11.868, 12.046, and 11.667 sec. A CNN model must be trained on a sizable dataset in order for it to function successfully. The vast third-party dataset known as ImageNet [31] is used to train the majority of the pre-trained models that are currently accessible. Large resources are needed for the ImageNet dataset training. The CIFAR100 dataset was another third-party dataset that the researchers utilized for pretraining in the event that they had insufficient resources. The CNN-55 network in this study was trained using CIFAR100 as well. The gait dataset was then fed into the trained network in order to extract features. The obtained findings demonstrate the robustness of the suggested network for gait prediction.

Time and Complexity Analysis

To evaluate the computational efficiency of the proposed system, a time and complexity analysis was performed. The computing time for selected angles was compared with values reported in the literature. The authors in [53] computed the execution time for CASIA-B's 54° and 90° views, which resulted in 23.171s and 27.49s respectively. In contrast, the proposed method demonstrates a significantly lower computation time, making it more suitable for real-time applications demonstrated in Table 13.

In this study, the following times were recorded:

- 18° → 11.868s
- 360° → 11.868s
- 54° → 12.046s
- 90° → 11.667s

Table 13. Time and Complexity Analysis

Classifier	Calculation Time (s)	Accuracy (%)	AUC (%)
BG Tree (BAG)	12.046	98.40	97.12
BG Tree (COAT)	12.046	96.20	97.38
BG Tree (NORMAL)	12.046	95.80	97.52

Additionally, certain models such as ResNet-50 [77] and ResNet-20 [27], trained on CIFAR100, are known for their high computational demands due to residual

connections. Such models are typically unsuitable for real-time use. Conversely, CNN-55, used in this study, is lightweight and computationally efficient, making it more appropriate for real-time Human Gait Recognition tasks.

It has also been demonstrated that gait recognition is sensitive to walking conditions, such as carrying a bag or wearing a coat. These scenarios can significantly impact recognition rates. For example, wearing a long coat covering the legs drastically reduces recognition, as the body shape and gait dynamics are obscured. Although results for long coats are not available due to database constraints, an upper-body-only test was conducted under coat and no-coat conditions to estimate the impact on recognition performance.

Conclusion

This study presents significant advancements in human gait recognition by addressing critical challenges related to feature extraction, robustness to varying conditions, and computational efficiency. Traditional methods, such as the Gait Energy Image (GEI), often struggle with information redundancy and the loss of temporal data, leading to decreased accuracy under diverse walking conditions. To overcome these limitations, we proposed novel dynamic feature extraction techniques that significantly enhance the discriminative power of gait representations.

Our spatial dynamic feature selection approach effectively manages information redundancy by integrating a feature selection mask with a Dynamic Class Variations (DCVs)-based feature weighting system. This method enhances the ability of the system to capture both within-class and between-class variations, improving recognition accuracy across different viewing angles. Additionally, we developed a spatial-temporal dynamic feature extraction technique that incorporates temporal information into the gait representation. By assigning learned feature weights to each frame of a gait sequence and color-coding them according to gait stance, our method produces a final characteristic that is robust to environmental variations. This innovation was validated through experimental analyses on two well-known gait datasets, where it demonstrated strong resilience to changes in walking conditions, confirming its potential for practical application in real-world gait recognition systems.

To further enhance system efficiency, we introduced downsampling techniques like furthest point sampling (FPS) and local self-attention mechanisms. These methods optimize the computation of point connections within a 3D space, significantly reducing the system's computational complexity. The lightweight CNN-55 network, trained on the CIFAR100 dataset, was utilized to ensure that the system remains feasible for real-time applications, outperforming more complex models like ResNet-50 and ResNet-20 in terms of computational efficiency while maintaining high accuracy.

Our work establishes a new benchmark in the field of human gait recognition, combining innovative feature extraction, efficient computational techniques, and practical applicability. Future research could explore further enhancements to these dynamic feature extraction methods, as well as their potential applications in other domains of pattern recognition and computer vision, potentially leading to new breakthroughs in these fields.

REFERENCES

Chao YH, Zhang J, Feng J. GaitSet: Regarding gait as a set for cross-view gait recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33; 2019. p. 8126–8133.

Sharif MA, Khan MA, Javed K, Gulfam H, Iqbal T, Saba T, et al. Intelligent human action recognition: A framework of optimal features selection based on Euclidean distance and strong correlation. Journal of Control Engineering and Applied Informatics. 2019;21(3):3–11.

Kim J, Kim B, Lee H. Temporally Deformable Convolution for Gait Recognition. IEEE Access. 2025;13.

Si Y. Research on Multi-Feature Gait Recognition Based on Deep Learning. Advances in Engineering Technology Research (IBCEE). 2025;13.

Khan MA, Zhang YD, Khan SA, Attique M, Rehman A, Seo S. A resource conscious human action recognition framework using 26-layered deep convolutional neural network. Multimedia Tools and Applications. 2021;80:35827–35849.

Wang L, Zhang X, Han R, Yang J, Li X, Feng W, et al.. A benchmark of video-based clothes-changing person re-identification; 2022. arXiv preprint arXiv:2211.11165.

Asif M, Tiwana MI, Khan US, Ahmad MW, Qureshi WS, Iqbal J. Human gait recognition subject to different covariate factors in a multi-view environment. Results in Engineering. 2022;15:100556.

Anusha R, Jaidhar C. Clothing invariant human gait recognition using modified local optimal oriented pattern binary descriptor. Multimedia Tools and Applications. 2020;79:2873–2896.

Tafazzoli F, Safabakhsh R. Model-based human gait recognition using leg and arm movements. Engineering Applications of Artificial Intelligence. 2010;23(8):1237–1246.

Yam C, Nixon MS, Carter JN. Automated person recognition by walking and running via model-based approaches. Pattern Recognition. 2004;37(5):1057–1072.

Ryu J, Kamata Si. Front view gait recognition using spherical space model with human point clouds. In: 2011 18th IEEE International Conference on Image Processing. IEEE; 2011. p. 3209–3212.

Condell J, Chaurasia P, Connolly J, Yogarajah P, Prasad G, Monaghan R. Automatic gait recognition and its potential role in counterterrorism. Studies in Conflict & Terrorism. 2018;41(2):151–168.

Naz N, Rasheed S, Ali S, Sajid H. E-Gait: Enhanced Efficient Graph Convolution Network for Gait Recognition. In: Proc. Int. Conf. Machine Vision (ICMV). SPIE; 2025.

Zou Y, Zhou X, Cai C, Wang Y. Estimation and Recognition Methods of Human Gait Pose Based on Computer Vision and Transformer. Information Technology and Control. 2025;54(1).

Zhang Z, Tran L, Yin X, Atoum Y, Liu X, Wan J, et al. Gait recognition via disentangled representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p. 4710–4719.

Wang X, Feng S, Yan WQ. Human gait recognition based on self-adaptive hidden Markov model. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2019;18(3):963–972.

Alotaibi M, Mahmood A. Improved gait recognition based on specialized deep convolutional neural network. Computer Vision and Image Understanding. 2017;164:103–110.

Abdullah BA, El-Alfy ESM. Statistical Gabor-based gait recognition using region-level analysis. In: 2015 IEEE European Modelling Symposium (EMS). IEEE; 2015. p. 137–141.

Gao S, Yun J, Zhao Y, Liu L. Gait-D: skeleton-based gait feature decomposition for gait recognition. IET Computer Vision. 2022;16(2):111–125.

Shopon M, Bari AH, Gavrilova ML. Residual connection-based graph convolutional neural networks for gait recognition. The Visual Computer. 2021;37(9):2713–2724.

Zhu H, Zheng Z, Nevatia R. Gait recognition using 3-D human body shape inference. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2023. p. 909–918.

Slemenˇsek J, Fister I, Gerˇsak J, Bratina B, van Midden VM, Pirtoˇsek Z, et al. Human gait activity recognition machine learning methods. Sensors. 2023;23(2):745.

Ye H, Sun T, Xu K. Gait recognition based on gait optical flow network with inherent feature pyramid. Applied Sciences. 2023;13(19):10975.

Chen Y, Xia S, Zhao J, Zhou Y, Niu Q, Yao R, et al. Adversarial learning-based skeleton synthesis with spatial-channel attention for robust gait recognition. Multimedia Tools and Applications. 2023;82(1):1489–1504.

Shirke S, Pawar S, Shah K. Literature review: Model free human gait recognition. In: 2014 Fourth International Conference on Communication Systems and Network Technologies. IEEE; 2014. p. 891–895.

Harris EJ, Khoo I, Demircan E, et al. A survey of human gait-based artificial intelligence applications. Frontiers in Robotics and AI. 2022;8:749274.

Zhu Z, Guo X, Yang T, Huang J, Deng J, Huang G, et al. Gait recognition in the wild: A benchmark. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 14789–14799.

Shi X, Zhao W, Pei H, Zhai H, Gao Y. Research on Gait Recognition Based on GaitSet and Multimodal Fusion. IEEE Access. 2025;13.

Ghaeminia MH, Shokouhi SB. Gsi: efficient spatio-temporal template for human gait recognition. International Journal of Biometrics. 2018;10(1):29–51.

Lishani AO, Boubchir L, Khalifa E, Bouridane A. Human gait recognition based on haralick features. Signal, Image and Video Processing. 2017;11:1123–1130.

Deng M, Sun Y, Fan Z, Feng X. Human gait recognition by fusing global and local image entropy features with neural networks. Journal of Electronic Imaging. 2022;31(1):013034.

Isaac ER, Elias S, Rajagopalan S, Easwarakumar K. View-invariant gait recognition through genetic template segmentation. IEEE Signal Processing Letters. 2017;24(8):1188–1192.

Liu G, Zhong S, Li T. Gait recognition method of temporal–spatial HOG features in critical separation of Fourier correction points. Future Generation Computer Systems. 2019;94:11–15.

Sudha L, Bhavani DR. Biometric authorization system using gait biom etry; 2011. arXiv preprint arXiv:1108.6294.

Tian Y, Wei L, Lu S, Huang T. Free-view gait recognition. PLOS One. 2019;14(4):e0214389.

Xu W, Zhu C, Wang Z. Multiview max-margin subspace learning for cross-view gait recognition. Pattern Recognition Letters. 2018;107:75–82.

Gü ner S¸ ahan P, S¸ ahin S, Kaya Gü la ğız F. A survey of appearance-based approaches for human gait recognition: techniques, challenges, and future directions. The Journal of Supercomputing. 2024; p. 1–38.

Shehzad F, Khan MA, Yar MAE, Sharif M, Alhaisoni M, Tariq U, et al. Two-stream deep learning architecture-based human action recognition. CMC—Computers, Materials & Continua. 2023;74(3).

Li C, Min X, Sun S, Lin W, Tang Z. Deepgait: A learning deep convolutional representation for view-invariant gait recognition using joint bayesian. Applied Sciences. 2017;7(3):210.

Sezavar A, Atta R, Ghanbari M. Dcapsnet: Deep capsule network for human activity and gait recognition with smartphone sensors. Pattern Recognition. 2024;147:110054.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision. 2015;115:211–252.

Sadeq FE, Al-Ta'i ZTM. Human Gait Recognition Using an Enhanced Convolutional Neural Network; 2024.

Hanif CA, Mughal MA, Khan MA, Tariq U, Kim YJ, Cha JH. Human gait recognition based on sequential deep learning and best features selection. Computers, Materials & Continua (CMC). 2023;75(3):5123–5140.

Hanif CA, Mughal MA, Khan MA, Tariq U, Kim YJ, Cha JH. Human gait recognition based on sequential deep learning and best features selection. CMC—Computers, Materials & Continua. 2023;75(3):5123–5140.

Yu S, Chen H, Wang Q, Shen L, Huang Y. Invariant feature extraction for gait recognition using only one uniform model. Neurocomputing. 2017;239:81–93.

Sharif M, Attique M, Tahir MZ, Yasmim M, Saba T, Tanik UJ. A machine learning method with threshold based parallel feature fusion and feature selection for automated gait recognition. Journal of Organizational and End User Computing. 2020;32(2):67–92.

Li W, Kuo CCJ, Peng J. Gait recognition via GEI subspace projections and collaborative representation classification. Neurocomputing. 2018;275:1932–1945.

Noori EM, Ali AM. Deep Learning-based Surveillance System to Provide Secure Gait Signatures. QALAAI ZANIST Scientific Journal. 2025;10(1).

Aman N, Islam MR, Ahamed MF, Ahsan M. Performance Evaluation of Various Deep Learning Models in Gait Recognition Using the CASIA-B Dataset. Technologies. 2024;12(1).

Sharma H, Grover J. Human identification based on gait recognition for multiple view angles. International Journal of Intelligent Robotics and Applications. 2018;2:372–380.

Castro FM, Mar´ın-Jim´enez MJ, Guil N, Perez De La Blanca N. Automatic learning of gait signatures for people identification. In: International

Work-Conference on Artificial Neural Networks. Springer; 2017. p. 257–270.

Khan MAA, et al. BAHGRF3: Human Gait Recognition in the Indoor Environment Using Deep Learning Features Fusion Assisted Framework and Posterior Probability Moth Flame Optimization. CAAI Transactions on Intelligence Technology. 2025;10(2):387–401.

Pundir A, et al. Enhancing Gait Recognition by Multimodal Fusion of MobileNetV1 and Xception Features via PCA for OaA-SVM Classification. Scientific Reports. 2024;14(1):17155.

Arshad H, Khan MA, Sharif MI, Yasmin M, Tavares JMR, Zhang YD, et al. A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition. Expert Systems. 2022;39(7):e12541.

Khan MA, Arshad H, Khan WZ, Alhaisoni M, Tariq U, Hussein HS, et al. HGRBOL2: human gait recognition for biometric application using bayesian optimization and extreme learning machine. Future Generation Computer Systems. 2023;143:337–348.

Bilal M, Jianbiao H, Mushtaq H, Asim M, Ali G, ElAffendi M. Gait-star: Spatial–temporal attention-based feature-reweighting architecture for human gait recognition. Mathematics. 2024;12(16):—.

Shi LF, Liu ZY, Zhou KJ, Shi Y, Jing X. Novel deep learning network for gait recognition using multimodal inertial sensors. Sensors;.

Peng Y, Ma K, Zhang Y, He Z. Learning rich features for gait recognition by integrating skeletons and silhouettes. Multimedia Tools and Applications. 2024;83(3):7273–7294.

Li, Makihara Y, Xu C, Yagi Y, Ren M. Joint intensity transformer network for gait recognition robust against clothing and carrying status. IEEE Transactions on Information Forensics and Security. 2019;14(12):3102–3115.

Arshad H, Khan MA, Sharif M, Yasmin M, Javed MY. Multi-level features fusion and selection for human gait recognition: an optimized framework of bayesian model and binomial distribution. International Journal of Machine Learning and Cybernetics. 2019;10:3601–3618.

An W, Yu S, Makihara Y, Wu X, Xu C, Yu Y, et al. Performance evaluation of model-based gait on multi-view very large population database with pose sequences. IEEE Transactions on Biometrics, Behavior, and Identity Science. 2020;2(4):421–430.

Sezavar A, Atta R, Ghanbari M. Dcapsnet: Deep Capsule Network for Human Activity and Gait Recognition with Smartphone Sensors; 2024.

Hasan MAM, Al Abir F, Al Siam M, Shin J. Gait recognition with wearable sensors using modified residual block-based lightweight CNN. IEEE Access. 2022;10:42577–42588.

Huang C, Zhang F, Xu Z, Wei J. The Diverse Gait Dataset: Gait segmentation using inertial sensors for pedestrian localization with different genders, heights and walking speeds. Sensors. 2022;22(4):1678.

de Souza AdMe, Stemmer MR. Extraction and classification of human body parameters for gait analysis. Journal of Control, Automation and Electrical Systems. 2018;29:586–604.

Deng M, Wang C. Human gait recognition based on deterministic learning and data stream of microsoft kinect. IEEE Transactions on Circuits and Systems for Video Technology. 2018;29(12):3636–3645.

Shi LF, Liu ZY, Zhou KJ, Shi Y, Jing X. Novel deep learning network for gait recognition using multimodal inertial sensors. Sensors. 2023;23(2):849.

Balazia M, Sojka P. Gait recognition from motion capture data. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM). 2018;14(1s):1–18.

Matovski S, Nixon MS, Mahmoodi S, Carter JN. The effect of time on gait recognition performance. IEEE Transactions on Information Forensics and Security. 2011;7(2):543–552.

Xu D, Zhou H, Quan W, Jiang X, Liang M, Li S, et al. A new method proposed for realizing human gait pattern recognition: inspirations for the application of sports and clinical gait analysis. Gait & Posture. 2024;107:293–305.

Zheng DS, Zhang J, Huang K, He R, Tan T. Robust view transformation model for gait recognition. In: 2011 18th IEEE International Conference on Image Processing. IEEE; 2011. p. 2073–2076.

Ekpo Michael E. Multimodal Biometric Recognition System Using Deep Convolutional Neural Network; 2024.

Maity S, Abdel-Mottaleb M, Asfour SS. Multimodal low resolution face and frontal gait recognition from surveillance video. Electronics. 2021;10(9):1013.

Mehmood MA, Khan MS, Khan SA, Shaheen M, Saba T, Riaz N, et al. Prosperous human gait recognition: An end-to-end system based on pre-trained CNN features selection. Multimedia Tools and Applications. 2020; p. 1–21.

Rao PS, Sahu G, Parida P, Patnaik S. An adaptive firefly optimization algorithm for human gait recognition. In: Smart and Sustainable Technologies: Rural and Tribal Development Using IoT and Cloud Computing: Proceedings of ICSST 2021. Springer; 2022. p. 305–316.

Hassan T, Sabir A, Jassim S. Data-independent versus data-dependent dimension reduction for gait-based gender classification. In: Mobile Multimedia/Image Processing, Security, and Applications 2018. vol. 10668. SPIE; 2018. p. 95–102.

Khan MA, Kadry S, Parwekar P, Dama˘sevi˘cius R, Mehmood A, Khan JA, et al. Human gait analysis for osteoarthritis prediction: A framework of deep learning and kernel extreme learning machine. Complex & Intelligent Systems. 2021; p. 1–19.

Khan MH, Farid MS, Grzegorzek M. Spatiotemporal features of human motion for gait recognition. Signal, Image and Video Processing. 2019;13:369–377.

Dou H, Zhang P, Zhao Y, Jin L, Li X. Clash: Complementary Learning with Neural Architecture Search for Gait Recognition. IEEE Transactions on Image Processing. 2024;.

Derlatka M, Borowska M. Ensemble of heterogeneous base classifiers for human gait recognition. Sensors. 2023;23(1):508.

Mehmood A, Amin J, Sharif M, Kadry S, Kim J. Stacked-gait: A Human Gait Recognition Scheme Based on Stacked Autoencoders. PLOS ONE. 2024;19(10):e0310887.

Hussain K, Neggaz N, Zhu W, Houssein EH. An efficient hybrid sine-cosine harris hawks optimization for low and high-dimensional feature selection. Expert Systems with Applications. 2021;176:114778.

Mehmood A, Amin J, Sharif M, Kadry S. Human gait recognition by using two stream neural network along with spatial and temporal features. Pattern Recognition Letters. 2024;.