CHAPTER 4

CNN-Based Gait recognition irrespective of covariate conditions

4.1 Introduction

One approach to gait recognition is through the use of Convolutional Neural Networks (CNNs) models. CNN is a type of neural network commonly used in computer vision tasks, such as image classification and object detection. They consist of several layers of neurons, including convolution layers, pooling layers, and fully connected layers. The convolution layers apply a set of learnable filters to the input data, extracting features at different spatial scales. The pooling layers then reduce the spatial resolution of the feature maps, making them more manageable for further processing. Finally, the fully connected layers use the extracted features to classify the input data into one or more classes.

In gait recognition, the CNN model can be trained to learn discriminative features from a set of gait sequences. A gait sequence is typically represented as a series of consecutive frames, each containing an image of the subject at a different point in their walking cycle. The CNN can then be used to classify the gait sequence into one of several classes, such as the identity of the subject or their walking condition (e.g., normal walking, limping, etc.).

To train a CNN-based gait recognition system, a large dataset of gait sequences is required. This dataset should include a diverse set of subjects with varying walking styles, as well as different walking conditions. The CNN can be trained using supervised learning, where the gait sequences are labeled with their corresponding class labels. Alternatively, unsupervised learning techniques can be used, where the CNN learns to cluster similar gait sequences together without explicit class labels.

Overall, CNN-based gait recognition has shown promising results in recent years and has the potential to be a useful biometric tool in a variety of applications, such as security systems, healthcare, and sports analysis.

This chapter investigates and proposes human gait recognition irrespective of covariate conditions using a CNN model and feature fusion technique. The study aims to overcome the limitations of existing CNN-based gait recognition methods by making the recognition process more robust to covariate conditions, which are factors that can affect the appearance of a person's gait, such as clothing, shoes, and walking surface. Additionally, a Feature fusion-based method is developed for gait recognition irrespective of covariate conditions. The CASIA-B standard dataset is used to perform analyses of the proposed methods and compare them to existing methods.

4.2 Related work

Gait recognition under covariate conditions for real-time applications poses a significant challenge. Changes in an individual's appearance, such as carrying an object or wearing different clothing, can notably complicate gait recognition. Researchers have traditionally concentrated on gait analysis for human identification through the recognition of a person's normal walk or adjustments in viewing angles. The recent studies on gait recognition indicate the potential to develop an efficient methodology that enhances recognition performance in the presence of covariates.

A convolutional neural network (CNN) and an evaluation method are proposed to analyze gait in covariate conditions for human identification. The performance of the proposed method is compared under various covariate conditions, such as viewing angle, clothing, and carrying objects [104]. In work on multiview gait recognition by Hu Ng et al [57]found that combining static and dynamic gait features with a Gaussian filter can reduce the effect of outliers and improve person identification accuracy to 92.1%. Shiqi et al [78] proposed a view-invariant gait feature extraction using a view transformation model based on auto-encoders. The model is robust to variations in viewing angle, clothing, and carrying conditions, and can greatly improve recognition rate in surveillance applications. Chao et al [14] proposes a novel approach for human identification from a set of discrete frames. The approach views a gait as a collection of discrete frames and then uses a new network called GaitSet to identify a human from the set. Liu et al [49] Gait-based gender recognition using two CNN models with CASIA B dataset is proposed. Ahmed R Hawas et al [?] proposes a gait-based human identification method using convolutional neural networks (CNNs) and optical flow Gait Energy Images (GEIs). The method achieved over 95% accuracy on a dataset of gait sequences with varying view angles. The results of this work suggest that it is promising to explore more gait patterns and feature representations in order to improve gait recognition performance. In 2021, Pose-based gait feature extraction using the GAN model to reduce covariates issues in silhouette images for efficient gait recognition is proposed [28]. The work mentions how pose analysis is performed using pose energy images to label the frames with a specific key pose. After the silhouette frames are labeled, the GAN model is trained by pairing them with the recognition of the same pose. Furthermore, the Covariate Factors Omitted Silhouette Image (CFOSI) a novel gait invariant feature is generated by this trained GAN model. The findings strongly show that the presence of covariates in silhouette images can effectively address covariate difficulties in gait identification, and that features fusion can be useful in doing so. Also, the recent work by Sanjay et al [29] suggests an effective gait feature template representation technique termed the DGEI(Dynamic Gait Energy Image). The extracted gait features are employed on a Generative Adversarial Network (GAN) to predict the corresponding DGEI templates without the covariates. This work could be useful for comparing different gait template costs and complexity [68]. The performance of CNN-based gait recognition systems can be evaluated using several metrics, such as accuracy, precision, recall, and F1-score. These metrics measure how well the system can correctly identify the subjects in the gait sequences and distinguish them from other subjects.

State-of-the-art CNN-based gait recognition systems typically use deep neural networks with several layers, such as (Residu)al Network(modelResNet), Inception model, or DenseNet model. These networks are trained on large datasets of gait sequences, such as the CASIA-B dataset, the OU-ISIR dataset, or the BIWI Walking Pedestrians dataset.

In recent years, several studies have reported high accuracy rates for CNN-based gait recognition systems. For example, a study by Liu et al. (2021) achieved an accuracy of 99.2% on the CASIA-B dataset using a deep ResNet-based network. Another study by Wu et al. (2020) achieved an accuracy of 96.1% on the OU-ISIR dataset using a DenseNet-based network.

However, it is important to note that the performance of CNN-based gait recognition systems can be affected by various factors, such as the quality of the gait sequences, the number of subjects in the dataset, and the variation in walking conditions. In addition, CNN-based gait recognition systems may also suffer from the "intra-subject" and "inter-view" variations, where the same subject can have different walking patterns under different conditions or captured from different angles. Therefore, further research is needed to address these challenges and improve the robustness and generalizability of CNN-based gait recognition systems.

4.3 Proposed CNN model for human gait identification irrespective of covariate issues

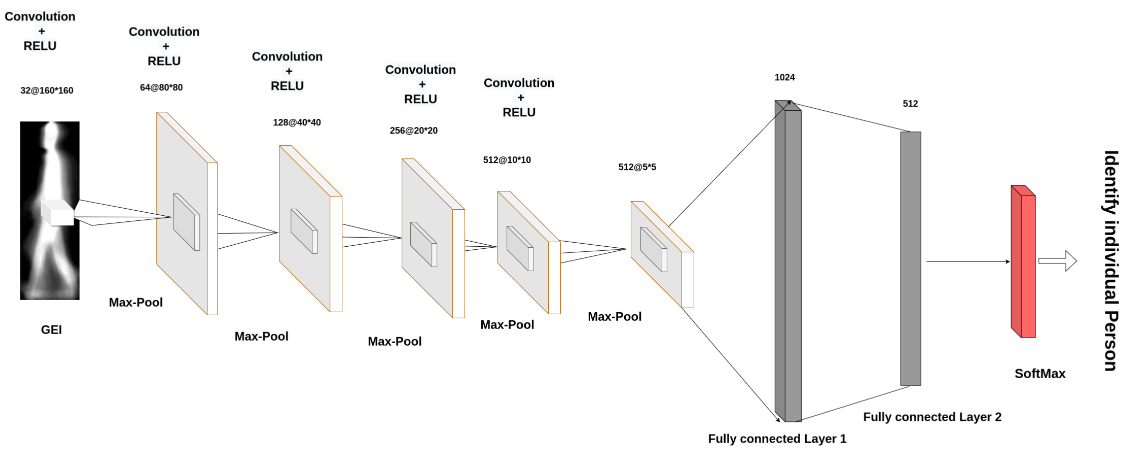


Figure 4.1: CNN with GEI as input

The proposed CNN model is based on LeNet architecture with more layers. The LeNet architecture is a multi-layer convolution neural network for image classification and was mainly developed for handwritten recognition problems [41]. This architecture is simple and straightforward to implement. The network is called LeNet-5 because it consists of five layers with learnable parameters. It utilizes a combination of average pooling, three sets of convolution layers, and two fully connected layers. Lastly, a Softmax classifier is employed to categorize the images into their respective classes. A recent work by Shao et al. [77] using LeNet architecture towards improving the architecture with residual activation function to reduce the influence of covariates condition in gait recognition is seen in the studies.

The work on using deeper networks towards improving the performance of network models by increasing the size of the network by Szegedy et al [83], it seen that more chains of layers and dense layers with dropout could be used for solving a complex problem. The idea behind proposing the CNN architecture for gait classification in covariate was to improve the performance of recognition rate affected by variable factors in gait patterns. Also to analyse human identification irrespective of covariates issues. Two gait representations are used to examine the proposed CNN architecture, and the framework is illustrated in Figure 4.1 and Figure 4.2.

The proposed deep CNN architecture has five convolution layers, and five subsampling (max-pooling) layers, two fully connected (dense layer), one dropout layer, and input and output. For input, we use GEI and GGMI gait templates of image size 160x160 respectively. We have considered a kernel size of 5x5 for the input layer and for the remaining layers of kernel size 3x3. The filter of sizes 32, 64, 128, 256, and 512 are used for feature mapping. The subsampling(max-pool) layer has a 2x2 pixel window size and stride size of two. The subsampling layer reduces the feature maps to half size of the input in each convolution layer. The model has a ReLU activation function and two fully connected (dense layer) of 1024 and 512 nodes with a dropout of 0.5 in between the first and the second fully connected layer, and a dropout of 0.2 is added before the classification task to avoid overfitting on the training data. A regularizer of L2 kernel with a learning rate of 0.0001 is added to both the dense layers to prevent over-fitting caused by large weights. The output layer in the model has a soft-max function with a categorical cross-entropy loss function. Adam optimizer algorithm [40] is used for training and testing the proposed model with a learning rate of 0.0001. The analysis of the model is performed for four different mini-batch sizes of 8,16,32 and 64 respectively. The proposed architecture is trained upto 100 epochs. The train and test set were randomly split into training(80%) and validation(20%) to train the network. The CASIA-B, a standard dataset is used for training and testing ton the proposed architecture. It is also observed from our analysis that out of the four mini batch-size, a good training and testing accuracy rate is achieved using mini batch-sizes for both gait representation methods. Consequently, the training and validation losses are also stable with larger mini-batch sizes.

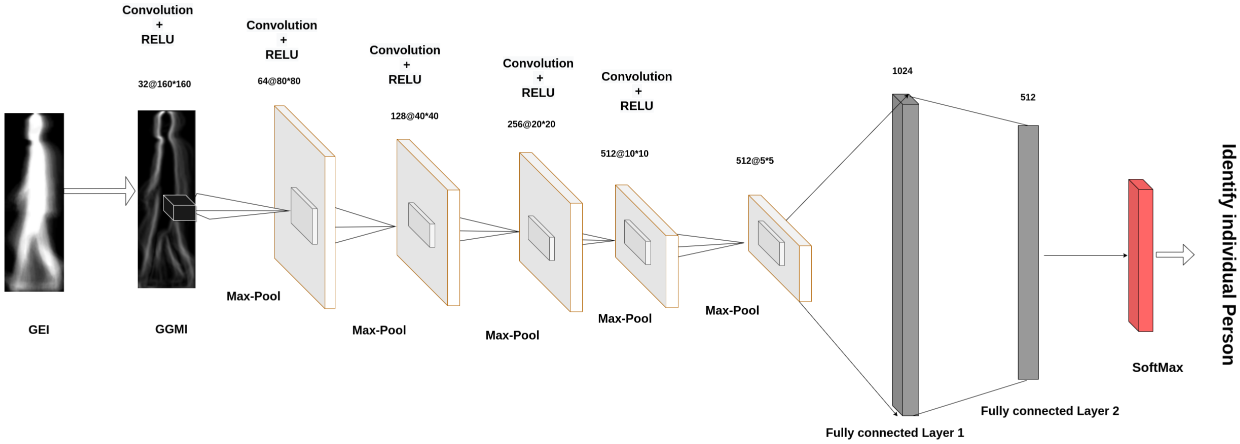


Figure 4.2: CNN with GGMI as input

4.4 EXPERIMENT ANALYSIS AND RESULTS

4.4.1 Dataset Arrangement

The CASIA B dataset of 124 subjects was used for the experimental analysis, and the GEI templates generated from silhouette sequences of each subject were kept irrespective of the covariates conditions. The CASIA Dataset-B is a multi-view gait database created in 2005, it has 124 subjects captured from 11 different angles [93]. The gait sequences of each person are collected in 10 perspectives, consisting of six usual walking sequences as Nm1, Nm2, Nm3, Nm4, Nm5 and Nm6, two sequences of walking with carrying a bag as Bg1 & Bg2, and clothing sequences as Cl1 & Cl2. The walking variation under different covariates conditions from a single view angle is shown in Figure 4.3.

(a) Normal walk (b) With jacket (c) With a bag

Figure 4.3: CASIA B Dataset 90 degree angle

4.4.2 Experimental analysis

4.4.2.1 Training and Validation analysis

GEI has the tendency to lose the temporal information from averaging the binary silhouette sequence. Further, the GEI template is used for extracting the GGMI templates to understand the GEI characteristic feature. Accordingly, the GEI and GGMI have a total image of 8150 of 124 subjects. The training set is randomly split into training(80%) and validation(20%) set to train the network. The subjects are stored in a training database irrespective of viewing angles covering all the covariates conditions. So the first training database consists of Nm1, Nm2, Nm3, Nm4, Cl1, and Bg1, and the remaining were kept for testing. The dataset is trained and validated irrespective of viewing angles on 8,16,32, and 64 batch-size for 200 epochs. The results obtained for different batch-size using the two gait representation methods mentioned in the task are shown in table 4.1 and table 4.2

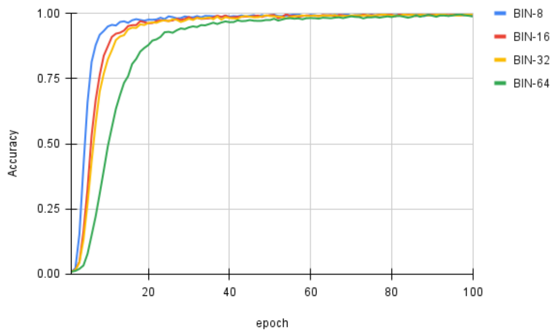
Table 4.1: Accuracy with GEI templates

Batch-Size	Train accuracy(%)	Validation accuracy(%)
8	99.34	97.20
16	99.72	97.39
32	99.86	97.91
64	98.77	98.33

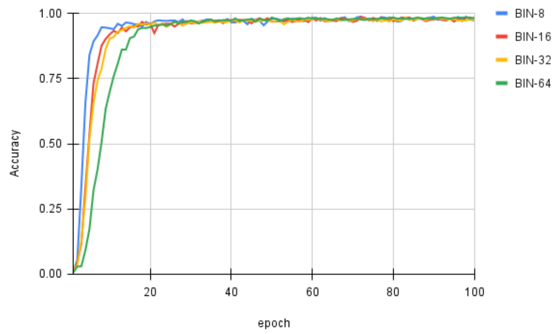
Table 4.2: Accuracy with GGMI templates

Batch-Size	Train accuracy(%)	Validation accuracy(%)
8	99.73	97.35
16	99.42	96.76
32	99.65	97.49
64	99.91	98.60

Using GGMI gait representation methods, the proposed CNN architecture performs better with 64 mini-batch size, with good train accuracy. It is also observed from various works on gait recognition that adopting GEI, the proposed CNN model gives improved results on using the GEI template, comparison is shown in Table 4.4. However, the GGMI gait representation is found to be more effective for handling covariates issues in gait patterns, with promising accuracy for both training and testing. The objective of combining GEI with GGMI is to capture both spatial and gradient information, reducing the issue of covariates in gait recognition algorithms. Accordingly, the training accuracy and validation accuracy using GEI gait representation with 8,16,32, and 64 mini-batch size is shown in Figure 4.4. Similarly, the results for GGMI are shown in Figure 4.5.

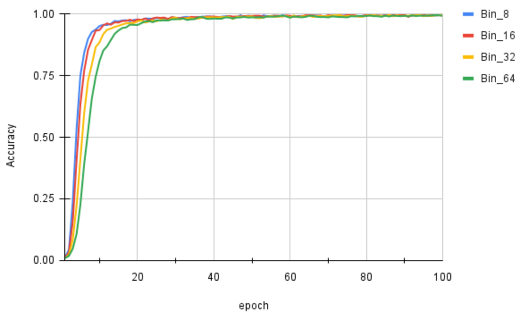


(a) GEI train accuracy

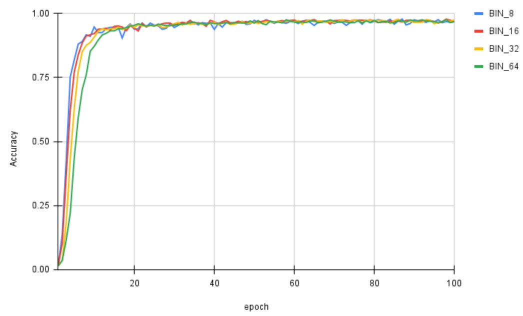


(b) GEI validation accuracy

Figure 4.4: GEI train accuracy and validation accuracy



(a) GGMI train accuracy



(b) GGMI validation accuracy

Figure 4.5: GGMI train accuracy and validation accuracy

4.4.2.2 Testing analysis

For the testing task, a test set was generated utilizing the remaining covariates—NM5 and NM6 for normal walking, Cl-02 and BG-02 for clothing and carrying conditions—as previously specified in the training and validation section. We proceeded to test this test set against the training set and observed a remarkable accuracy rate for gait recognition, regardless of the covariate conditions. This chapter presents a methodology for training and testing the dataset, regardless of viewing angles, using a single known covariate from the training database. Consequently, testing was conducted with the assumption that the CNN model incorporates at least one known covariate, and only known variables were tested against the training set. Table 4.3 displays the average test accuracy, encompassing all covariate considerations.

Table 4.3: Testing accuracy for all covariate conditions

Method	NM	CL	BG	Average
GEI	98.35	95.80	96.39	96.84
GGMI	99.27	96.19	95.88	97.29

The proposed method is compared with existing method proposed by Alotobi et.al(2017), Liu et,al(2018), Hawas. et.al (2019), Shao et.al (2020) and Xiaofang Wu et.al(2020). Our proposed technique produces better results irrespective of covariate factors. The comparison table is shown in Table 4.4. According to the findings of the analysis, our proposed model outperformed the other methods for both the gait representation methods.

Table 4.4: Comparision table with recent works on gait recognition in covariates

Author	Gait representation	Dataset	Covariates	Test Accuracy
altobi etl [2]	GEI	CASIA-B	singleview	92
Liu et al. [49]	GEI	CASIA-B	no covariates	89.62
Shao et al[77]	Silhouette	CASIA-B	cloating carrying normal walk	94.43
Hawas et al. [33]	optical flow	CASIA-B	cross view	95
Xiaofang Wu et al. [96]	GEI	CASIA-B	no covariates	95.96
Proposed method	GEI GGMI	CASIA-B CASIA-B	irrespective of viewing covariates	96.84 97.29

The main objective of this task is on gait representation and gait classification in covariate conditions. The effectiveness of the GEI gait representation for gait-based human identification has been demonstrated in numerous prior research. We have integrated GEI and GGMI the existing gait representation techniques in this research work for more effective human gait analysis with covariates issues. A CNN architecture for gait recognition irrespective of covariates condition using basic parameter tuning is discussed. The average recognition rate covering all the covariates conditions with the proposed CNN architecture for the gait representation technique was 96.84% , 97.29% for GEI and GGMI respectively. According to the findings, improving gait recognition performance requires accurate gait representation methodologies and efficient learning approaches. Our CNN architecture shows a better accuracy for normal walk, carrying, and clothing irrespective of view angles. In the future, the proposed CNN architecture can be considered for other gait covariates issues for performance analysis with advanced hyper-parameters for better accuracy.

4.5 Feature Fusion-based gait recognition

Identification of a person based on their walking pattern is affected by factors such as camera viewing angle, clothing, carrying a bag, walking surface, and complex situations. These are the covariate conditions in human gait recognition. The covariate conditions change an individual's gait pattern, making it difficult to implement gait recognition in a realistic environment. This work addresses the issue of covariate conditions in gait identification through a fusion of features. Using a pre-trained VGG16 model with four fully connected layers, the dynamic features are extracted and merged with HoG (Histogram of Oriented Gradients) features extracted from the raw GEI (Gait Energy Image) gait templates. PCA (Principal Component Analysis) is then used to lower the dimension of the combined features in order to select the discriminant feature vectors. Using the CASIA-B dataset, which examines the effect of carrying and clothing factors under known and unknown conditions, the efficacy of the suggested technique is examined. The findings show that the proposed technique, which employs an MLP (Multi-layer Perceptron) classifier, outperforms other existing approaches in terms of accuracy while walking normally, wearing a coat, and carrying a bag under identical viewing conditions.

4.5.1 Related works

Covariate conditions in human gait have a significant effect on the effectiveness of Gait-based human identification in intelligent surveillance monitoring. Model-free gait recognition is one technique that is used mostly in gait recognition. because it is inexpensive and has a low computational complexity. In model-free, gait features are estimated from the whole image or contour image. GEI is a model-free approach to generating enhanced spatial-temporal features. The GEI is calculated by taking the average of a series of binary silhouette images collected over the duration of one gait cycle. The information in the GEI is likely to drastically change the person's appearance under different covariate conditions[106]. In the past, GEI model-free gait templates have been adopted in a number of empirical studies of gait recognition algorithms[99, 98, 84]. In our recent work on an effective method to reduce covariate issues in gait-based human identification, an efficient gait representation is suggested[38]. In this work, a good classification accuracy is achieved for multi-view angles for all the covariate conditions by using GEI and GGMI(Gait gradient magnitude image)[4] with a Nearest Centriod(NC) classifier. Many works on gait recognition have indicated that the result analysis was primarily performed on transforming models with known covariate conditions. The uniqueness of gait recognition is that it serves as an effective distance-based human identification system without subject cooperation. So, it is required to evaluate the gait recognition approach with a mixture of gait sequences in various covariate conditions. Bashir.et. al,[8] works on gait recognition without subject has shown remarkable gait recognition performance for covariate-invariant using Gait entropy Image (GEIn) from a side view walking posture of 90 degrees. Using a keyframe extraction method for local feature detection between the frames shows improvement in extracting distinctive features[35]for gait recognition. An effective method for reducing covariate conditions in gait recognition is a part-based gait that considers only the less affected part by covariate factors [72].

In recent times, the recognition of gait through the use of convolutional neural networks (CNN) has been the main focus of a significant number of cutting-edge human identification systems. Gait recognition based on CNN networks such as Deepgait[43], GEINET[78], GaitPart[25], GaitSet[14],etc, has shown remarkable results. According to the above works, the introduction of deep convolutional network-based feature learning methods has resulted in the use of pre-trained models for gait recognition. The studies also show that combining the pre-trained model with other static model feature descriptors could be an alternative method for overcoming covariate factor issues in gait recognition.

vspace-5mm

4.5.2 Proposed framework

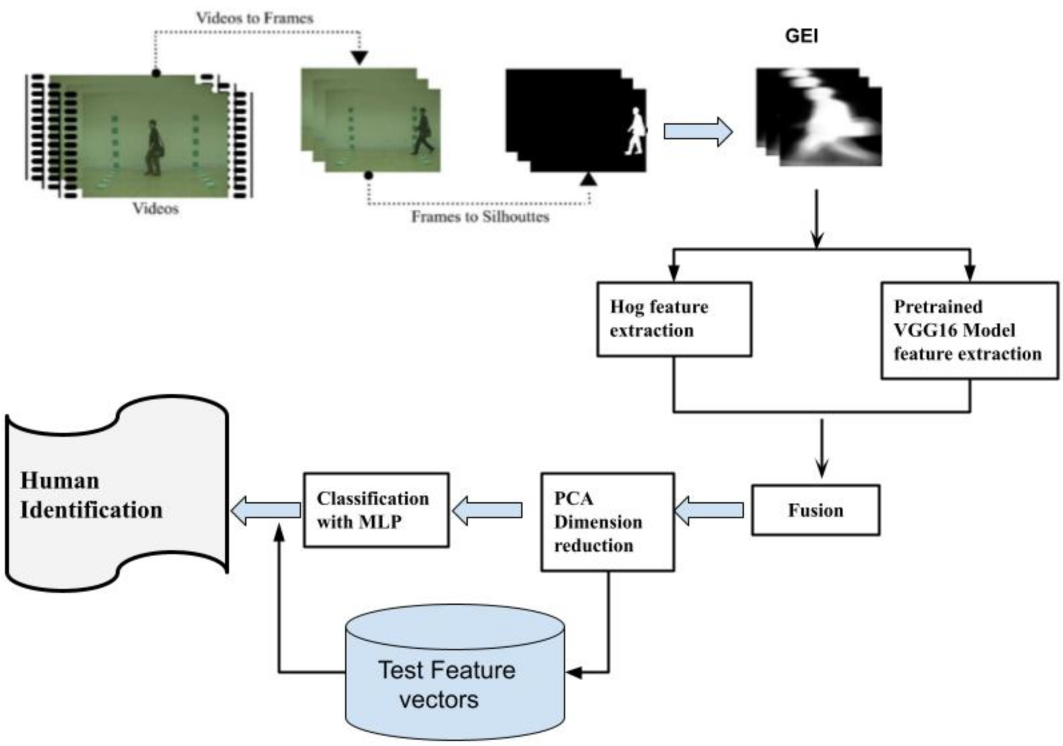


Figure 4.6: The Proposed framework

The proposed framework for reducing the covariate factors in gait-based human identification is described in detail in this section, the proposed framework is shown in Figure 4.6. The framework is designed to provide a robust and accurate identification system, even in the presence of covariate factors such as wearing different clothes and carrying different conditions. The first step in the process is to input video files, which undergo preprocessing to extract silhouette frames. A deep Convolutional Neural Network (CNN) image segmentation model is then used for background subtraction, to further extract the silhouette frames.

Once the silhouette frames have been extracted, a Gait Energy Image (GEI) gait template is created using the silhouette averaging technique suggested by Han and Bhanu et.al[31]. This GEI gait template is then subjected to feature extraction using a HoG(Histogram of Oriented Gradients) of a 16x16 bin size with nine orientations. The resulting feature vectors are 7056 in number.

A VGG16 pre-trained model is then used to extract 1024 feature vectors by unfreezing

(a) Normal

(b) Bag

(c) coat

Figure 4.7: Casia-B 90 Degree dataset Sample

the last 4 layers with 2048, 2048, 1024, and 1024 neurons, and using a Rectified Linear Unit (ReLU) activation function. The extracted features are combined to create fused feature vectors with 8080 features.

To further reduce the feature vectors, PCA(Principal Component Analysis) is applied as a dimension reduction method. The purpose of this is to retain only the discriminant features that characterize the classes. The resulting reduced feature vectors are then used for classification with an MLP (Multilayer Perceptron) classifier. This classifier has three dense layers of 2048, 1024, and 512 neurons, each with a ReLU activation function and an Adam optimizer. The goal of this classifier is to correctly classify the subject in various covariate conditions.

4.6 Experimental and Results analysis

The CASIA-B a standard dataset, which consists of 124 individuals from 11 different view angles with a variation of 18 degrees each, is used for the experiment. The dataset includes ten different perspectives, with six of them being normal walking, two being walking while carrying a bag, and two being wearing a long or short coat. A sample of each walking normally, carrying a bag, and wearing a coat at a 90-degree angle is shown in Figure 4.7.

To evaluate the impact of various covariate conditions on gait-based human identification, the dataset is divided into three parts. For normal walking conditions, 6 normal walking sequences were selected and 4 of them were kept in the gallery set. The gallery set was used to estimate the correct classification for the remaining 2 sequences of normal walking conditions, as well as for the baggage covariate conditions and clothing covariate conditions.

The Multi-layer Perceptron (MLP) classifier was used for classification and the results

Table 4.5: Experimental analysis result indicating normal walk, wearing coat and carrying bag condition when gallery and probe angle are identical

Gallery / Probe Angle	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	MEAN
Normal Vs Normal	100	100	97.97	96.61	99.32	**97.29**	98.64	96.62	98.64	100	100	98.64
Normal Vs Coat	63.01	66.89	64,14	69.38	73.64	**74.32**	68.91	63.26	67.34	59.31	67.56	67.36
Normal Vs Bag	97.26	90.54	87.16	84.35	90.54	**94.59**	95.27	90.47	90.47	91.09	95.91	91.60

Table 4.6: Comparison table with existing methods

Author	**Gait representation**	**Nm-Nm**	**Nm-Bg**	**Nm-Cl**	**Mean**
K. Bashir et.al [8]	GEnI	100.0	78.30	44.40	74.20
Margaret et al [38]	GGMI+Hog	99.18	83.06	37.77	73.33
M. Rokanujjaman et.al[72]	EnDFT	97.61	83.87	51.61	77.69
G. Huang et.al[35]	LFF	91.90	80.30	72.30	81.50
MuhammadAsi etal[6]	HOG-GEI	82.30	79.80	87.90	83.33
Our method	**HOG+VGG16**	**97.29**	**94.59**	**74.32**	**88.73**

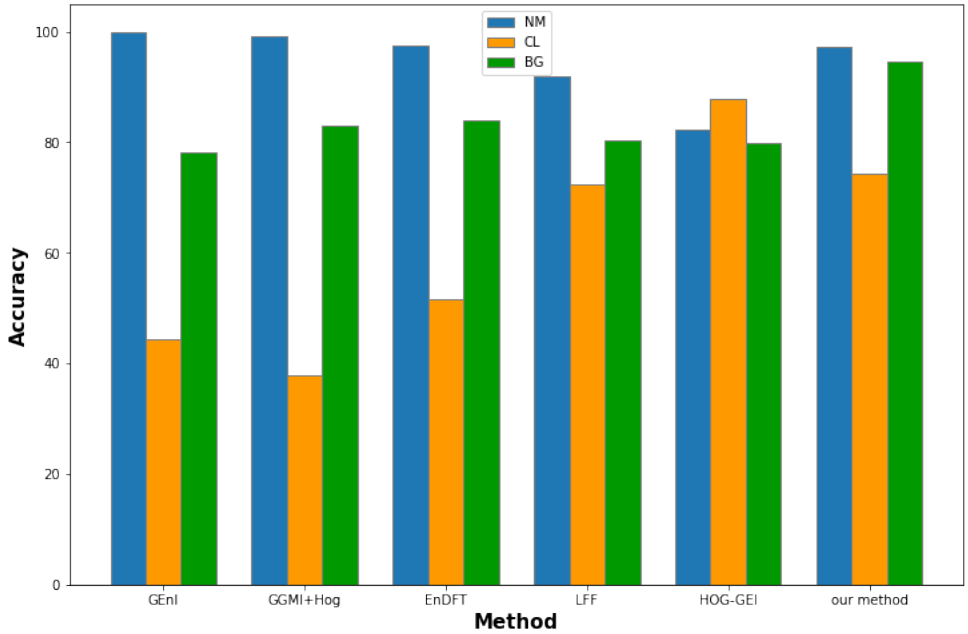


Figure 4.8: Comparison graph with other methods

obtained from the experiment are shown in Table 4.5for different viewing angles. The analysis reveals that the proposed method demonstrates significant improvement for normal walking and carrying conditions when the gallery and probe angles are identical. However, the clothing covariate condition appears to be a critical issue even with identical viewing angles. Table 4.6 compares the results with existing works.

4.7 Summary

This chapter highlights Convolutional Neural Networks (CNNs) as an approach for gait recognition, describing their structure and functionality. CNNs can be trained on a dataset of gait sequences to classify individuals based on their walking patterns. To achieve robust recognition, the study proposes a new method using a CNN model and feature fusion technique to account for covariate conditions that can affect a person's gait appearance. The evaluation of the proposed method was performed in a realistic scenario where the gallery and probe sets had varying covariate conditions. The proposed methods are evaluated using the CASIA-B dataset and compared with existing approaches. The results showed a better result when both the gallery and probe had identical viewing angles for normal walking and carrying conditions. The proposed method achieved higher classification accuracy compared to current state-of-the-art approaches, with normal walking accuracy at 97.29%, carrying bag accuracy at 94.59%, and wearing coat accuracy at 74.32% at a 90-degree viewing angle. The results indicate that clothing covariate conditions play a significant role in gait recognition, making it a challenge to implement in real-world scenarios.

Overall, CNN-based gait recognition holds promise for diverse applications, including security systems, healthcare, and sports analysis. Further to improve the robustness of gait recognition and make it suitable for use in automatic human identification systems, future research can focus on developing more robust deep models.

The impact of gait recognition under diverse viewing conditions is also one of the challenges of this research work, given the ability to observe human walking from various angles. Analyzing gait features from multiple viewpoints poses significant challenges, discussed in the upcoming chapter. The next chapter aims to devise effective techniques to address these challenges in multiview gait recognition, striving for a robust system that can recognize gait irrespective of the viewing angle.