

Article

Age Estimation in Human Gait Extraction Using A Combination of Multi-Energy Image with Invariant Moment

Nasser Al Musalhi *, Erbug Celebi

Department of Computer Engineering, Cyprus International University,
Nicosia, North Cyprus, Mersin 99258

* Correspondence: nhsoman2015@gmail.com

Abstract: Accurately estimating a person's age is crucial for identity verification at all critical checkpoints, including airports, land borders, and seaports. Human gait may be used as a biometric identifier and indicator of age, among other things. This study aims to develop methods for estimating a person's age by observing their walk. In this paper, a novel technique for preprocessing the proposed gait dataset has been used by utilizing a combination of the gait energy image (GEI), the accumulated frame difference energy image (AFDEI), and the invariant moment of the image. The proposed technique outperformed state-of-the-art methods in terms of accuracy. The proposed method was tested and evaluated using a convolutional neural network (CNN), and it achieved an average accuracy of 90.35% across 14 different view angles within 5 K-Fold, the proposed method resulted in 94.68% and 94.54% in 30° and 75° view degree respectively. concluding that the approach is effective and promising for estimating age using human gait.

Keywords: Age estimation; AFDEI; CNN; human gait; Neural networks; GEI; Invariant moments

1. Introduction

In today's society, the development of technology and computer use has transformed security into one of most challenges. As human beings, we always look for ways to overcome security. Passwords, personal identification numbers (PINs), and a variety of additional safeguards were the norm in the past when it came to protecting sensitive information. Due to the growing number of threats, we are always brainstorming new ways to positively identify individuals and confirm their authenticity. Given that it is feasible for each password and PIN to be shared, a system was developed that relies on the unique traits of the person who is granted access to the system. The word "biometrics" was coined to describe this academic field. Human gait is only one example of the various biometric identifications that rely on the automated use of a person's physiological or behavioral traits to establish or validate their identity [1].

For the last decade, a significant amount of investigation has been put into gait recognition, which is essentially an attempt to categorize people based on the way they walk. When compared to other forms of biometric identification like fingerprints and irises, gait has the distinct benefit of being non-intrusive while yet being simple to obtain from a distance makes it a particularly desirable kind of identification. These benefits are significant for video-based such as intelligent security surveillance systems in which gait detection is the primary approach. The gait recognition system is an innovative piece of technology that does not necessarily need to be in a certain position or in standing in front of biometric devices. Instead, it relies on stealthy technology that analyzes human gait gestures to determine whether an individual poses a threat to security or is acting unusually [1], [2]. Both model-free and model-based approaches can be utilized in the process of gait identification; nevertheless, the two styles of gait analysis are the most prevalent at present. Methods that are model-based and use gait recognition model human movements by employing physiological features. The

dynamics of gait, as well as human kinematics, may be explained with the use of explicit characteristics such as the stick-figure model. Methods of gait recognition known as “model-free gait analysis” process gait sequences without using any models. In most cases, the motion features of silhouettes as well as their spatio-temporal forms are studied. A gait cycle detection technique, a collection of training data, a feature extractor, and a classifier are the four components that make up a standard model-free gait identification approach [3].

Human gait age estimation is an exciting new area of study that attempts to determine a person's age only from observations of their gait. Several elements, including physiological, biomechanical, and neuromuscular, contribute to the inevitable alterations in gait that occur with advancing years. Researchers can create methods to precisely and non-invasively determine a person's age by examining and measuring these age-related changes in gait. The ability to accurately determine a person's age based on their stride has several uses in fields like criminology, medicine, biometrics, and geriatrics. Collecting gait data from people of varying ages is essential to the field of gait age estimate.

Gait age estimation has the potential to have far-reaching positive effects, such as the enhancement of security systems, providing more targeted healthcare interventions, and shedding light on the effects of aging on human mobility.

1.1. The Gait Cycle

Walking is the most natural and efficient way of movement for human beings to use when they want to go independently and quickly. The term “gait” refers to a movement that may be described as a movement that consists of a translation of the complete body and is made possible by a repetition of movements of body segments while maintaining equilibrium. The concept of a cycle may be derived from this cyclical pattern of motion. The term “gait cycle” refers to the amount of time that elapses between any two similar events that occur during the gait process. These two technically distinct events often begin and end with the first and second contact of one foot with the ground (called initial contact or IC) [4]. The lower extremity went through alternating stages of stance (when the foot was in contact with the ground) and swing (when the foot was in the air) as part of the gait cycle (when the foot was not in contact with the ground) [5]. Figure 1 shows the periods of the gait cycle or as it's known the gait phases.

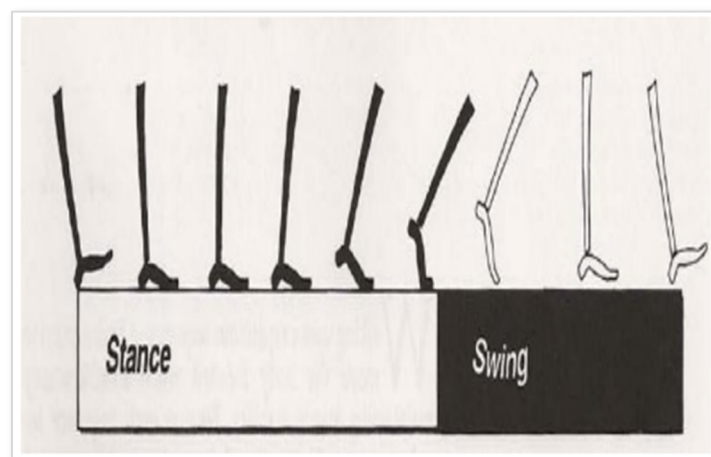


Figure 1: The two Periods of the Gait Cycle [4].

The gait cycle starts when one foot strikes the ground, and it concludes when the same foot, after taking one stride, is placed back on the ground. On average, 60% of the cycle occurs during the stance phase (the heel-to-toe contact process of the foot), whereas 40% occurs during the swing phase (the foot is suspended and progresses through the air) [6]. Figure 2 shows the events of the gait cycle:

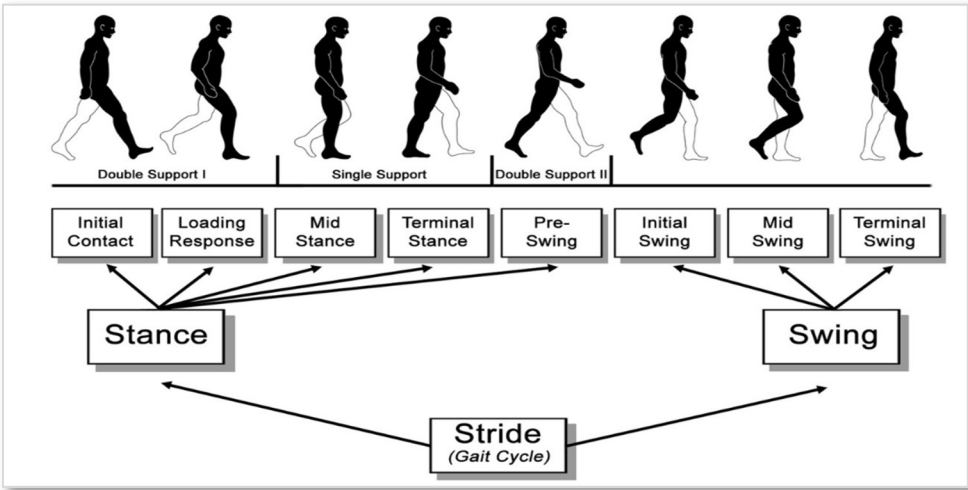


Figure 2. The Events of the Gait Cycle [6].

1.2. Gait Recognition Processes

It is the recognition algorithm that makes or breaks a human gait recognition system. Multiple capturing devices, such as a video camera, sensors, wearables, and so on, might provide data for a gait recognition system. The efficiency with which the system can make use of this raw data depends on the accuracy of the recognition algorithm. Overall, a gait recognition system [7], as shown in Figure 1, will carry out the following processes [8,9]:

- Capturing gait data: The video sequence of the gait is gathered by detection and tracking, with data collected by cameras at varying distances and angles.
- Analyzing pictures: Preprocessing and analysis are used to retrieve the gait features. Important gait identification operations including locating and segmenting gait motion in an image sequence, extracting features from that motion, etc., are all part of this.
- Feature extraction: Convert the gait data into the same format as the gait information already in the database.
- Comparing data: Review the database of gait characteristics and see whether any of the newly obtained features match up.

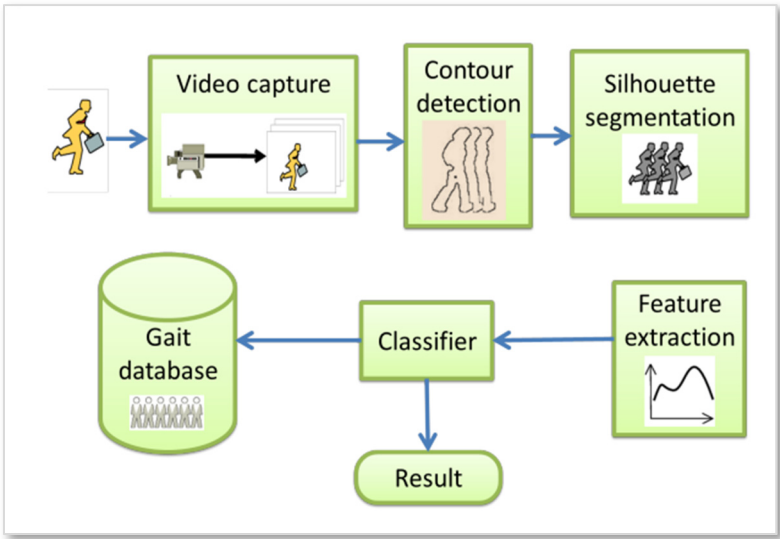


Figure 3. Gait Recognition Phases [7].

2. Related works

Recently human gait recognition has been highly researched in various applications including age estimation problems. In studies, such as [10–15] different well-known individual and combined machine learning and deep learning models are employed to extract the ages from the human gait, for example, the convolutional neural network (CNN), the support vector machine (SVM), the support vector regression (SVR), etc. They used various evaluation metrics to assess the performance, including mean absolute error (MAE), correct classification rate (CCR), etc. Also, various gait recognition datasets are adopted for evaluation purposes, such as CASIA-B, OU-MVLP gait datasets, etc. However, each model has its benefits and drawbacks related to accuracy and computational complexity.

In [16], the authors of this paper propose a gait recognition method based on a combination of GEI and AFDEI in which time features can be reflected. According to their research, the gait features were chosen by combining the invariant moments extracted from GEI and AFDEI. Subsequently, the nearest neighbor classifier based on the Euclidean distance was employed to complete gait recognition. Furthermore, their approach is applied to the CASIA-B gait dataset. The experimental results showed that their method outperformed other previous studies related to the same field.

Domagoj et al. [17] developed a new method in their study that makes use of self-supervised learning (SSL) to complete the gait identification test they use the vision transformer (ViT) architecture presented in the self-supervised DINO approach for image classification. To learn the useful gait features, training samples that have not been annotated are used. Following this, gait features are extracted and fed into a feature extractor [18], from which a fully connected neural network (FCNN) classifier is trained using a supervised method. The CASIA-B and OU-MVLP gait datasets are used to implement their technique, and the walking scenarios include “normal walking,” “walking with a bag,” and “walking with a coat.” Both datasets provided encouraging findings for their experimental approaches. Moreover, in another study [19], a novel concept is presented in which an innovative network named GaitSet is presented that can learn and gather identity information from the set. Their approach, which relies on a set perspective, is immune to frame permutation and can simply combine images from movies shot under varying situations (such as different camera angles or with the subjects wearing or not wearing specific clothing). For experimental results, both gait datasets, CASIA-B and OU-MVLP, have been employed. Their method achieved good accuracy results in various situations (both normal and complex) compared to state-of-the-art techniques.

A self-supervised gait identification system named SelfGait was proposed in [20] to improve spatiotemporal backbone representation skills, it employs the massive quantity of unlabeled gait data as a pre-training phase by using horizontal pyramid mapping (HPM) and micro-motion template building (MTB) as our spatiotemporal backbones to capture multi-scale spatial and temporal representations. SelfGait is demonstrated to outperform existing state-of-the-art gait recognition approaches in experiments conducted on the CASIA-B and OU-MVLP benchmark gait datasets.

Many popular deep learning techniques today make use of loss functions, including contrastive loss and triplet loss, which are commonly employed in the facial recognition challenge. According to [21], these loss functions suffer from the issue of “hard negative mining.” Therefore, in their article, they propose using a loss function called angle center loss (ACL) to learn discriminative gait features. This loss function is efficient, reliable, and related to gait. If the temporal window or the local sections are altered, the suggested loss function will remain unchanged. The proposed loss function is different from center loss in that it learns many sub-centers, one for each identity angle. The use of a simpler spatial transformer network is proposed for locating the correct horizontal body portions. Each horizontal segment of gait is analyzed separately, and then that data is combined to generate the overall description. For example, they recommend using long short-term memory (LSTM) units as the temporal attention model to learn the attention score for each frame, such as prioritizing discriminative frames and neglecting low-quality ones. When compared to temporal average pooling and gait energy pictures, the temporal attention model performs exceptionally well (GEI). When

applied to the CASIA-B and OU-MVLP gait datasets, the combination of the three achieves state-of-the-art performance on a variety of cross-view gait identification criteria.

In [22], a multi-view gait generative adversarial network (MvGGAN) is used to artificially produce gait samples and supplement existing gait datasets, making it possible for deep learning-based cross-view gait identification algorithms to work. The recommended MvGGAN technique only requires the training of a single generator, regardless of the number of datasets that are concatenated. To mitigate the negative effects of distribution divergence resulting from sample generation, domain alignment based on the predicted maximum mean discrepancy is also utilized. Experiments on the CASIA-B and OUMVLP datasets show that the proposed MvGGAN approach for generating false gait samples greatly enhances the efficiency of existing state-of-the-art cross-view gait recognition systems in both single-dataset and cross-dataset assessment scenarios. An approach to gait identification using Graph Convolutional Networks (GCNs) that combines higher-order inputs with residual networks is presented by the authors in [23]. Extensive investigation of the two most prominent gait datasets, CASIA-B and OUMVLP-Pose, demonstrates a considerable improvement (3X) of the state-of-the-art (SotA) on the biggest gait dataset, OUMVLP-Pose, and robust temporal modeling abilities.

3. Proposed Gait Methodology

In this paper, we propose to use a combination of GEI and Accumulative frame difference energy image AFDEI using a weighted average calculation to generate a combined energy image CEI which will feed to the neural network. A block diagram of the proposed method, representing the steps from reading the dataset to feeding it to the neural network and classifying the image, is shown in Figure 4 below.

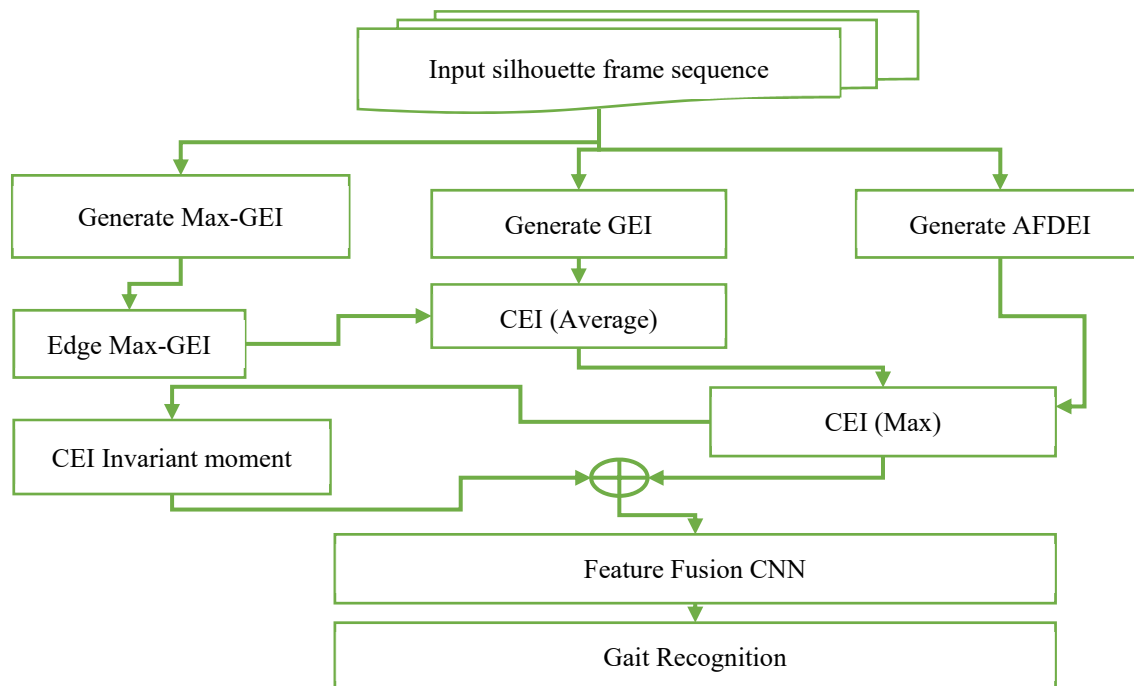


Figure 4. Proposed Methodology.

3.1. Dataset

The OU-ISIR Gait Database, Large Population Dataset with Age [24], is the proposed dataset for this paper. This dataset is a preprocessed video-based human gait. The information was gathered in conjunction with a video-based gait analysis exhibit at a museum dedicated to science and technology. Everyone who participated in this study gave their signed, official, informed permission. There are 10,307 individuals in total in the dataset, with 5,114 men and 5,193 females with ages

ranging from 2 to 87 years old, as seen from 14 distinct perspectives (0°-90°, 180°-270°). Seven network cameras (Cam1–7) are arranged in a quarter circle with the walking track at its center, each at an azimuth angle of 15 degrees with a diameter of 8 meters and a height of 5 meters. Each camera captures 25 frames per second of 1,280 by 980-pixel gait data.

3.2. Preprocessing Operation

The detection of moving objects, the extraction of features, and the identification of gait are the three stages typically involved in the process of normal gait recognition. Dataset [dataset] recommended the sequence of silhouette images to be used to generate the gait energy images. Therefore, this paper used the same recommended image list.

3.2.1. Preparing OUMVLP with age image dataset

The OUMVLP with age image dataset has been divided into 18 classes (0 -17) in 5 years slices to meet our aim as shown in Figures 5 and 6 shows the total number of images in each class after the data is categorized.

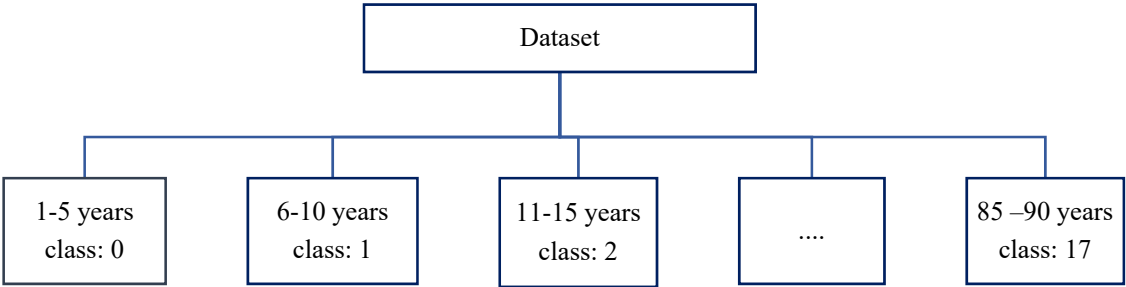


Figure 5. Dataset images classes.

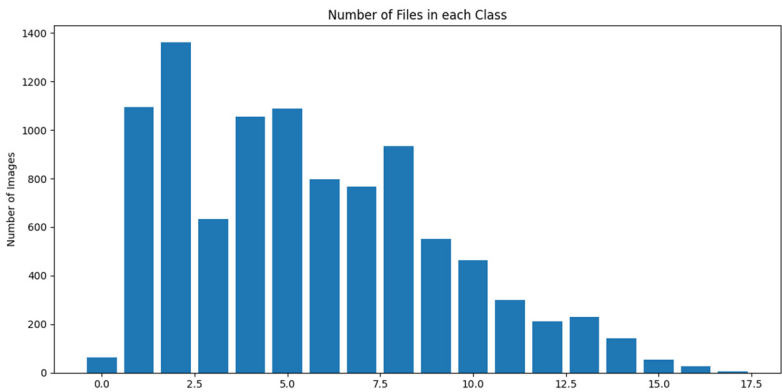


Figure 6. number of images in each class.

3.2.2. Gait Energy Image

One of the most commonly used representations is known as the gait energy image (GEI), and the reason for this is that it strikes a good balance between the amount of computing effort required and the amount of recognition it provides. A spatio-temporal description of the gait, known as GEI, may be produced by averaging the silhouettes throughout a gait cycle. GEI can then be used to analyze gait patterns. Although this visualization is one of the most often used gait representations, it has been discovered that clothes and carrying circumstances have an impact on detection accuracy [25].

The weighted average approach is used to create the gait energy picture, which is then used to represent the gait sequence of a cycle in a straightforward energy image. Processing is done on the

gait sequences that are included inside a gait cycle to align the binary silhouette. The following equation may be used to get the gait energy picture when the gait cycle image sequence is given as $B(x, y, t)$ Equation (1):

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

where $B(x, y, t)$ denotes the gait cycle pictures, N denotes the number of frames in a cycle's gait series, and t denotes the current frame of the gait cycle. Figure 7a shows the silhouette gait cycle frame sequences and Figure 7b represents GEI extracted from the gait frame sequence.

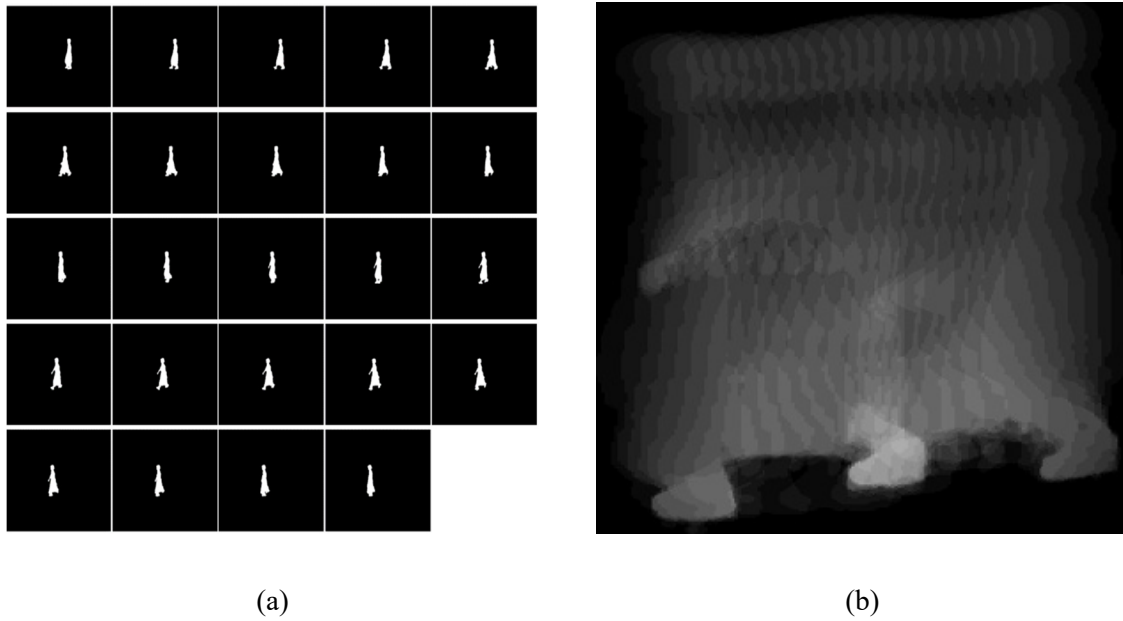


Figure 7. (a) Silhouette gait cycle frame sequences, (b) GEI image extracted from gait cycle.

3.2.3. Accumulated Frame Difference Energy Image

To create the AFDEI, preprocessed silhouette sequences are used. The proposed FDEI approach aims to minimize the effect of the silhouette's incompleteness while preserving the majority of the shape's attributes and reducing temporal volatility. Recovering shape information from an incomplete frame requires completing additional frames to compensate for the incomplete one [26]. When the forward frame difference image and the reverse frame difference image are combined, the energy picture is produced as the frame difference. image computation for the forward direction is shown by Equation (2), computation for the backward direction is shown by Equation (3), and computation for the frame difference is shown by Equation (4):

$$F_a(x, y, t) = \begin{cases} 0, & \text{if } B(x, y, t) \leq B(x, y, t-1) \\ B(x, y, t) - B(x, y, t-1), & \text{otherwise,} \end{cases} \quad (2)$$

where $F_a(x, y, t)$ is the forward frame difference image.

$$F_b(x, y, t) = \begin{cases} 0, & \text{if } B(x, y, t) \geq B(x, y, t-1) \\ B(x, y, t-1) - B(x, y, t), & \text{otherwise,} \end{cases} \quad (3)$$

where $F_b(x, y, t)$ is the backward frame difference image.

$$F(x, y, t) = F_a(x, y, t) + F_b(x, y, t) \quad (4)$$

where $F(x, y, t)$ is the frame difference image.

The cumulative frame differential energy picture may be created using the weighted average approach [26]. This image can depict the temporal characteristic. The cumulative frame difference picture may be calculated as shown in Equation (5), which is as follows:

$$A(x, y) = \frac{1}{N} \sum_{t=1}^N F(x, y, t) \quad (5)$$

where $F(x, y, t)$ is the frame difference image. N is the total frames difference generated by Equation (4) and t is the current frame difference.

The information frames are shown in following Figure 7a, the frame difference energy picture is displayed in the following Figure 8a, and the cumulative frame difference energy image is displayed in Figure 8b below.

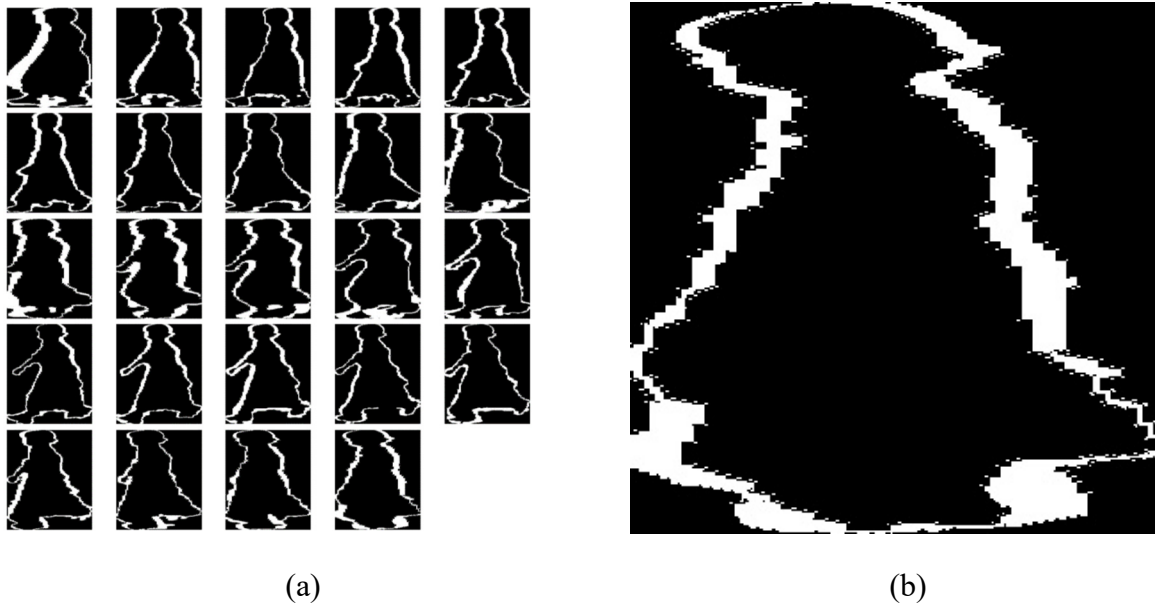


Figure 8. (a) Accumulative frame difference Sequence, (b) AFDEI.

3.2.4. Gait Energy Image using Max Pixel (GEI-Max)

GEI is the average of the total number of gait cycle sequence images. However, in the proposed method we fetch the highest value among the whole cycle of gait frames. Equation (6) represents the calculation of maximal pixel values in position. Figure 10a represents the result of GEI-Max generated from Silhouette frame sequences shown in Figure 7a.

$$A(x, y) = \max(F_1(x, y), F_2(x, y), F_3(x, y), \dots, F_n(x, y)) \quad (6)$$

where $F_n(x, y)$ represents the frame number in the silhouette gait sequence.

3.2.5. Edge Detection:

Edge detection is a method that identifies the boundaries between two homogenous picture areas with varying brightness levels [27]. Detecting edges and corners in a single picture may help reduce the size of the data stream and also aid in well-matched operations such as image restoration and so on [28]. There are many approaches to edge detection, like Prewitt, Canny, Laplacian, and Sobel edge detection. In this study, we used Sobel edge detection to preprocess our dataset. The Sobel technique of edge detection for picture segmentation discovers edges using the Sobel approximation to the derivative. When the gradient is steepest, it appears ahead of the edges. The conjecture suggests that at least two 3x3 complexity kernels (Figure 9) are used by the operator [29]. Sobel edge detection is an easy and computationally efficient method that offers reliable localization, bidirectional differentiation, noise suppression, and adaptability. It works with photographs of varying sizes and

formats, and it’s capable of processing both smooth and jerky changes in brightness, the Sobel edge detection method was employed to detect the edge of GEI-Max, and the results are shown in Figure 10 below.

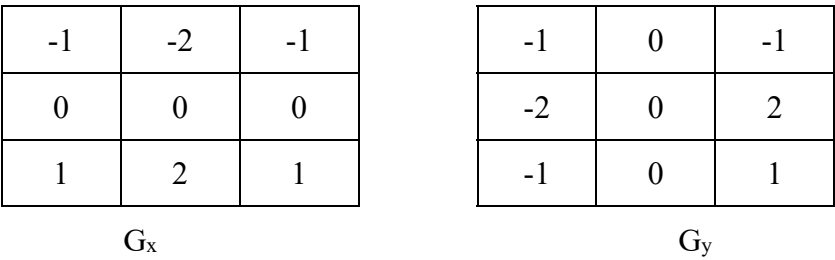


Figure 9. Sobel X & Y Kernels.

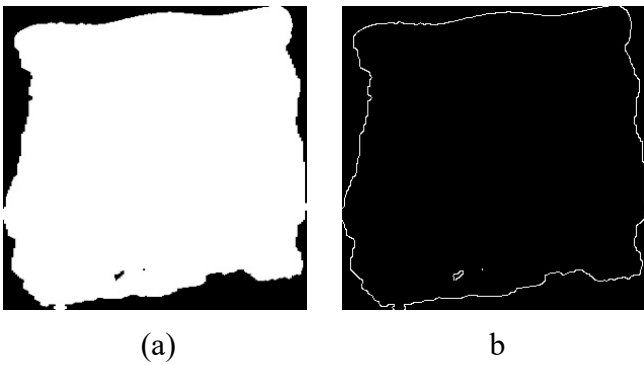


Figure 10. (a) GEI-Max, (b) Edge of GEI-Max.

3.2.6. Combined Energy Image (CEI)

Here we first merge Edge GEI-Max with GEI using the average calculation presented in Equation (1) result can be seen in Figure 11b.

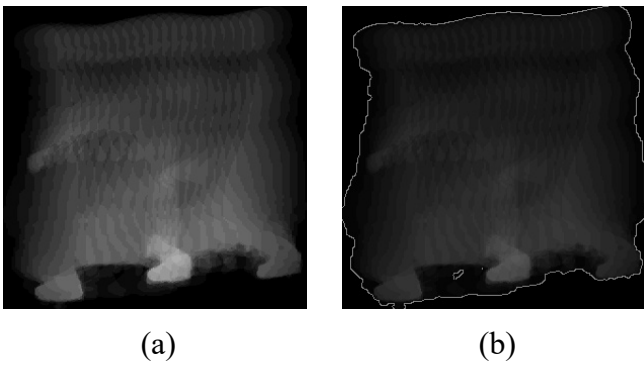


Figure 11. (a) GEI image, (b) Combined CEI-Average image.

Then we merge the resulting image from Figure 11b with the AFDEI image illustrated in Figure 7b using Equation (6) to combine the two images as shown in Figure 12.

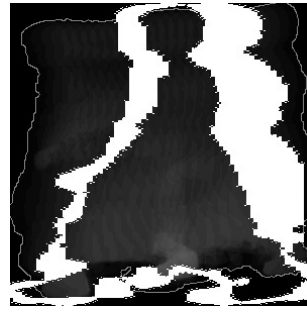


Figure 12. Combined CEI-Max.

3.2.7. Invariant Moment of Image

As a method of feature extraction, moment invariants are put to use in the fields of shape recognition and identification. With its ability to generate feature vectors that accurately reflect a picture, it has found widespread use in a variety of contexts, particularly recognition. The form attributes of binary images can be extracted using the moment-invariant method. Moments that have been computed from binary images are referred to here as “silhouette moments” [30]. For any conceivable moment, type p and q , intensity image functions $f(x,y)$ with $N \times M$ pixel dimensions can be computed, as can their characterization in the context of universal computation, as in Equation (7).

$$m_{pq} = NF \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \Pi_{pq}(x,y) f(x,y) \quad (7)$$

where NF is the normalization factor and Π_{pq} is the moments, kernel formed by the product of the specified polynomials of order p and q that serve as the orthogonal basis. Given the variety of moment types that emerge from the various types of polynomials in the Kernel, we give this new family of moments a descriptive name.

3.2.8. Hu invariants moment

Hu [31] introduced the first ever 2-D moments in 1962. As a logical component of the “Moment Invariants,” he proposed the 2-dimensional geometric moments of an image’s distribution function. Hu Moments invariants are a set of seven numbers determined by calculating central moments that remain unchanged upon changing the underlying image. Invariance under translation, scaling, rotation, and reflection has been demonstrated for the first six moments. But the sign for image reflection flips at the seventh moment. The calculation of Hu invariants’ moments is represented by Equation (8).

$$\begin{aligned} M_0 &= m_{20} + m_{02}, \\ M_1 &= (m_{20} - m_{02})^2 + 4m_{11}^2, \\ M_2 &= (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2, \\ M_3 &= (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2, \\ M_4 &= (m_{30} - 3m_{12})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) \\ &\quad + (3m_{21} - m_{03})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2), \\ M_5 &= (m_{20} - m_{02})((m_{30} + m_{12})^2 - (m_{21} + m_{03})^2) + 4m_{11}(m_{30} + m_{12})(m_{21} + m_{03}), \\ M_6 &= (3m_{21} - m_{03})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) \\ &\quad - (m_{30} - 3m_{12})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2) \end{aligned} \quad (8)$$

Hu’s approach is providing seven vector values (M_i) for a given image as the calculated results for random five images in the proposed dataset shown in the Table 1 below.

Table 1. Invariant moment results.

CEI Image	M ₀	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆
1	3.176056	7.516807	12.32739	13.01199	-25.7417	-17.2446	25.99022
2	3.153797	7.597082	10.45985	12.06968	23.6814	-16.463	23.38354
3	3.178445	8.46446	10.98882	12.87856	-24.9104	-17.1599	25.03184
4	3.164626	7.574627	10.68473	13.04071	-25.1183	-17.1062	-25.0044
5	3.174391	7.810398	11.3712	14.33523	28.11615	-18.3237	27.19149

3.2.9. Proposed Combined Technique using CEI with Invariant moment

The result of Hu invariant moments which is mentioned in Table 1 is combined with CEI in Figure 12 to produce the final preprocessed image which then feeds it to the neural network. To implement such a combination, we employed a novel technique with the following equations, (Equation (9) to Equation (11)) are used to implement such combination:

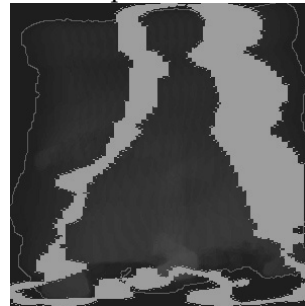
$$A = \left\lfloor \frac{(Ci + 1) \sum_{m=1}^7 Hm}{7} \right\rfloor, \quad (9)$$

where A represents the combined class number with Hu invariant moments, $\sum_{m=1}^7 Hm$ is the sum of seven invariant moments, Ci represents the class index of the given image. An activation formula is applied as seen in Equation (10), and finally, generate the new pixel by implementing Equation (11).

$$A' = \begin{cases} 255, & \text{for } A > 255 \\ A, & \text{otherwise} \end{cases} \quad (10)$$

$$pi(x, y) = \frac{pi(x, y) + A'}{2} \quad (11)$$

where A' represents the activation of A , $pi(x, y)$ is the current pixel in the image. Figure 13 shows the result of applying the above-mentioned equations.

**Figure 13.** Final CEI.

4. Neural Networks

Artificial neural networks (ANNs) are a key part of deep learning techniques. One type of ANN called a recurrent neural network (RNN), is designed to process data in the form of a sequence or time series. useful for tasks including multilingual translation, speech recognition, and automatic picture captioning [32]. Convolutional Neural Networks (CNNs) are an artificial neural network subfield of deep learning that is frequently employed in the aforementioned domains of image and object recognition and classification. Therefore, deep learning uses CNN to identify items in a picture. CNN, like other neural networks, can efficiently mine time series and visual data for insights. As a result, it excels at image-related jobs like recognition and classification [33]. CNNs are trained to recognize patterns in images by employing linear algebraic techniques such as matrix multiplication. CNNs can also be used to categorize data involving sound and signals. The three primary layers that make up a CNN are a convolutional layer, a pooling layer, and a fully connected layer.

4.1. Class imbalance

In machine learning, one of the most common difficulties is a class imbalance, which occurs when some classes contain far fewer samples than others. There are a few ways to address class imbalance issues such as Class weighting, undersampling, oversampling, hybrid methods, and resampling data. Data undersampling entails reducing the sample size of the majority group so that it is comparable to the minority group, while data oversampling involves raising the sample size of the minority group [34], [35]. Both Class Weighting and Hybrid Methods provide more weight to the minority class in training by giving more samples from that class. As it appears in Figure 6 the classes are not equal and fatal differences between classes to overcome this issue we have used the class weighting approach to train the model, as well as the resampling method means either removing some classes images or increasing the minority by flipping, zooming or rotating the images which are not fit our aims. Figure 14 represents the class weights for 75 view degrees.

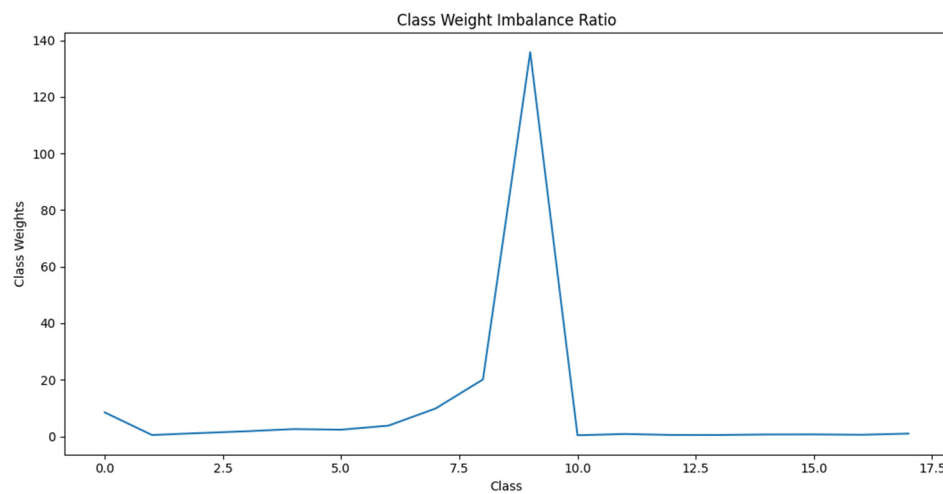


Figure 14. Class weighted imbalance ratio on 75 degree.

4.2. Convolutional Neural Network Model

In this research, we implement the CNN model to extract features from the OUMVLP image dataset after preprocessing. The model input layer with an image size of 100x100 and four hidden layers with several feature-extracting filters of 16, 32, 64, and 256, respectively. The activation function used is "LeakyRelu". Furthermore, the pooling layer is set to default as (2x2). The dense layer is set to 512, and finally, the classification layer uses the SoftMax activation function. The full architecture diagram is shown in Figure 15 below.

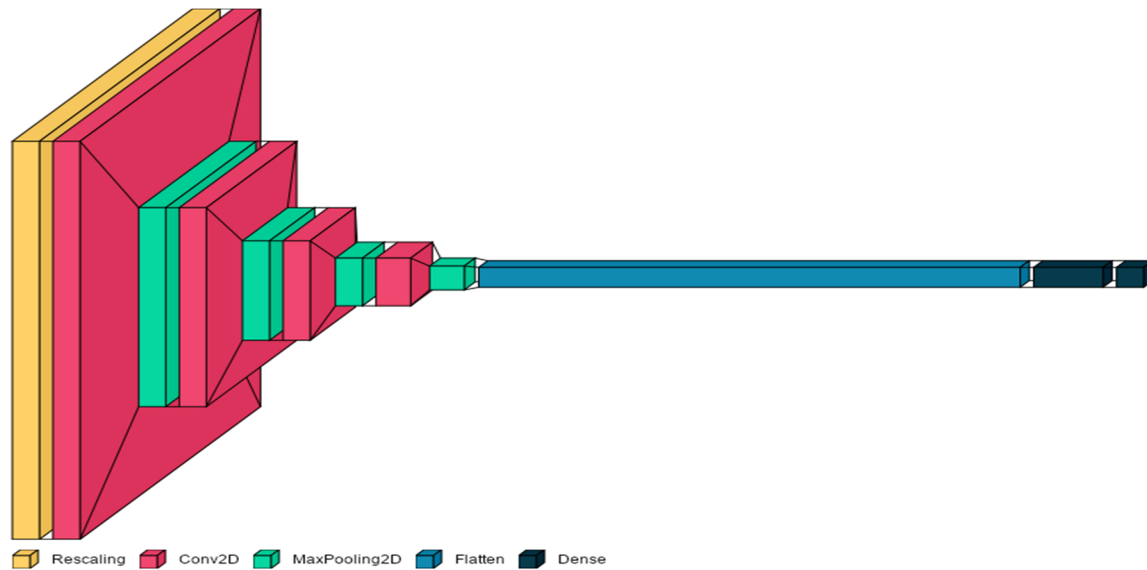


Figure 15. CNN Model Architecture.

4.3. Cross-Validation testing

Cross-validation is a method for evaluating the efficacy of machine learning models that involves training different models on separate subsets of the input data and comparing their results on the full dataset. Cross-validation is useful for identifying cases of overfitting, which is the inability to generalize an observed pattern. [36]. Validation refers to the procedure of determining whether or not numerical results measuring supposed relationships between variables are suitable as representations of the data. The evaluation of residuals, also known as an error estimate, is typically performed following model training. Here, the training error, which is the difference between the actual responses and the ones that were anticipated, is estimated numerically[37]. There are many techniques used for cross-validation, like holdout, k-fold, and leave-one-out. [38] In this paper, we choose to use K-fold cross-validation. In K-fold the data is divided into k subsets and tested independently using the k-Fold cross-validation method. Then, k rounds are carried out of the holdout technique, with a distinct subset of the data being used for testing and validation and the remaining data being used as a training set. Taking the average of the error estimates over all k trials allows us to determine our model's overall performance. It is demonstrated that each data point appears precisely k times in the training set and k-minus once in the validation set. As the same data is utilized for both the fitting and the validation, bias, and variance are drastically cut down. Alternating between training and test sets can further improve the method's effectiveness. As a general rule of thumb supported by observational data, $K = 5$ or 10 is recommended; however, this is not set in stone and can be modified to any value. [37,39].

5. Experimental Results

Many experiments are done using the proposed CNN model. The model was trained with 9777, divided into 18 classes. The dataset was split randomly into training and testing for 80 percent and 20 percent, respectively. The model is implemented over 20 epochs to train and test in 10 K-fold iterations. Furthermore, the model was trained and tested using the generated images after they had been preprocessed as described in Section 3.2.6. We observed that the experimental results show non-competitive accuracy performance. Meanwhile, an enhancement of preprocessing was done on the image dataset by combining the images from CEI-Max with invariant moments. The experimental results showed that the preprocessing technique had successfully enhanced the accuracy of the proposed model. The model also trains in all degrees available in the OUMVLP dataset separately. The experimental results (worst and best) for each view degree are recorded in Table 2.

Table 2. Accuracy score of 5 K-Fold for all view degrees.

View Degree	Worst Score	Best Score	Average Score
0°	74.49	98.19	91.26
15°	72.21	98.34	90.40
30°	91.14	98.82	94.68
45°	79.74	96.15	90.03
60°	13.75	98.00	76.64
75°	90.49	98.87	94.54
90°	83.02	99.23	93.27
180°	88.84	97.35	93.31
195°	77.17	97.52	89.59
210°	81.17	97.4	87.99
225°	64.93	96.50	89.24
240°	81.29	97.22	92.66
255°	73.96	98.32	88.97
270°	87.79	97.00	92.30

Also, Figure 15 shows the best K-fold accuracy progress during the training and testing operations for 50 epochs in 75 degrees and Figure 17 shows the cross-validation accuracy performance for all other tests in 75 degrees. Figure 18 shows the accuracy Evaluation on 5 k-folds for 75 degrees.

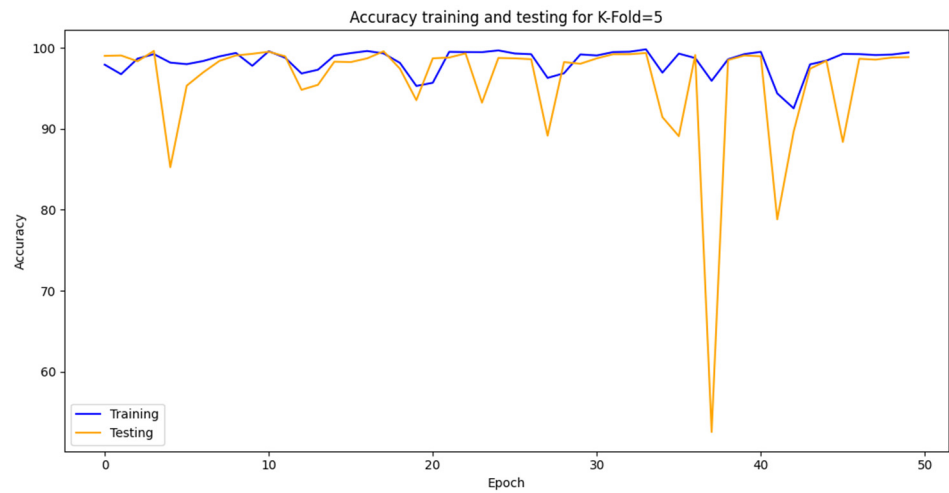


Figure 16. Best crossover Accuracy performance of 5 k-Folds on 75°.

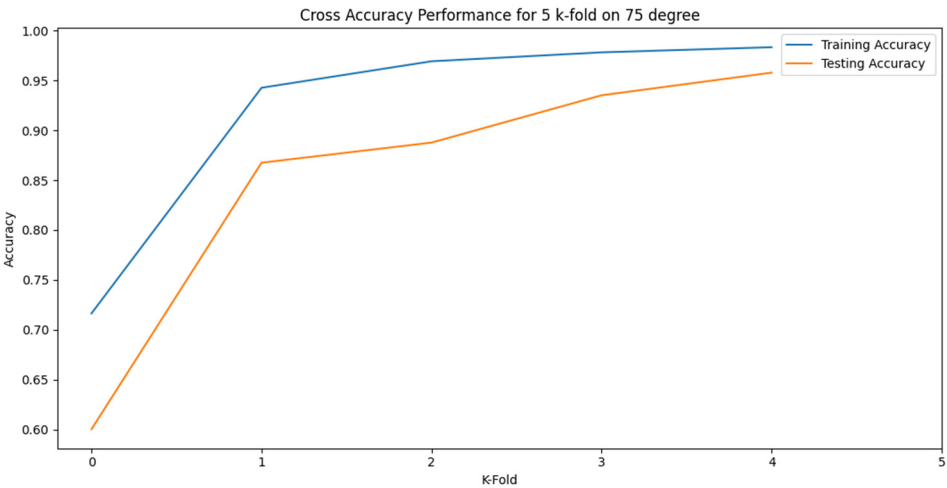


Figure 17. crossover Accuracy performance of 5 k-Folds on 75°.

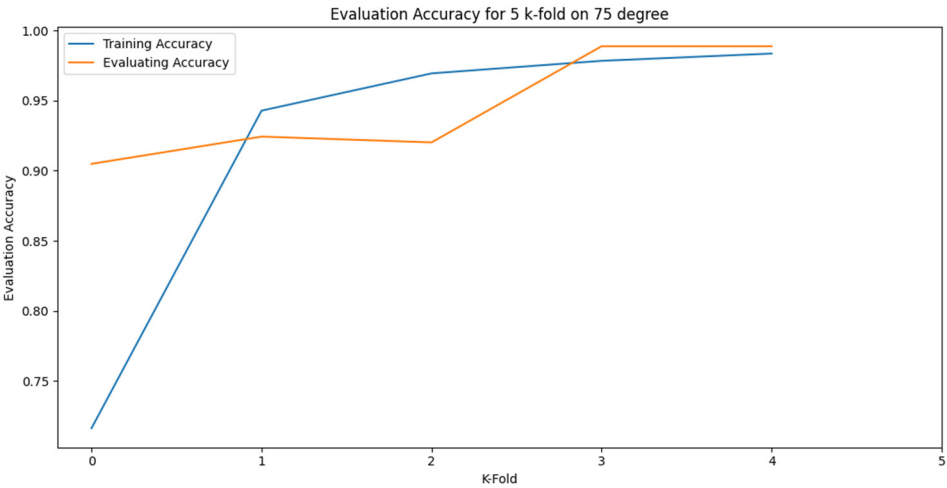


Figure 18. Evaluation Accuracy of 5 k-Folds on 75°.

Moreover, the model is compared with other previous studies’ approaches related to the same field (age estimation) using the same dataset for all degrees. The proposed technique achieved higher accuracy compared to the state-of-the-art methods. Table 3 shows the best accuracy score results for the proposed approach compared with other studies.

Table 3. Comparison results Accuracy best score for all degrees.

Approach	0°	15°	30°	45°	60°	75°	90°	180°	195°	210°	225°	240°	255°	270°	AVG
Proposed ViTs16[17]	78.00	88.06	91.11	90.69	86.80	87.69	85.86	83.20	90.57	92.24	91.88	87.10	88.32	86.36	87.71
GaitSet [19]	79.50	87.90	89.90	90.20	88.10	88.70	87.80	81.70	86.70	89.00	89.30	87.20	87.80	86.20	87.14
SelfGait [20]	85.10	89.30	92.00	94.30	89.10	90.2	90.90	87.40	91.80	89.30	88.70	90.80	91.60	87.70	89.87
Zhang et al. [21]	74.00	88.30	94.60	95.40	88.00	91.30	90.00	76.70	89.50	95.00	94.90	88.00	90.80	89.80	89.02
Our Approach	91.26	90.40	94.68	90.03	76.64	94.54	93.27	93.31	89.59	87.99	89.24	92.66	88.97	92.3	90.35

6. Conclusion

In this study, we developed a novel method for estimating a person's age by observing their stride. To preprocess the proposed gait dataset in a new way, We combined the GEI, the AFDEI, and the invariant moment of the picture. The proposed method achieved higher accuracy compared with state-of-the-art methods. The proposed method was tested using CNN within 5 K-fold, and its average accuracy across 14 distinct camera views was 90.35%, suggesting that this method has great potential for use in predicting a person's age based on their gait. The highest results score was 94.68% at 30° view degree, and the minimum score was 76.64 % at 244° view degree.

References

1. N. A. Makhdoomi, T. S. Gunawan, and M. H. Habaebi, "Human gait recognition and classification using similarity index for various conditions," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 53, no. 1, 2013, doi: 10.1088/1757-899X/53/1/012069.
2. B. Jawed, O. O. Khalifa, and S. S. Newaj Bhuiyan, "Human Gait Recognition System," *Proc. 2018 7th Int. Conf. Comput. Commun. Eng. ICCCE 2018*, pp. 89–92, 2018, doi: 10.1109/ICCCE.2018.8539245.
3. X. Wang, J. Wang, and K. Yan, "Gait recognition based on Gabor wavelets and (2D)2PCA," *Multimed. Tools Appl.*, vol. 77, no. 10, pp. 12545–12561, 2018, doi: 10.1007/s11042-017-4903-7.
4. A. Kharb, V. Saini, Y. Jain, and S. Dhiman, "A review of gait cycle and its parameters," *IJCEM Int J Comput Eng Manag*, vol. 13, no. July, pp. 78–83, 2011.
5. A. Bonnefoy-Mazure and S. Armand, "Orthopedic management of children with cerebral palsy," in *Pediatric Clinics of North America*, vol. 40, no. 3, 2015, pp. 645–657. doi: 10.1016/S0031-3955(16)38556-X.
6. Microgate Srl, "Gait Analysis," 2014. <http://www.optojump.com/Applications/Gait-Analysis.aspx>
7. D. S. Kumar, N. Nagarajan, and A. T. Azar, "Bio-inspiring Cyber Security and Cloud Services: Trends and Innovations," *Intell. Syst. Ref. Libr.*, vol. 70, no. November, pp. 249–264, 2014, doi: 10.1007/978-3-662-43616-5.
8. RECFACES BLOG, "Gait Recognition System: Deep Dive into This Future Tech," 2021. <https://recfaces.com/articles/what-is-gait-recognition>
9. D. THAKKAR, "Gait Recognition Systems Can Identify You with Your Manner of Walking," *Bayometric*, 2022. <https://www.bayometric.com/gait-recognition-identify-with-manner/>
10. X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, "Gait-based human age estimation using age group-dependent manifold learning and regression," *Multimed. Tools Appl.*, vol. 77, no. 21, pp. 28333–28354, 2018, doi: 10.1007/s11042-018-6049-7.
11. A. Sakata, N. Takemura, and Y. Yagi, "Gait-based age estimation using multi-stage convolutional neural network," *IPSJ Trans. Comput. Vis. Appl.*, vol. 11, no. 1, 2019, doi: 10.1186/s41074-019-0054-2.
12. N. S. Russel and A. Selvaraj, "Gender discrimination, age group classification and carried object recognition from gait energy image using fusion of parallel convolutional neural network," *IET Image Process.*, vol. 15, no. 1, pp. 239–251, 2021, doi: 10.1049/ipr2.12024.
13. C. Xu *et al.*, "Uncertainty-Aware Gait-Based Age Estimation and its Applications," *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 3, no. 4, pp. 479–494, 2021, doi: 10.1109/TBIOM.2021.3080300.
14. X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, "Make the Bag Disappear: Carrying Status-invariant Gait-based Human Age Estimation using Parallel Generative Adversarial Networks," *2019 IEEE 10th Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2019*, 2019, doi: 10.1109/BTAS46853.2019.9185973.
15. H. Zhu, Y. Zhang, G. Li, J. Zhang, and H. Shan, "Ordinal distribution regression for gait-based age estimation," *Sci. China Inf. Sci.*, vol. 63, no. 2, pp. 1–14, 2020, doi: 10.1007/s11432-019-2733-4.
16. J. Luo, J. Zhang, C. Zi, Y. Niu, H. Tian, and C. Xiu, "Gait Recognition Using GEI and AFDEI," *Int. J. Opt.*, vol. 2015, no. 4, pp. 1–5, 2015, doi: 10.1155/2015/763908.
17. D. Pinčić, D. Sušanj, and K. Lenac, "Gait Recognition with Self-Supervised Learning of Gait Features Based on Vision Transformers," *Sensors*, vol. 22, no. 19, 2022, doi: 10.3390/s22197140.
18. A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Oct. 2020, [Online]. Available: <http://arxiv.org/abs/2010.11929>
19. H. Chao, Y. He, J. Zhang, and J. Feng, "GaitSet: Regarding gait as a set for cross-view gait recognition," *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, no. 16, pp. 8126–8133, 2019, doi: 10.1609/aaai.v33i01.33018126.
20. Y. Liu, Y. Zeng, J. Pu, H. Shan, P. He, and J. Zhang, "Selfgait: A spatiotemporal representation learning method for self-supervised gait recognition," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2021-June, pp. 2570–2574, 2021, doi: 10.1109/ICASSP39728.2021.9413894.
21. Y. Zhang, Y. Huang, S. Yu, and L. Wang, "Cross-view gait recognition by discriminative feature learning," *IEEE Trans. Image Process.*, vol. 29, no. July 2019, pp. 1001–1015, 2020, doi: 10.1109/TIP.2019.2926208.

22. X. Chen, X. Luo, J. Weng, W. Luo, H. Li, and Q. Tian, "Multi-View Gait Image Generation for Cross-View Gait Recognition," *IEEE Trans. Image Process.*, vol. 30, pp. 3041–3055, 2021, doi: 10.1109/TIP.2021.3055936.
23. T. Teepe, J. Gilg, F. Herzog, S. Hormann, and G. Rigoll, "Towards a Deeper Understanding of Skeleton-based Gait Recognition," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2022-June, pp. 1568–1576, 2022, doi: 10.1109/CVPRW56347.2022.00163.
24. C. Xu, Y. Makihara, G. Ogi, X. Li, Y. Yagi, and J. Lu, "The OU-ISIR Gait Database comprising the Large Population Dataset with Age and performance evaluation of age estimation," *IPSJ Trans. Comput. Vis. Appl.*, vol. 9, no. 1, 2017, doi: 10.1186/s41074-017-0035-2.
25. T. Stöckel, R. Jacksteit, M. Behrens, R. Skripitz, R. Bader, and A. Mau-Moeller, "The mental representation of the human gait in young and older adults," *Front. Psychol.*, vol. 6, Jul. 2015, doi: 10.3389/fpsyg.2015.00943.
26. C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian, "Frame difference energy image for gait recognition with incomplete silhouettes," *Pattern Recognit. Lett.*, vol. 30, no. 11, pp. 977–984, 2009, doi: 10.1016/j.patrec.2009.04.012.
27. D. Kim, "Sobel Operator and Canny Edge Detector," pp. 1–10, 2013, [Online]. Available: <http://www.egr.msu.edu/classes/ece480/capstone/fall13/group04/docs/danapp.pdf>
28. V. Torre and T. A. Poggio, "On Edge Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 2, pp. 147–163, 1986, doi: 10.1109/TPAMI.1986.4767769.
29. M. . M.Radha, "Edge Detection Techniques for," *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 6, pp. 259–267, 2011.
30. M. W. Nasrudin, N. S. Yaakob, N. A. Abdul Rahim, M. Z. Zahir Ahmad, N. Ramli, and M. S. Aziz Rashid, "Moment Invariants Technique for Image Analysis and Its Applications: A Review," *J. Phys. Conf. Ser.*, vol. 1962, no. 1, 2021, doi: 10.1088/1742-6596/1962/1/012028.
31. J. Flusser, T. Suk, and B. Zitová, *Moments and Moment Invariants in Pattern Recognition*. 2009. doi: 10.1002/9780470684757.
32. K. Yamazaki, V. K. Vo-Ho, D. Bulsara, and N. Le, "Spiking Neural Networks and Their Applications: A Review," *Brain Sci.*, vol. 12, no. 7, pp. 1–30, 2022, doi: 10.3390/brainsci12070863.
33. J. Guo *et al.*, "CMT: Convolutional Neural Networks Meet Vision Transformers," pp. 12165–12175, 2022, doi: 10.1109/cvpr52688.2022.01186.
34. M. Lango and J. Stefanowski, "What makes multi-class imbalanced problems difficult? An experimental study," *Expert Syst. Appl.*, vol. 199, no. January, p. 116962, 2022, doi: 10.1016/j.eswa.2022.116962.
35. J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0192-5.
36. Amazon, "Amazon Machine Learning Developer Guide." 2022.
37. C. Barile, C. Casavola, G. Pappalettera, and V. P. Kannan, "Damage Progress Classification in AlSi10Mg SLM Specimens by Convolutional Neural Network and k-Fold Cross Validation," *Materials (Basel)*, vol. 15, no. 13, 2022, doi: 10.3390/ma15134428.
38. Great Learning Team, "What is Cross Validation in Machine learning? Types of Cross Validation," 2022. <https://www.mygreatlearning.com/blog/cross-validation/>
39. V. Lyashenko and A. Jha, "Cross-Validation in Machine Learning: How to Do It Right," 2022. <https://neptune.ai/blog/cross-validation-in-machine-learning-how-to-do-it-right>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.