IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 5, October, 2023, pp. 2179-2190
Journal homepage: http://iieta.org/journals/ts

Integration of Face and Gait Recognition via Transfer Learning: A Multiscale Biometric Identification Approach

Dindar M. Ahmed*, Basil Sh. Mahmood

Information Technology Department, Technical College of Informatics-Akre, Duhok Polytechnic University, Duhok 42001, Iraq

Corresponding Author Email: Dindar.ahmed@dpu.edu.krd

https://doi.org/10.18280/ts.400535

Received: 18 March 2023
Revised: 12 July 2023
Accepted: 22 August 2023
Available online: 30 October 2023

Keywords:
recognition, face, gait, transfer learning, fusion

ABSTRACT

The ubiquity of biometric identification systems and their applications is evident in today's world. Among various biometric features, face and gait are readily obtainable and thus hold significant value. Advances in computational vision and deep learning have paved the way for the integration of these biometric features at multiple scales. This study introduces a system for biometric recognition that synergises face and gait recognition through the lens of transfer learning. Feature extraction was accomplished using Inception_v3 and DenseNet201 algorithms, while classification was performed employing machine learning algorithms such as K-Nearest Neighbours (KNN) and Support Vector Classification (SVC). A unique dataset was constructed for this research, consisting of face and gait information extracted from video clips. The findings underscore the efficacy of integrating face and gait recognition, primarily through feature and score fusion, resulting in enhanced recognition accuracy. Specifically, the Inception_v3 algorithm was found to excel in feature extraction, and SVC was superior for classification purposes. The system achieved an accuracy of 98% when feature-level fusion was performed, and 97% accuracy was observed with score fusion using Decision Trees. The results highlight the potential of transfer learning in advancing multiscale biometric recognition systems.

1. INTRODUCTION

Face recognition, one of the most extensively studied biometric features within the realm of computer vision, has found applications in a myriad of areas, ranging from security and verification to tracking [1, 2]. The widespread usage of face recognition as a biometric feature is attributed to its accessibility, facilitated by inexpensive cameras, and its general acceptance by the public [3]. However, despite these advantages, face recognition is not without its limitations. Factors such as ageing, facial hair, cosmetics, glasses, and masks can alter facial appearance, while external influences such as lighting, noise, image quality, camera positioning, and the angle of photography could compromise the system's discriminatory capacity [4].

In parallel, gait, another key biometric feature, has been the focus of numerous studies investigating its utility in distinguishing individuals [5-7]. Similar to facial recognition, gait is a visual signal extractable from video footage, thereby offering comparable benefits [8]. A significant advantage of gait over face recognition is that it can be ascertained from a sequence of low-resolution images of individuals captured from a distance, where the person's body occupies relatively few pixels. This allows for the extraction of other biometric information [9, 10]. Nevertheless, gait, being a behavioural attribute, is also susceptible to variations due to factors such as clothing, footwear, environmental conditions, emotions, fatigue, inebriation, pregnancy, injury, disease, and age [11, 12]. Additionally, gait recognition grapples with common challenges related to visual signal extraction from videos, such as segmentation, inadequate separation of the walking subject from the background scene, and poor recording conditions [13, 14].

Recent years have seen the emergence of methods integrating face and gait recognition, combining physical and behavioural biometrics to explore whether this amalgamation can enhance the performance of systems that utilise only one of these biometrics [15, 16]. Although this multi-biometric combination is still in its nascent stage, with relatively few studies published on the topic, the findings thus far are promising, showing clear potential for this approach in differentiating individuals [17].

This study proposes a dual face and gait recognition system for distinguishing individuals, integrating at the level of feature extraction using deep learning and at the output level of separate face and gait models. The present paper reviews the existing literature on face and gait recognition, provides a theoretical overview of the algorithms and techniques employed, and discusses the dataset created for the fusion of face and gait.

The main contributions of this work are fourfold:

1. The creation of a robust and replicable dataset for face and gait recognition.
2. The application of transfer learning principles for feature extraction from the dataset.
3. The construction and efficacy evaluation of a combined face and gait recognition system, compared to isolated face and gait recognition systems.

4. The harnessing of transfer learning techniques during the feature extraction stage, and the testing of multiple classifiers to identify the most accurate model.

2. LITERATURE REVIEW

The field of biometric recognition systems, utilizing facial and gait data, has been the subject of extensive scholarly exploration. A number of studies have proposed the combination of these two biometrics, aiming to enhance the accuracy and robustness of person identification.

Zakaria et al. [18] proposed a Convolutional Neural Network (CNN) consisting of 15 layers. The efficacy of this CNN was assessed using the YouTube Faces database and was subsequently compared to the Principal Component Analysis (PCA) and FHMM methods utilizing the ORL face database.

In a different study, Tabassum et al. [19] applied Discrete Wavelet Transform (DWT) coupled with PCA for feature extraction. They proposed a fusion of the CNN results using detection probability entropy and a fuzzy system, achieving an identification rate of 89.56% in the worst-case scenario and 93.34% in the best case.

A novel approach was implemented by Mehmood et al. [20], who proposed Human Gait Recognition (HGR) from various viewing angles. Their methodology involved the use of the pre-trained Densenet-201 CNN model for feature extraction, feature reduction based on a mixed selection method, and learning through four major stages of supervised learning techniques. The crucial stage was the extraction of CNN features, with the most active features being simultaneously combined from the second and third layers. Subsequently, the Firefly algorithm and skew-based technology were employed for feature selection. They achieved an accuracy of 94.7% at an angle of 180 degrees.

Mogan et al. [21] proposed a hybrid model merging multilayer perceptrons with the pre-trained DenseNet-201 for the CAS, OU-ISIR D, and OU-ISIR databases. Their approach involved extracting the gait energy image, obtaining representative features, and applying transfer learning from the pre-trained DenseNet-201 model. A multilayer perceptron was used to identify correlations between these features, which were then assigned to appropriate class labels by a classification layer, achieving an accuracy of 92.22%.

In a study by Liu and Liu [22], they used the CASIA-B gait dataset and the UCMP-GAIT dataset. Their Two-Stream neural network (TS-Net) model concurrently extracted dynamic deep features from gait images and static invariant features from multiple resolutions. The goal of person recognition was transformed into a binary classification problem through similarity learning techniques, achieving an accuracy of 92.22%.

Wang and Yan [23] proposed a method based on convolutional long short-term memory (Conv-LSTM) and applied it to the OU-ISIR LP and CASIA datasets. They introduced a modification of Gait Energy Images (GEI), referred to as frame-by-frame GEI (ff-GEI), to increase the available gait data and relax the restrictions of gait cycle segmentation. Their study demonstrated the effectiveness of ff-GEI by analysing the cross-covariance of a single person's gait data, achieving a Correct Recognition Rate (CRR) average of 95.9%.

Aung and Pluempitiwiriyawej [24] applied a deep convolutional neural network (CNN) technique for gait biometric person identification, using Gait Energy Images (GEI) of individuals, and utilized the CASIA-B gait dataset. Their empirical findings indicated superior recognition efficacy when compared to other advanced machine learning models.

Punyani et al. [25] introduced "double-level fusion" that combined data and output levels. PCA and MLG were applied to the OU-ISIR Gait, Large Population, and USF Gait datasets to reduce the dimensionality arising from integrating facial features with gait. The KNN algorithm was utilized for recognition. Their results showed that the Mean Absolute Error (MAE) reached 6.11.

Rahman et al. [26] configured a biometric system combining facial and gait biometrics at two levels (rank and score level fusion). The Histogram of Oriented Gradients (HOG) method was applied for facial feature extraction, while structural features were used for gait feature extraction. Their highest fusion result was 96.67% for logistic regression rank-level fusion.

Aung et al. [27] proposed the use of a CNN algorithm with transfer learning to extract gait and facial features. They applied the feature-level fusion method to merge facial and gait features, and then applied various classification algorithms, achieving the highest accuracy of 97.3% using the logistic regression (OvR) algorithm.

In the study by Maity et al. [28], a system was proposed for accurate multimodal human identification from low-resolution video surveillance footage, using a single biometric data source for low-resolution face and frontal gait recognition. The Adaboost detector was used for automatic detection of low-resolution face images, while a quick object segmentation algorithm was used to segment frontal gait binary silhouettes. The low-resolution face images were pre-processedusing super-resolution techniques to generate a high-resolution representation. This was followed by normalization of lighting and poses and image synthesis via registration. Gabor and Local Binary Patterns (LBP) features were then extracted from the synthetic face images. The nearest neighbor classifier was employed to accomplish rank-1 recognition for each modality. Subsequently, score-level fusion was used to combine the results of each separate recognition process. Their results showed that the combined use of low-resolution face and frontal gait modalities led to the highest rank-1 recognition accuracy, compared to the performance of each modality independently.

Manssor et al. [29] employed the YOLO-face algorithm along with the Eigenfaces approach for recognizing faces. For gait recognition, they applied the YOLO algorithm in tandem with Hidden Markov Models (HMMs). The outputs of the face and gait recognition algorithms were combined using decision-level fusion. The YOLOv3-Human model was trained using the DHU Night, FLIR, and KAIST databases.

In summary, these studies provide a comprehensive overview of the advancements and methodologies utilized in the field of biometric recognition systems, specifically focusing on facial and gait data. The diverse approaches employed in these studies have contributed to enhancing the accuracy and robustness of person identification systems, thereby paving the way for future research in this domain.

3. THEORETICAL BACKGROUND

This section summarizes the theoretical background of the algorithms used in data initialization, feature extraction, and classification.

3.1 Face and gait detection and tracking

A set of detection and tracking algorithms were used to extract the gait and face images. To extract the gait images from the video clips. The Mixture of Gaussian (MOG) algorithm was used. After background separation, the gait images were tracked using a Histogram of Oriented (HOG). As for extracting images of faces, the Haar Cascade Algorithm was used.

3.1.1 Haar Cascade algorithm

Haar Cascade is an Object Detection Algorithm that detects faces in images and real-time videos. "haar" refers to a rectangle-shaped mathematical function [30]. Haar is a single rectangular wavelet (one high and one low interval). It features one bright and one dark side in two dimensions (2D) [31]. The purpose of cascade classification is to incorporate additional features efficiently. Initially, haar's image processing only depends on each pixel's RGB value, after which the picture is processed in rectangle forms with specific pixels in each shape. Each form is analyzed, and the limit level (threshold) is reached, revealing the dark and bright areas [32]. The haar feature formula states that if the average value of the result is more than the threshold, the haar feature exists. To train, the algorithm is given many positive photos with faces and many negative images with no faces. The pixels with values 1 are darker in the haar feature, whereas those with 0 are brighter. Each is in charge of identifying a particular feature in the picture: an edge, a line, or any other structure where the intensities abruptly shift [33].

3.1.2 Mixture of gaussian (MOG) algorithm

It is a background/foreground segmentation algorithm based on a Gaussian mixture. Each backdrop pixel is modeled using a technique combining K Gaussian distributions. The complexity reduction threshold is returned [34].

3.1.3 Histogram of oriented (HOG) algorithm

It is one of the feature descriptors. The histogram of the directed gradients descriptor may characterize the distribution of intensity gradients or edge directions in a picture. The picture is split into cells, which are little linked areas [35]. A cell may include many pixels, creating a gradient histogram for each pixel. Each pixel's gradient is represented as a histogram in the description. This is done by calculating intensity across a wider region of many cells known as a block and then using that value to normalize all cells inside that block. The adjusted result performs better under different lighting and intensity conditions [36]. HOG descriptors provide several benefits over other descriptors, including that they are invariant to geometric and photometric modifications except for object orientation [37].

3.2 Transfer learning

When faced with specific application scenarios in image classification, it is often impossible to obtain labeled data of the scale required to build a neural network model [38]. Transfer learning proposes solving such cross-domain learning problems by extracting useful information from related domains and transferring it for use in target tasks. In addition to data in the target domain, related data in different fields can also be included to extend the availability of prior knowledge of the target future data [39]. The data set is far from enough to retrain the deep neural network. Therefore, with the help of transfer learning, the pre-trained model can be used when the training data set is small. By truncating the bottleneck layer of the pre-trained network, the useful neurons of the reusable layer are retained to mine more classification features. In addition, migration learning can keep the training data in the same feature space or have the same distribution as the future data, avoiding the problem of overfitting [40].

3.2.1 Inception_v3

The Inception-v3 model (see Figure 1) improves three Inception modules based on v2: use two 3×3 convolutions instead of each 5×5 convolution, and decompose the n×n convolution into one-dimensional n×1 and Concatenation of 1×n convolutions [41]. Compressing the dimensionality of features facilitates the representation of high-size images and alleviates the overfitting phenomenon. Fully connected layers are replaced by global average pooling layers, greatly reducing the number of parameters [42]. The general architecture of Inception-v3 is shown in Figure 1.

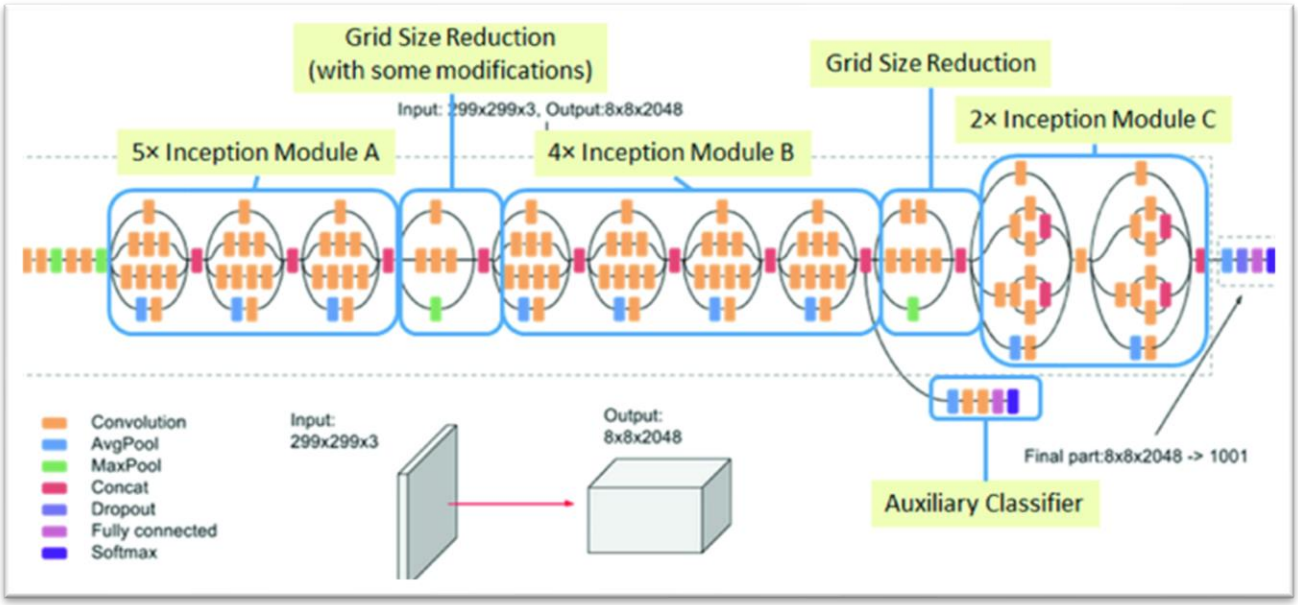

Figure 1. The inception V3 model's structure [43]

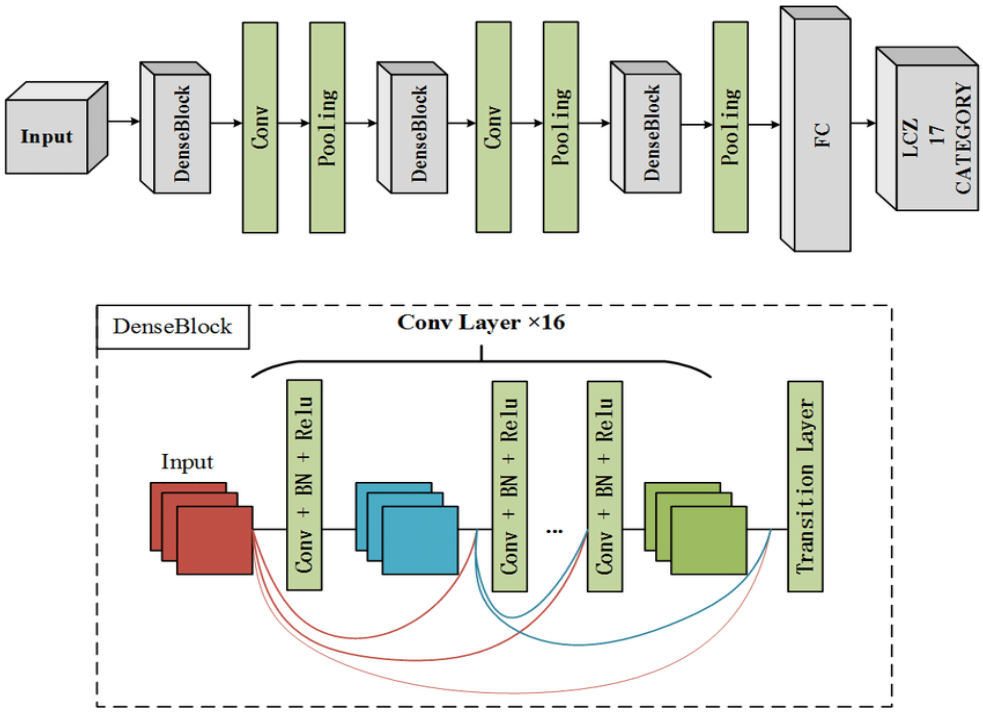

Figure 2. The DenseNet-201 module's structure

3.2.2 DensNet201

The inception V3 model's structure to solve gradient disappearance to a greater extent, enhance feature transfer, use features more effectively, and reduce a certain number of parameters, the DenseNet201 network structure directly connects all layers to ensure the maximum complete transmission between layers in the network (See Figure 2) [44]. In recent years, deep learning has achieved excellent results in video and image processing [45].

The DenseNet-201 model includes sequentially connected convolutional layers, pooling layers, first dense block, first transition layer, second dense block, second transition layer, third dense block, third transition layer, fourth Dense blocks, and classification layers [46].

3.3 Machine learning

After feature extraction using transfer learning, machine learning algorithms were used for classification. This study compared K-Nearest Neighbors and Support Vector Machine at the feature merging level, while decision trees were used for merging at the model's level.

3.3.1 K-nearest neighbors (KNN) algorithm

The K-nearest neighbors (KNN) algorithm is a well-known statistical method for pattern identification and holds a significant position in machine learning classification algorithms. It is a supervised classification algorithm that can be rationalized and selected based on its simplicity, effectiveness, and ability to handle non-linear decision boundaries. The fundamental tenet of the KNN algorithm is that a sample belongs to a category and exhibits the traits of models in that category if the majority of its k nearest neighbor samples in the feature space do the same. This approach makes the classification decision based on the category of one or more nearest neighbor samples, which allows for flexibility in capturing complex patterns in the data [47, 48].

3.3.2 Support vector machine (SVM) algorithm

A support vector machine is a supervised learning model and related learning algorithm for analyzing data in classification and regression analysis [49]. A support vector classifier (SVC) creates a hyperplane or set of hyperplanes in high-dimensional space to divide training data into various categories. These are then utilized to categorize the complete image or a collection of images [50].

3.3.3 Decision tree

The decision tree algorithm is one of the supervised learning algorithms. It can be used to solve both regression and classification problems [51]. The purpose of using a decision tree is to create a trained model that classifies the value of a target variable by learning simple decision rules inferred from previous data (training data). It represents a mapping relationship between object attributes and object values. Each node in the tree represents an object, and each branch path represents a possible attribute value. Each leaf node corresponds to the object's value represented by the path experienced from the root node to the leaf node [52].

4. EXPERIMENT AND SUGGESTED METHODOLOGY

A framework has been proposed to identify humans using the face and gait with several stages. In the first stage, a dataset of the face and gait is created for the same people, where two digital cameras were used (to record) video clips. The first camera (the front camera) photographed the person as he passed from the front, and the second (recorded) video clips from a side view of the person for Recognition of the gait. The viewing area of the front and side cameras was ten meters, and 10-second video clips were recorded. The experiment was conducted at the Duhok Technical Institute and the Shekhan Technical Institute / Polytechnic University of Duhok for 65 volunteers. In the second stage, the videos for each person were divided into two

videos, the first for the face and the second for the gait. After creating the videos, the face and gait videos were processed separately. To extract facial features, the Haar Cascade algorithm was used to track and detect facial images from front camera videos and performed image pre-processing that included standardization of size, centering, and rotation according to the position of the face and using 100 facial images for each person. After pre-processing, pre-training algorithms were applied to the face images using (inception_v3 and DenseNet201) algorithms. They extracted the features specific to the face and the number of faces particular to each person. As for gait and extracting gait features, the background was initially isolated from the person's body using the MOG2 algorithm to convert the video clip of the gait to white as the person's body and a black background. Then, the HOG algorithm will track the person's gait within the video clip. The result is a set of sequential cut images of the gait with sizes according to the distance and proximity of the person from the camera's position. At this stage, image pre-processing was used to standardize the sizes and focus the images. The GEI gait energy finding technique was used to convert gait images into distinguishable images, which collects chromatic units between images and adopts a new gait energy image for each image. The same facial feature extraction algorithms were used in the final stage of feature extraction. As shown in Figure 3.

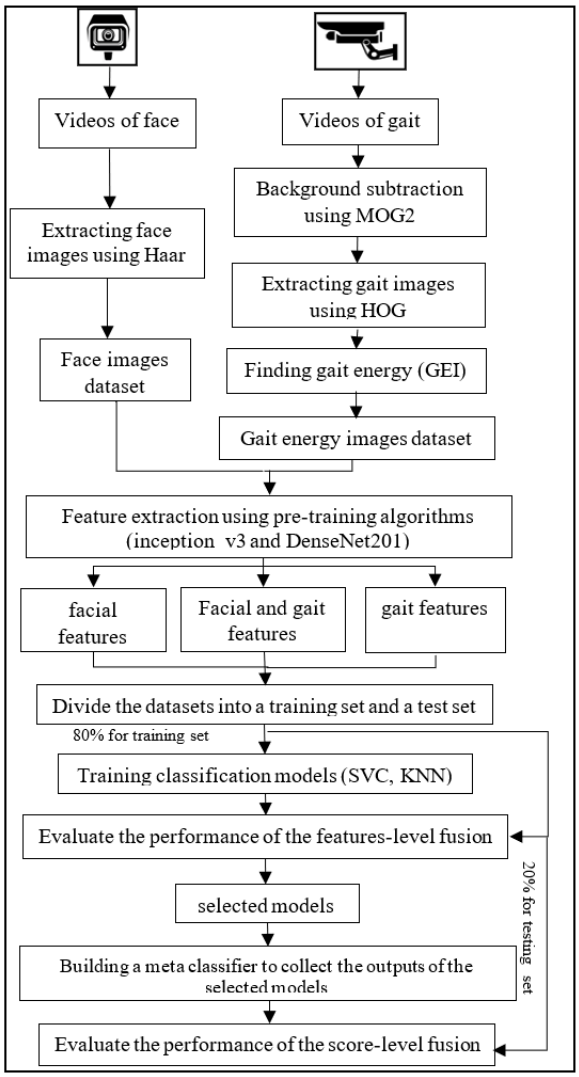

Figure 3. Suggested methodology for face and gait recognition

After obtaining the facial and gait features, three sets of features (face, gait, face, and gait) are established. This stage is called (features-level fusion). After the feature extraction phase, the data set was divided into (80:20) a training set and a test set. The selected machine learning algorithms are trained for classification and person identification. The algorithms (SVC, KNN) were trained on the training data set for face once, gait energy once, facial features, and gait once, testing the models using test and comparison data and determining the most accurate model for each case. In the last stage, the outputs of the selected models are combined with a meta-classifier to create a model that combines the biometric features at the model output level, called (score-level fusion) and testing the model using the test data.

4.1 Create a dataset for face and gait

65 volunteers from the Polytechnic Institute in Dohuk – Information Technology Department were used to create a data set to distinguish humans by face and gait. Two digital cameras with a resolution of 108 megapixels and a recording space of 10 * 7 meters were installed. A special camera for recording video clips of the face was installed at the end. Recording space While the gait recording camera was placed in the center of the recording space to take the complete steps of the gait. As shown in Figure 4.

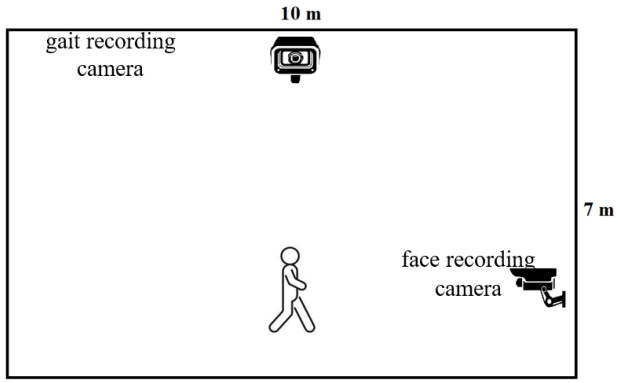

Figure 4. Recording space environment

Humans were passed in the recording space at a rate of 10 seconds per person and a size of 75 megabytes per video clip, and videos are stored in MP4 format. Experiments were conducted in the natural environment of the institute's corridors. Table 1 shows a description of the recorded video clips.

Table 1. Description of the recorded video clips

Characteristic	Value
Average Length Per Clip	00:07:00
Frame Width (Pixcel)	1920
Frame Height (Pixcel)	1080
Data Rate	21962kbps
Frame Rate (Kbps)	30.00 frames/second
Number Of Clip	65*2
Format	mp4
Data Size For Face	4.49 GB
Data Size For Gait	3.37 GB

4.2 Feature extraction

To integrate more than one biometric feature at the level of

features, the gait, and face must be detected and tracked as an initial stage for the recognition system to work.

4.2.1 Face images detection and tracking

Haar Cascade is an Object Detection Algorithm that detects faces in images and real-time videos. "haar" refers to a rectangle-shaped mathematical function [30]. Haar is a single rectangular wavelet (one high and one low interval). It features one bright and one dark side in two dimensions (2D) [31]. The purpose of cascade classification is to incorporate additional features efficiently. Initially, haar's image processing only depends on each pixel's RGB value, after which the picture is processed in rectangle forms with specific pixels in each shape. Each form is analyzed, and the limit level (threshold) is reached, revealing the dark and bright areas [32]. The haar feature formula states that if the average value of the result is more than the threshold, the haar feature exists. To train the algorithm, a large number of positive photos with faces and a large number of negative images with no faces are provided. The pixels with values 1 are darker in the haar feature, whereas those with 0 are brighter. Each is in charge of identifying a particular feature in the picture: an edge, a line, or any other structure where the intensities abruptly shift [33]. It is important to note that the training results have been verified using a significant amount.

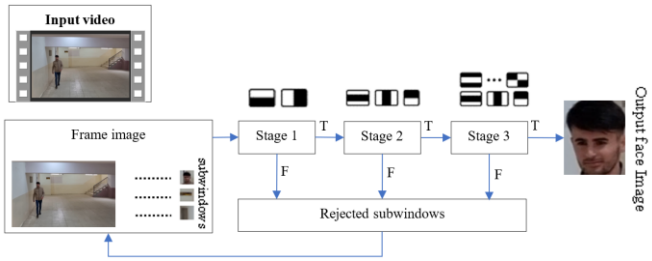

Figure 5. Extracting face images using Haarcascade method

Figure 5 shows that the learning algorithm's definition of a set of Haar-like characteristics makes up each cascade level. In the first phases, classifiers are trained to recognize almost all candidates resembling faces while rejecting most negative sub-windows. This design can significantly speed up the discovery process because most negatives may be eliminated during the first two or three stages. Most computing resources are concentrated on face-like sub-windows. Stage classifiers systematically assess each sub-window, combining the results into the stage for each Haar-like feature. To assess if the present sub-window is a face-like filter, all the characteristics at a location are computed, and the aggregated value is compared with the stage boundary value. Only if the preceding step is active is the succeeding stage triggered.

4.2.2 Gait images detecting and tracking

The stage of extracting gait energy images goes through several stages, starting with background subtraction, converting the video clip to a black background, and representing the person walking in white color using the MOG2 algorithm. In the second stage, the rectangle of the gait image is detected and converted into a gait image for each frame in the gait video. This process is done using HOG. This stage is followed by calculating gait energy and creating a gait energy dataset. This study used the function createBackgroundSubtractorMOG2 provided in the OpenCV package for Background subtraction. Figure 6 shows the gait segments before and after background subtraction.

Figure 6. Background subtraction for gait image

To track and extract images of the person's gait HOGDescriptor() function was used. The default human detector size is 64×128, so the size of the person who wants to detect it must be at least 64×128. Figure 7 shows the images extracted from the video clip.

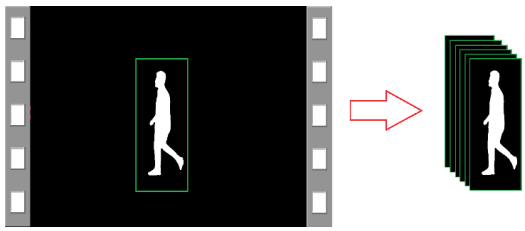

Figure 7. Extracting gait image

Then, the set SVMDetector function was used to set the Support Vector Machine to pre-train the gait detector and load it with the cv2.HOGDescriptor_get Default People Detector() function. Another important function used is DiscoverMultiScale(). This function implements a detection phenomenon through a multi-scale window. To extract the gait energy, the center of mass of the image must be determined before it is collected. p=(i; j) represents the pixel value in row i and column j. To find the gait center of mass, the brightest area must be determined, and it can be calculated by taking the average value of x and y and determining the place with the most density. Eqs. (1), (2), and (3) can be used to find d and, $\hat{y}$ which are the coordinates of the gait center of mass.

$$\hat{x} = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n} j * p(i,j)}{r} \quad (1)$$

$$\hat{y} = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n} i * p(i,j)}{r} \quad (2)$$

$$r = \sum_{i=0}^{m}\sum_{j=0}^{n} p(i,j) \quad (3)$$

The calculation method of each feature, when the brightness value at the pixel (x, y) of the frame of the normalized silhouette sequence is cap I. open paren *x, y, t*, close paren, And the walking cycle in *P*, is as follows. GEI is created by finding the average of the silhouette series within one walking cycle:

$$GEI(x,y) = \frac{1}{p}\sum_{t=1}^{p} I(x,y,t) \quad (4)$$

In GEI, the moving part becomes smaller by averaging the brightness values, so the pixel's shade expresses the movement's magnitude, and the part with less movement is expressed brightly.

As shown in Figure 8, the gait energy was found for every four frames in the video clip, and the gait energy was extracted as an image for recognition.

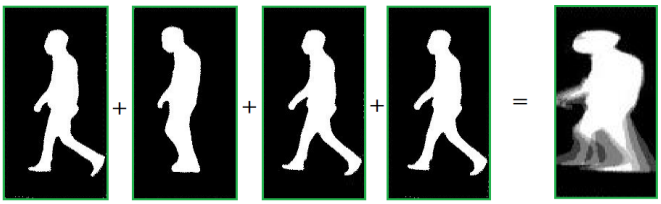

Figure 8. Gait energy image

4.3 Feature extraction using transfer learning

In this study, transfer learning is adopted, whereby a pre-trained model can be used as a feature extractor for any image of visual objects classified in computer vision. Transfer learning is a machine learning method that uses a pre-trained neural network. Two types of pre-trained deep learning algorithms (inception_v3 and DenseNet201) were used to extract gait and facial features for identification purposes.

4.3.1 Feature extraction using Inception_v3

Inception-v3 is a deep neural network, and the network consists of 11 units starting with five types in total. A pre-trained model was used to extract gait and facial features. The trained model was used in Python and Keras deep learning framework. Table 2 determines the algorithm's parameters and values.

Table 2. Inception_v3 pre-trained parameters

Parameters	Imagenet Vector	Employ Trained Weights
weights	Imagenet Vector	employ trained weights
include_top	False	determines whether to include the top, fully linked layer as the network's final.`
input_shape	(128, 128, 3)	Specifies the format of the network input. In terms of the number of pixels of the x-axis and the y-axis in addition to the color formula used
output_shape	(2, 2, 2048)	the vector represents the extracted values
Total params	21,802,784	numbers of weights during backpropagation training that are updated and not updated.
Trainable params	21,768,352	numbers of weights during backpropagation training that are updated.
Non-trainable params	34,432	numbers of weights during backpropagation training that are not updated.

The input to the algorithm was 100 images of both gait and face energy. RGB images with a size of 128 * 128 were used. The output of the mixed10 layer is adopted by the InceptionV3 algorithm. (2, 2, 2048) were extracted, equivalent to 8,192 features for each image for each image of faces and gait energy.

4.3.2 Feature extraction using DenseNet201

DenseNet201 is a deep neural network. This paper used a pre-trained model to extract gait and facial features. The Conv5 Block32 Concat layer was selected from the DenseNet model as the output layer. The trained model was used in Python and Keras deep learning framework. The parameters and values of the DenseNet201 algorithm were determined in Table 3.

Table 3. Densenet201 pre-trained parameters

Parameters	Imagenet Vector	Employ Trained Weights
weights	Imagenet Vector	employ trained weights
include_top	False	determines whether to include the top, fully linked layer as the network's final.`
input_shape	(128, 128, 3)	specifies the format of the network input. In terms of the number of pixels of the x-axis and the y-axis in addition to the color formula used
output_shape	(4, 4, 1920)	the vector represents the extracted values
Total params	18,321,984	numbers of weights during backpropagation training that are updated and not updated.
Trainable params	18,092,92	numbers of weights during backpropagation training that are updated.
Non-trainable params	229,056	numbers of weights during backpropagation training that are not updated.
Non-trainable params	229,056	numbers of weights during backpropagation training that are not updated.

The input to the algorithm was 100 images of both gait energy and face. RGB images with a size of 128 * 128 were used. The DenseNet201 algorithm adopts the output of the Conv5 Block32 layer. (4, 4, 1920) were extracted features for each image for each image of faces and gait energy.

4.4 Facial and gait fusion

The merger is done at one of two levels (features-level fusion and score-level fusion) to identify humans using multiple biometric features. Two models (SVC and KNN) were used to model the data and compare them for the first level, and for the second level, decision trees were used to combine the classifiers' outputs. Experiments were conducted using Python to build classification models within the Sklearn framework.

4.4.1 Recognize using features-level fusion

The combination of feature-level information provides more information about the person to be verified. The proposed feature level fusion combines pre-match biometric information of the face and gait. Feature-level merging is not widely used because it is difficult to integrate incompatible feature vectors from multiple methods. Therefore, in this study, it was proposed to use gait energy images with images of faces and to use the same feature extraction algorithms and create two datasets that are identical in terms of face and gait composition. The simplest form of feature-level fusion is the collection of data sets at the feature level. This type of merging doubles the number of distinctive features.

As in Figure 9, two biometric features of a person were combined to distinguish face and gait energy. In the first stage, facial features and gait energy were extracted using pre-training algorithms (inception_v3 and DenseNet201). The output of this stage is two datasets for extracted features; then, the two feature datasets are combined to create a new dataset that integrates facial features and gait energy. In the next stage, the selected machine learning models are trained to develop a classifier model.

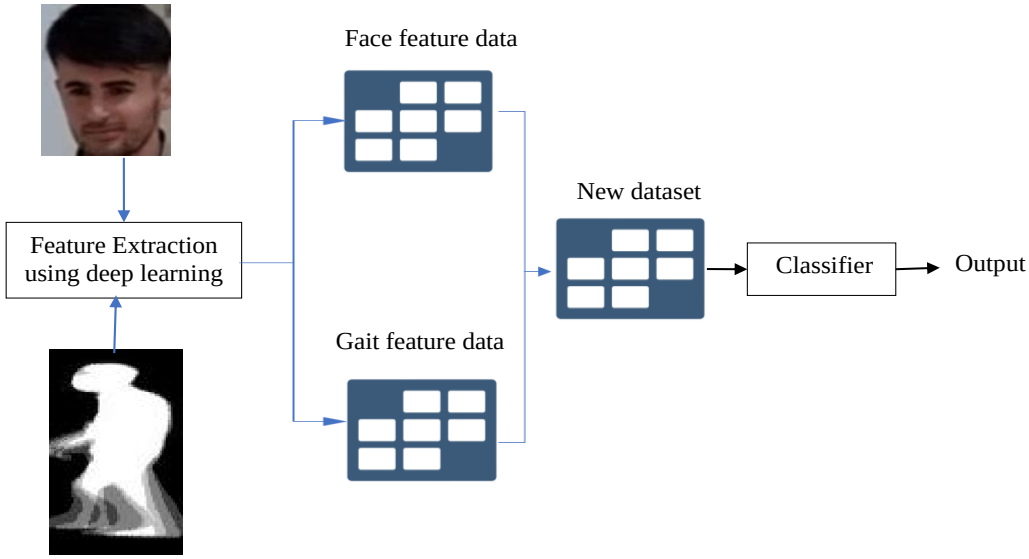

Figure 9. Block diagram of feature-level fusion

4.4.2 Recognize using score-level fusion

In the case of score-level fusion, the outputs of the classification models are merged using a supporting model. This method enables to collect the predictive values of several classifiers and give a final result. In this paper, as in the case of features-level fusion, facial features, and gait energy were extracted using deep learning, and two data sets were created, as in Figure 10.

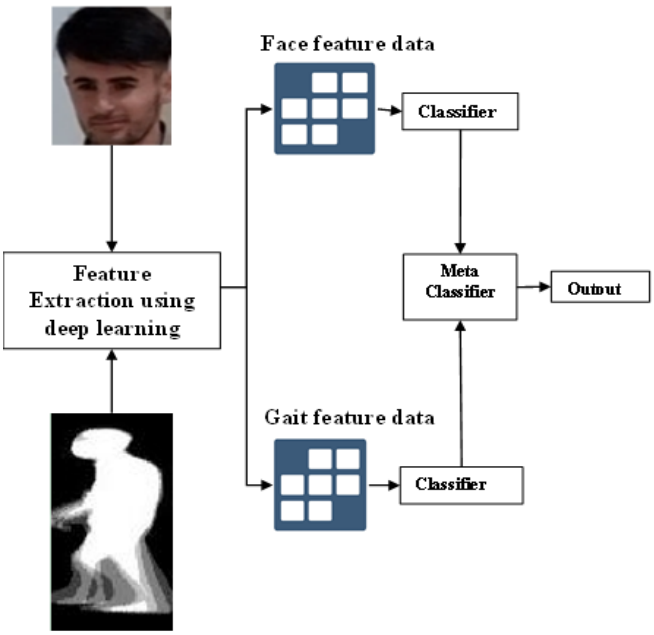

Figure 10. Block diagram of the score-level fusion

After extracting the facial features and gait energy, two models for classification are trained. The first model is trained on features extracted from the face, while the second model is trained on features extracted for gait energy. In the last stage, a final model is trained on the outputs of the two models.

5. RESULT AND DISCUSSION

5.1 Result

The suggested model was evaluated based on several scales: accuracy, recall, precision, and F-score. The Performance measures of identification shown in Equations (5, 6, 7, and 8) are calculated based on the number of correctly identified class instances (TP, true positives), the number of correctly identified class instances that do not belong to the class (TN, true negatives), and the number of instances that were either incorrectly assigned to the class (FP, false positives), or were not recognized as class instances (FN, false negatives).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

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

$$recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = 2 * \frac{Precision * recall}{precision + recall} \quad (8)$$

During testing and validation, the model is adjusted and reviewed. The face and gait were distinguished separately in the first stage. The Table 4 shows the result of face recognition.

Table 4. Algorithms performance based on face recognition

Method	Accuracy	Precision	Recall	f1-Score
Face recognition Using DenseNet201 and SVC	0.94	0.94	0.94	0.94
Face recognition Using DenseNet201 and KNN	0.96	0.96	0.96	0.96
Face recognition Using inception_v3 and KNN	0.94	0.94	0.94	0.94
Face recognition Using inception_v3 and KNN	0.95	0.95	0.95	0.95

Table 5. Algorithms performance based on gait regocnation

Method	Accuracy	Precision	Recall	f1-Score
Gait recognition Using DenseNet201 and SVC	0.96	0.70	0.68	0.67
Gait recognition Using DenseNet201 and KNN	0.77	0.78	0.79	0.77
Gait recognition Using inception_v3 and SVC	0.88	0.88	0.88	0.87
Gait recognition Using inception_v3and KNN	0.72	0.72	0.72	0.72

Table 6. Comparing the performance of algorithms when integrating biometric features (Face and gait)

Method	Accuracy	Precision	Recall	f1-Score
F. & GR Using DenseNet201 and SVC	0.92	0.92	0.92	0.92
F. & GR Using DenseNet201 and KNN	0.91	0.92	0.92	0.91
F. &G.R. Using inception_v3and KNN	0.92	0.93	0.92	0.92
F. &G. R. Using inception_v3and SVC (Feature Level Fusion)	0.98	0.98	0.98	0.98
F. &G. R. Using inception_v3and SVC (Score Level Fusion)	0.97	0.97	0.97	0.97

As shown in Table 4, the best performance is in the case of using DenseNet201 in feature extraction with KNN as a classifier, as it achieves 0.96 in various measures, followed by inception_v3 in feature extraction with KNN, where it was 0.95. The same algorithm were applied for gait recognition, as presented in Table 5.

And when distinguishing the gait using the same algorithms, the performance compared to the face was low, as the inception_v3 algorithm achieved 0.88 with SVC and 0.72 with KNN. At the same time, the DenseNet201 algorithm earned 0.77 using KNN and 0.67 using SVC using the accuracy scale. It is also noted that there is a slight imbalance when using DenseNet201 with SVC, where the precision value was 0.70, while the recall was 0.68.

In the next stage, facial and gait features were merged using both algorithms, DenseNet201, and inception_v3, and using both classifiers SVC and KNN, merging was also done at the output level of the best classifier using decision tree algorithms. Table 6 presents the performance of models when combining face and gait on features- and score-level fusion.

It is noted from the merging Table 6 that the results of excellence were better at both levels, The best merging results were when using the inception_v3 algorithm with SVC, as it reached 0.98 and that is for merging different measures on Feature level fusion and 0.97 for the various merging measures on Score level fusion. However, score-level fusion achieved less performance than feature-level fusion, fusion is better than facial or gait distinction performance separately.

n this study, the final comparative accuracy measure was adopted between facial recognition and gait separately, and the combination between them on two levels (features, score). As shown in Figure 11, merging the face and the gait using the inception_v3 algorithm to extract features with the SVM as a classifier achieved the best performance with an accuracy of 0.98. While fusion on score level using decision trees achieved 0.97. The results of separate face and gait recognition using the same algorithms for feature extraction and classification are 0.94 and 0.82, respectively.

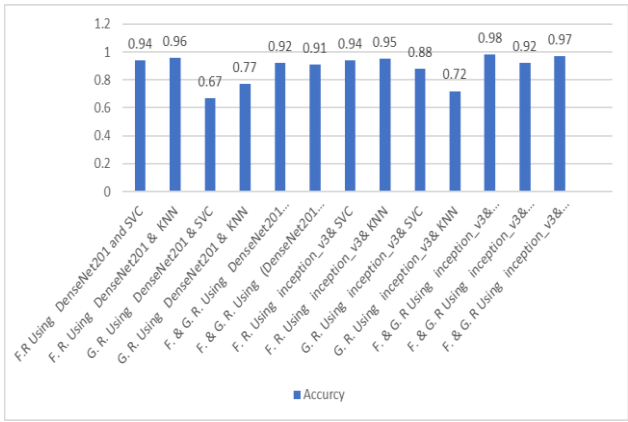

Figure 11. Comparison based on accuracy

Combining the face with the gait using the inception_v3 algorithm and KNN achieved an accuracy of 0.92, while face recognition using the same algorithm achieved an accuracy of 0.95 and for gait was 0.92. On the other hand, using DenseNet201 algorithms, the merger achieved 0.92 and 0.91 for SVM and KNN algorithms, respectively. Although merging the face with the gait shows higher results where recognition gait, but this result while the result of the merging was lower in the case of using the face only, where the DenseNet201 with SVM achieved 0.94, and the DenseNet201 with KNN 0.96 for face recognition. Table 7 provides a comparison of the findings obtained in this study with other research within the same domain.

Table 7. Comparisons between the proposal methods and others

Years	Authors'	Dataset	Fusion	Methods	Result
2018	Punyani et al. [25]	'OU-ISIR, 'USF Gait dataset'	Double-Level Fusion	KNN	MAE=6.1
2019	Rahman et al. [26]	KINECT Gait, KINECT Eurocom Face datasets	Score Level, Rank Level Fusion	Logistic regression	Accuracy =0.96
2021	Maity et al. [28]	Face and Ocular Challenge Series (FOCS) dataset	Score Level Multimodal Fusion	(Gabor), with(CNN)	Accuracy=.95.9
2022	Aung et al. [27]	public datasets	Feature-Level Fusion	Deep CNN with transfer learning	Accuracy=0.97
2023	Dindar M. Ahmed	Our dataset	Feature Fusion Level	inception_v3, with SVC	Accuracy =0.98
2023	Dindar M. Ahmed	Our dataset	Score Fusion Level	inception_v3, with SVC	Accuracy =0.97

5.2 Results discussion

The study concluded that the merging process improves recognition accuracy in general, as the best result was when merging the face with the gait using inception_v3 in extracting the features and svc as a classifier. Or in the case of using DenseNet201 algorithms in extracting features with both KNN and SVC classifiers, it was less in the case of using the same face classifier. At the same time, the accuracy of its use of gait was less, as it was 0.76 using SVC and 0.77 using KNN. The results show that the best algorithm for feature extraction is inception_v3, and the best classifier is SVC. And the process of integrating between the face and the gait improves the system in general in the various algorithms for both levels (Feature level fusion and Score level fusion).

6. CONCLUSIONS

In this study, a gait and face dataset has been created to create a multi-biometric recognition system to create an accurate recognition system using security system recorders. The data was pre-processed, and the faces and gait images were extracted. The study concluded that using more than one biometric feature increases recognition accuracy and reliability. Using transfer learning and feature fusion at the feature fusion level to integrate the dataset and then feed it into a machine learning model is a promising solution in biometric feature integration. Achieve the best model in feature extraction (inception_v3) with SVC when combining face and gait at the feature level, followed by the same algorithms for face and gait and using decision trees as a support classifier. In future work, more than one data set will be taken, and a more comprehensive range of deep learning and machine learning algorithms will be used.

ACKNOWLEDGMENT

Duhok Polytechnic University has supported this paper, and I would like to thank my supervisor for their guidance and support throughout the process. Their feedback and advice have been invaluable in shaping my research and refining my ideas.

REFERENCES

[1] Wang, M., Deng, W. (2021). Deep face recognition: A survey. Neurocomputing, 429: 215-244. https://doi.org/10.1016/j.neucom.2020.10.081

[2] Deng, J., Guo, J., An, X., Zhu, Z., Zafeiriou, S. (2021). Masked face recognition challenge: The insightface track report. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1437-1444.

[3] Zeng, D., Veldhuis, R., Spreeuwers, L. (2021). A survey of face recognition techniques under occlusion. IET Biometrics, 10(6): 581-606. https://doi.org/10.1049/bme2.12029

[4] Anda, F. (2021). Digital forensics: Leveraging deep learning techniques in facial images to assist cybercrime investigations. School of Computer Science, University College Dublin.

[5] Makihara, Y., Nixon, M.S., Yagi, Y. (2020). Gait recognition: Databases, representations, and applications. Computer Vision: A Reference Guide, 1-13. https://doi.org/10.1007/978-3-030-03243-2_883-1

[6] Zhang, C., Chen, X.P., Han, G.Q., Liu, X.J. (2023). Spatial transformer network on skeleton-based gait recognition. Expert Systems, e13244. https://doi.org/10.1111/exsy.13244

[7] Santos, C.F.G.D., Oliveira, D.D.S., Passos, L.A., Pires, R.G., Santos, D.F.S., Valem, L.P., Colombo, D. (2022). Gait recognition based on deep learning: A survey. arXiv preprint arXiv:2201.03323. https://arxiv.org/abs/2201.03323

[8] Kumar, M., Singh, N., Kumar, R., Goel, S., Kumar, K. (2021). Gait recognition based on vision systems: A systematic survey. Journal of Visual Communication and Image Representation, 75: 103052. https://doi.org/10.1016/j.jvcir.2021.103052

[9] Khan, M.H., Farid, M.S., Grzegorzek, M. (2021). Vision-based approaches towards person identification using gait. Computer Science Review, 42: 100432. https://doi.org/10.1016/j.cosrev.2021.100432

[10] Wang, Z., Yang, J., Zhu, Y. (2021). Review of ear biometrics. Arc hives of Computational Methods in Engineering, 28: 149-180. https://doi.org/10.1007/s11831-019-09376-2

[11] Chib, A.I., Lin, S., Li, C. (2022). Wearables for Health Promotion: An Interdisciplinary Review. https://dx.doi.org/10.2139/ssrn.4104254

[12] Singh, J.P., Jain, S., Arora, S., Singh, U.P. (2021). A survey of behavioral biometric gait recognition: Current success and future perspectives. Archives of Computational Methods in Engineering, 28: 107-148. https://doi.org/10.1007/s11831-019-09375-3

[13] Jafari, M., Schumacher, A.M., Snaidero, N., Ullrich Gavilanes, E.M., Neziraj, T., Kocsis-Jutka, V., Kerschensteiner, M. (2021). Phagocyte-mediated synapse removal in cortical neuroinflammation is promoted by local calcium accumulation. Nature Neuroscience, 24(3): 355-367. https://doi.org/10.1038/s41593-020-00780-7

[14] Sprager, S., Juric, M.B. (2015). Inertial sensor-based gait recognition: A review. Sensors, 15(9): 22089-22127. https://doi.org/10.3390/s150922089

[15] Catruna, A., Cosma, A., Radoi, I.E. (2021). From face to gait: Weakly-supervised learning of gender information from walking patterns. In 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), pp. 1-5. https://doi.org/10.1109/FG52635.2021.9666987

[16] Annbuselvi, K., Santhi, N., Sivakumar, S. (2022). A competent multimodal recognition using imperfect region based face and gait cues using Median-LBPF and Median-LBPG based PCA followed by LDA. Materials Today: Proceedings, 62: 4869-4879. https://doi.org/10.1016/j.matpr.2022.03.505

[17] Jaddoh, A., Loizides, F., Rana, O. (2023). Interaction between people with dysarthria and speech recognition systems: A review. Assistive Technology, 35(4): 330-338. https://doi.org/10.1080/10400435.2022.2061085

[18] Zakaria, E., Rahman, W.A., Twakol, A., Shawky, A. (2019). Face recognition using deep neural network technique. In SL International Conference Giza, pp. 1-32.

[19] Tabassum, F., Islam, M.I., Khan, R.T., Amin, M.R. (2022). Human face recognition with combination of DWT and machine learning. Journal of King Saud University-Computer and Information Sciences, 34(3):

546-556. https://doi.org/10.1016/j.jksuci.2020.02.002
[20] Mehmood, A., Khan, M.A., Sharif, M., Khan, S.A., Shaheen, M., Saba, T., Ashraf, I. (2020). Prosperous human gait recognition: An end-to-end system based on pre-trained CNN features selection. Multimedia Tools and Applications, 1-21. https://doi.org/10.1007/s11042-020-08928-0
[21] Mogan, J.N., Lee, C.P., Anbananthen, K.S.M., Lim, K.M. (2022). Gait-DenseNet: A hybrid convolutional neural network for gait recognition. IAENG International Journal of Computer Science, 49(2): 393-400.
[22] Liu, X., Liu, J. (2020). Gait recognition method of underground coal mine personnel based on densely connected convolution network and stacked convolutional autoencoder. Entropy, 22(6): 695. https://doi.org/10.3390/e22060695
[23] Wang, X., Yan, W.Q. (2020). Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory. International Journal of Neural Systems, 30(1): 1950027. https://doi.org/10.1142/S0129065719500278
[24] Aung, H.M.L., Pluempitiwiriyawej, C. (2020). Gait biometric-based human recognition system using deep convolutional neural network in surveillance system. In 2020 Asia Conference on Computers and Communications (ACCC), pp. 47-51. https://doi.org/10.1109/ACCC51160.2020.9347899
[25] Punyani, P., Gupta, R., Kumar, A. (2018). Human age-estimation system based on double-level feature fusion of face and gait images. International Journal of Image and Data Fusion, 9(3): 222-236. https://doi.org/10.1080/19479832.2018.1423644
[26] Rahman, M.W., Zohra, G.F., Gavrilova, M.L. (2019). Score level and rank level fusion for kinect-based multi-modal biometric system. Journal of Artificial Intelligence and Soft Computing Research, 9(3): 167-176.
[27] Aung, H.M.L., Pluempitiwiriyawej, C., Hamamoto, K., Wangsiripitak, S. (2022). Multimodal biometrics recognition using a deep convolutional neural network with transfer learning in surveillance videos. Computation, 10(7): 127. https://doi.org/10.3390/computation10070127
[28] Maity, S., Abdel-Mottaleb, M., Asfour, S.S. (2021). Multimodal low resolution face and frontal gait recognition from surveillance video. Electronics, 10(9): 1013. https://doi.org/10.3390/electronics10091013
[29] Manssor, S.A., Sun, S., Elhassan, M.A. (2021). Real-time human recognition at night via integrated face and gait recognition technologies. Sensors, 21(13): 4323. https://doi.org/10.3390/s21134323
[30] Taunk, P., Jayasri, G., Priya, J.P., Kumar, N.S. (2020). Face detection using Viola Jones with Haar cascade. Test Engineering and Management, 83: 19146.
[31] Al-Zubaidy, H.A.K. (2020). Real-time detection to solve traffic congestion problems. Master's thesis, Altınbaş Üniversitesi/Lisansüstü Eğitim Enstitüsü.
[32] Medic, T., Ruttner, P., Holst, C., Wieser, A. (2023). Keypoint-based deformation monitoring using a terrestrial laser scanner from a single station: Case study of a bridge pier. http://ocs.editorial.upv.es/index.php/JISDM/JISDM2022/paper/view/13812.
[33] Juneja, S., Jain, S., Suneja, A., Kaur, G., Alharbi, Y., Alferaidi, A., Dhiman, G. (2021). Gender and age classification enabled blockschain security mechanism for assisting mobile application. IETE Journal of Research, 1-13. https://doi.org/10.1080/03772063.2021.1982418
[34] Ran, Q., Liu, Z., Sun, X., Sun, X., Zhang, B., Guo, Q., Wang, J. (2021). Anomaly detection for hyperspectral images based on improved low-rank and sparse representation and joint Gaussian mixture distribution. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 6339-6352. https://doi.org/10.1109/JSTARS.2021.3087588
[35] Hussein, I.J., Burhanuddin, M.A., Mohammed, M.A., Benameur, N., Maashi, M.S., Maashi, M.S. (2022). Fully-automatic identification of gynaecological abnormality using a new adaptive frequency filter and histogram of oriented gradients (HOG). Expert Systems, 39(3): e12789. https://doi.org/10.1111/exsy.12789
[36] Chen, T., Gao, T., Li, S., Zhang, X., Cao, J., Yao, D., Li, Y. (2021). A novel face recognition method based on fusion of LBP and HOG. IET Image Processing, 15(14): 3559-3572. https://doi.org/10.1049/ipr2.12192
[37] Zhang, Z., Zou, C., Han, P., Lu, X. (2020). A runway detection method based on classification using optimized polarimetric features and HOG features for PolSAR images. IEEE Access, 8: 49160-49168. https://doi.org/10.1109/ACCESS.2020.2979737
[38] Yuan, Y., Chen, L., Wu, H., Li, L. (2022). Advanced agricultural disease image recognition technologies: A review. Information Processing in Agriculture, 9(1): 48-59. https://doi.org/10.1016/j.inpa.2021.01.003
[39] Pan, J., Cui, T., Le, T. D., Li, X., Zhang, J. (2020). Multi-group transfer learning on multiple latent spaces for text classification. IEEE Access, 8: 64120-64130. https://doi.org/10.1109/ACCESS.2020.2984571
[40] Lin, X., Qin, F., Peng, Y., Shao, Y. (2021). Fine-grained pornographic image recognition with multiple feature fusion transfer learning. International Journal of Machine Learning and Cybernetics, 12: 73-86. https://doi.org/10.1007/s13042-020-01157-9
[41] Zheng, X., Chen, J., Wang, H., Zheng, S., Kong, Y. (2021). A deep learning-based approach for the automated surface inspection of copper clad laminate images. Applied Intelligence, 51: 1262-1279. https://doi.org/10.1007/s10489-020-01877-z
[42] Song, R., Li, T., Wang, Y. (2020). Mammographic classification based on XGBoost and DCNN with multi features. IEEE Access, 8: 75011-75021. https://doi.org/10.1109/ACCESS.2020.2986546
[43] Ahn, J.M., Kim, S., Ahn, K.S., Cho, S.H., Lee, K.B., Kim, U.S. (2018). A deep learning model for the detection of both advanced and early glaucoma using fundus photography. PloS One, 13(11): e0207982. https://doi.org/10.1371/journal.pone.0207982
[44] Moon, W.K., Lee, Y.W., Ke, H.H., Lee, S.H., Huang, C.S., Chang, R.F. (2020). Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Computer Methods and Programs in Biomedicine, 190: 105361. https://doi.org/10.1016/j.cmpb.2020.105361
[45] Li, Y. (2022). Research and application of deep learning in image recognition. In 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), pp. 994-999.

https://doi.org/10.1109/ICPECA53709.2022.9718847
[46] Sanghvi, H.A., Patel, R.H., Agarwal, A., Gupta, S., Sawhney, V., Pandya, A.S. (2023). A deep learning approach for classification of COVID and pneumonia using DenseNet-201. International Journal of Imaging Systems and Technology, 33(1): 18-38. https://doi.org/10.1002/ima.22812
[47] Aamir, M., Zaidi, S.M.A. (2021). Clustering based semi-supervised machine learning for DDoS attack classification. Journal of King Saud University-Computer and Information Sciences, 33(4): 436-446. https://doi.org/10.1016/j.jksuci.2019.02.003
[48] Mahfouz, M.A., Shoukry, A., Ismail, M.A. (2021). EKNN: Ensemble classifier incorporating connectivity and density into kNN with application to cancer diagnosis. Artificial Intelligence in Medicine, 111: 101985. https://doi.org/10.1016/j.artmed.2020.101985
[49] Zhang, X.D. (2020). Support Vector Machines. In A Matrix Algebra Approach to Artificial Intelligence; Springer: Singapore, pp. 617-679.
[50] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408: 189-215. https://doi.org/10.1016/j.neucom.2019.10.118
[51] Charbuty, B., Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01): 20-28. https://doi.org/10.38094/jastt20165
[52] Luo, X., Wen, X., Zhou, M., Abusorrah, A., Huang, L. (2021). Decision-tree-initialized dendritic neuron model for fast and accurate data classification. IEEE Transactions on Neural Networks and Learning Systems, 33(9): 4173-4183. https://doi.org/10.1109/TNNLS.2021.3055991