



Human gait analysis for osteoarthritis prediction: a framework of deep learning and kernel extreme learning machine

Muhammad Attique Khan¹ · Seifedine Kadry² · Pritee Parwekar³ · Robertas Damaševičius⁴ · Asif Mehmood¹ · Junaid Ali Khan¹ · Syed Rameez Naqvi⁵

Received: 23 August 2020 / Accepted: 25 November 2020 / Published online: 19 January 2021
© The Author(s) 2021

Abstract

Human gait analysis is a novel topic in the field of computer vision with many famous applications like prediction of osteoarthritis and patient surveillance. In this application, the abnormal behavior like problems in walking style is detected of suspected patients. The suspected behavior means assessments in terms of knee joints and any other symptoms that directly affected patients' walking style. Human gait analysis carries substantial importance in the medical domain, but the variability in patients' clothes, viewing angle, and carrying conditions, may severely affect the performance of a system. Several deep learning techniques, specifically focusing on efficient feature selection, have been recently proposed for this purpose, unfortunately, their accuracy is rather constrained. To address this disparity, we propose an aggregation of robust deep learning features in Kernel Extreme Learning Machine. The proposed framework consists of a series of steps. First, two pre-trained Convolutional Neural Network models are retrained on public gait datasets using transfer learning, and features are extracted from the fully connected layers. Second, the most discriminant features are selected using a novel probabilistic approach named Euclidean Norm and Geometric Mean Maximization along with Conditional Entropy. Third, the aggregation of the robust features is performed using Canonical Correlation Analysis, and the aggregated features are subjected to various classifiers for final recognition. The evaluation of the proposed scheme is performed on a publicly available gait image dataset CASIA B. We demonstrate that the proposed feature aggregation methodology, once used with the Kernel Extreme Learning Machine, achieves accuracy beyond 96%, and outperforms the existing works and several other widely adopted classifiers.

Keywords Gait recognition · Osteoarthritis · Video surveillance · Public security · Feature fusion · Transfer learning · Deep learning · Extreme learning machine

Introduction

Nowadays, medical and biomedical image processing has attracted a lot of attention due to its importance in health care. [1, 2]. The primary goal of medical and biomedical image processing is to detect abnormalities in organs and body of patients [3, 4]. It allows us to detect many dangerous diseases such as cancer. The abnormalities can be controlled through the patient monitoring system. In the area of computer vision, human gait analysis (HGA) is a new research area for patients monitoring. In this approach, the patients are detected based on their walking styles such as their knee joint issue or any other symptoms which are affected by the patient's walking style. This change in patient walking style is described as gait analysis [5]. Osteoarthritis (OA) is the most common joint disorder in the elderly. In the United States alone, symptomatic knee arthritis occurs in 10% of

✉ Seifedine Kadry
skadry@gmail.com

¹ Department of Computer Science, HITEC University Taxila, Taxila, Pakistan

² Faculty of Applied Computing and Technology, Noroff University College, Kristiansand, Norway

³ Department of Computer Science and Engineering, SRM Institute of Science and Technology, Modinagar, Ghaziabad, UP 201204, India

⁴ Department of Applied Informatics, Vytautas Magnus University, 44404 Kaunas, Lithuania

⁵ Department of Computer and Electrical Engineering, COMSATS University Islamabad, Wah Campus, Islamabad, Pakistan

men and 13% of women aged sixty [6]. OA often causes persistent pain and poor quality of life for the elderly. Furthermore, it also makes it more difficult to walk for them. Today, with advances in technology, HGA is an effective method to predict OA [7]. Human gait analysis for classifying OA on the elderly plays an important role in applications of computer vision for medicine. In this article, we focus on the problem of HGA for patients based on their walking style. Moreover, it has also been used for a wide variety of tasks such as patients monitoring under the sensitive condition and any injury. The identification of suspect behavior of patients in Closed Circuit Television (CCTV) footage is important to ensure public safety in both indoor and outdoor locations [8]. For example, in the case of a pandemic such as COVID, it is possible to track people through their gait and if any unusual circumstances are found, one can act promptly [9]. A similar search may be performed over multiple records from other locations, to reconstruct the travel history of the suspected patients, or to match it with the gait patterns of known individuals stored in the hospital databases [10]. The gait patterns or some external characteristics of subjects (such as specific clothes or carried objects) can be used for real-time monitoring of crowds to identify their moving history [11].

Although, many methods exploit unique attributes of a person, such as facial, ear and iris, and Electroencephalography (EEG) for biometric recognition [12], the gait analysis enjoys an advantage in that it does not require the subject's cooperation to assist in the recognition process. Analyzing someone's unique walking patterns, also, allows identifying them at larger distances [13]. The gait analysis has become an active research area for medical and assisted living applications [14], but also user identity verification biometric applications because of its robustness and usefulness in many domains such as clinical analysis, airports, forensic, bus stations, and bank surveillance systems [15, 16]. Tracking and identification of subjects between different un-calibrated non-overlapping stationary CCTV cameras based on gait analysis have been shown in [17].

The gait extraction is usually rather easy, making the recognition process quite convenient [18]. However, there are many factors, such as different clothing [19], variation in view angle [20], carrying conditions [21], and poor lighting, which degrade the performance of the analysis system. Several Machine Learning (ML) and Computer Vision (CV) techniques are available for HGA; they are primarily classified into two wide categories: model-based [22] and model-free [23] approaches. In the former, a model based on the structure of the human body is used for recognition. The parameters of such a model are used as attributes like the angle of joints. These techniques work well for factors such as variation in view, clothing, carrying luggage, and shadow effects that degrade the performance of recognition [24]. Although, it is advantageous to have a high-level model, it

bears a high computational cost. The model-free approach, in contrast, works on the silhouette of the human body. This approach usually proves more cost-effective. This approach is more sensitive towards different covariants such as shadows, carrying conditions, and different clothing [25]. Therefore, we need to find and justify a tradeoff between the two components in the time-and-accuracy argument—making the HGR systems still an active area of research.

In the literature, various techniques are available for HGA to overcome the problem of different covariants, as listed above. Generally, a simple HGR method involves several steps, including preprocessing of image frames through different approaches [19], applying different methods of segmentation on the silhouette of the image frames [26], extraction of gait attributes and recognition of the gait [27]. Since an image may include several problems, such as low resolution, noise, and complex background, the preprocessing step is supposed to rectify these issues, and enhance the quality of the image for the next step—feature extraction [28, 29]. Since, the irrelevant features may drastically degrade the performance of the system, the main concern, in the features extraction step is to extract the most relevant and robust features for reasonably accurate recognition. Unfortunately, the larger the dimensionality of the features, the smaller the system's accuracy and the higher the computational cost will be [30]. To address this disparity, several features reduction methods have been reported in the literature. Some of the famous reduction and selection techniques are entropy-based [31], correlation-based [32], Wavelet Transform [33], Genetic Algorithm-based [34], nature-inspired optimization-based [35], and a few more [36].

The aggregation of features is another important step that increases the information of an object in the image. The main purpose of this step is to improve the classification accuracy of the system. But on the other side, this step decreases the system performance due to high dimensional features [37]. The aggregated features are finally embedded in a classification engine for the classifying selected classification problem.

Major contributions

Here, we propose a framework for human gait analysis that exploits the aggregation of robust deep learning features in Kernel Extreme Learning Machine (KELM). Two different angles of CASIA B dataset are used for the validation of the proposed scheme. In this dataset, three different situations are considered: wearing a coat, carrying a bag, and a normal walk. A few sample frames of both angles are shown in Figs. 1 and 2.

The principal contributions of this work are enlisted below.

Fig. 1 Sample frames from CASIA B dataset (90° angle) [38]

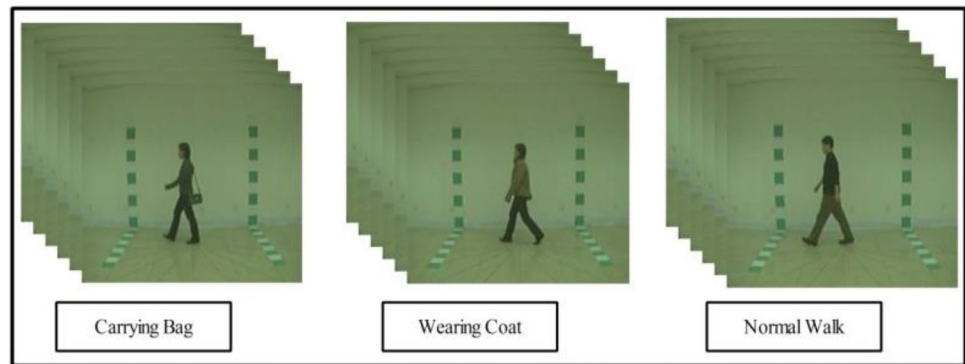
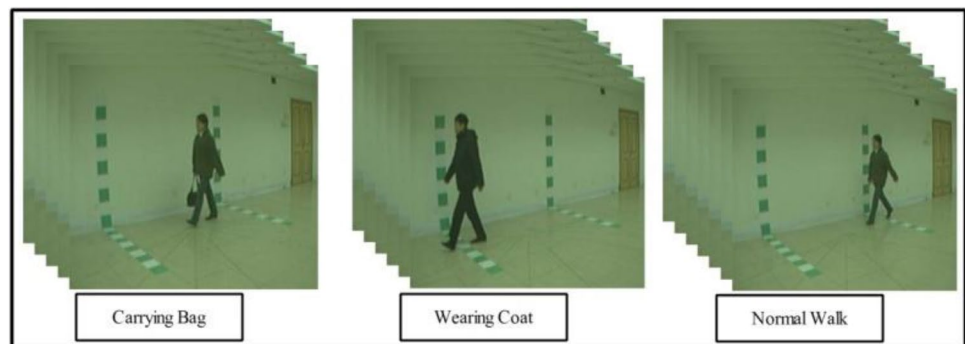


Fig. 2 Sample frames from CASIA B dataset (54° angle) [38]



- (i) We modify the VGG16 Net and AlexNet Convolutional Neural Networks (CNN) models for gait recognition and trained on the CASIA B database (54° and 90°) using Transfer Learning (TL). Subsequently, we extract the deep learning features from the Fully Connected (FC) layers instead of middle layers.
- (ii) A novel Euclidean Norm and Geometric Mean Maximization along with Conditional Entropy (ENGM-wCE) approach is proposed for the selection of maximum score features. A Fine-KNN classifier is used as a fitness function for the selection of robust features.
- (iii) We perform aggregation of the selected deep learning features using the Canonical Correlation Analysis based approach and embed the aggregated vector in KELM for final recognition.

Related work

Deep learning is a hot research area of machine learning and is employed in several applications such as biometrics, visual surveillance, medical, and image classification. Gait recognition is an important biometric process, and several techniques in this regard have been developed and presented in the literature. These existing techniques are specific and have been developed to overcome various gait recognition challenges such as clothing, carrying conditions, shadow, and view angles [24]. Castro et al. [39] presented a new

technique for gait recognition in video sequences using a CNN based approach. For the learning of high-level gait features, activation was performed on the fully connected (FC) layer of the CNN. Next, the spatio-temporal cuboids were fed in the CNN for final recognition. The TUM GAID dataset was considered for the evaluation of the presented techniques, which managed to achieve recognition accuracy of 88.9%. Habiba et al. [40] presented an optical flow-based framework for gait recognition along with Bayesian Model and Normal Distribution. The motion vectors were calculated using optical flow and then quartile deviation was used to segment the human region. Later, the texture information was extracted from the segmented regions, and the important features were selected using Bayesian Modeling. The presented method was validated on CASIA B dataset and achieved an accuracy of 87.7%. Li et al. [41] presented an HGR approach named DeepGait to overcome the problem of covariant factors. A Joint Bayesian (JB) model was used to deal with the problem of variation in a viewpoint. Firstly, the gait cycle was estimated using the Normalized Auto Correlation (NAC) to represent the deep convolution of gait. The VGG16 pre-trained architecture was used for the learning process. Next, the JB model was used for gait identification. The OULP gait dataset was used for the experimental process, and an accuracy of 89.3% was achieved. Mehmood et al. [42] presented a novel approach to overcome the problems associated with clothing variation and walking style. A four-step method was developed. In the first step, the

preprocessing was performed, which was followed by the features extraction step using the pre-trained CNN model named DenseNet201. Next followed the dimensionality reduction using skewness and firefly algorithm. In this step, the authors tried to select only the relevant features, which were embedded in “One-vs-All SVM” for final recognition. The evaluation of this technique was conducted on CASIA B dataset and attained an accuracy of 94.3%, 93.8%, and 94.7% for 180°, 360°, and 540° angles, respectively. Arshad et al. [43] presented a deep learning framework for HGR, in which they tried to resolve the problems brought in by clothing and view. The feature extraction was performed using two pre-trained deep learning models named AlexNet and VGG19. Secondly, entropy and skewness were calculated to construct a fused feature vector. Next, a novel concept called Fuzzy Entropy Controlled Skewness was proposed for selection of the best features. The presented framework was evaluated on four HGR databases CASIA A, CASIA B, CASIA C, and AVAMVG gait, and accuracies of 99.7%, 93.3%, 92.2%, and 99.8% were achieved respectively. Alotaibi et al. [44] also presented a CNN based HGR system. This work claimed to resolve the problems of common degradation and small data handling. The presented CNN model was based on four max pool and four fully connected layers. For evaluation, the CASIA B dataset was used, and accuracies of 98.3%, 83.87%, and 89.12% were achieved respectively for the walking normally (nm), wearing a coat (cl), and carrying a bag (bg) cases. Zhang et al. [45] presented an HGR system based on an encoder architecture to overcome the problems of variations such as clothing, view, and carrying things. The CNN and LSTM networks were used for feature extraction. Later, information on both systems was combined and performed the final recognition. Three HGR datasets named CASIA B, USF, and FVG were used for the evaluation, and accuracies of 81.8%, 99.5%, and 87.8% respectively were achieved. Yu et al. [46] introduced a novel technique to conquer the problems of different variations. Features based on CNN are extracted in this method. An auto encoder based on stack progressive method is used to address the problem of variation. PCA is utilized for the selection of best features and leaving the irrelevant features. Finally, KNN algorithm is used for the recognition. The system is assessed using SZU RGB D and CASIA-B datasets and achieved improved performance.

Marcin et al. [47] introduced a novel method to analyze the walking style of individual wearing different types on shoes. Different 81 individual and 2700 walking periods were used for the analysis and it is assessed that the style of walking changes according to the types of shoes. The system was evaluated based on the database of 81 individual and attained the accuracy of 99%. Khan et al. [48] presented a HGR system in which sequence of video is used for extraction of features. Codebook generation is done in this method

and after that; vector is encoded using encoding based on fisher vector. Linear SVM is used for final gait recognition. The presented HGR method is assessed using CASIA-A and TUM GAID databases. In case of CASIA-A the attained accuracy was 100% and in case of TUM GAID the recognition rate was 97.74%. Few other studies also used CASIA-B dataset and showed significant performance [49, 50]. In summary, the above listed techniques are tries to address the problem of HGA under different variations. The main challenge which they were faced is walking style like speed etc. To resolve these issues, few researchers focused on the region of interest detection for features extraction and rest of them passed raw video frames directly for feature extraction.

Proposed methodology

Here, a novel method is proposed for gait recognition using aggregation of deep learning features. The proposed design, as shown in Fig. 3, consists of a series of steps. First, two pre-trained CNN models (AlexNet and VGG16) are retrained on gait datasets using a transfer learning approach, and the features are extracted from the second last Fully Connected Layer (FC7) in each. Second, we select the most discriminant features using the proposed probabilistic approach ENGMwCE. Third, the aggregation of these features is performed using Canonical Correlation Analysis (CCA) and the features are subjected to KELM for final recognition.

Dataset collection

Two datasets are used in this work named CASIA-B [38] and CASIA-A [51]. The CASIA-B is a dataset comprising sample images from an indoor environment. It comprises images with 124 actors, including 93 male and 31 female, and consists of 11 view angles such as 0°–360°. Three variations are considered in this dataset like a normal walk (nm), walk with a bag (bg), and walk with a coat (cl). Each subject records 10 videos–6 videos of a normal walk, 3 videos for a walk with a coat, and 2 videos for a walk with a bag. A subject has three statuses of walking: normal walk (nm), wearing a coat (cl), and carrying a bag (bg). Each video is taken at 25 frames per second (fps), with each having a resolution of 352 × 240 pixels. In this work, we have selected 54° and 90° camera view angles for the evaluation of our work. A few sample frames are shown in Figs. 1 and 2.

The CASIA-A dataset consists of a total of 240 video sequences. Twenty subjects are involved in this dataset, and each subject records 12 videos in three different directions like parallel, 45°, and 90°, respectively. The length of each video in this dataset is based on the subject walking speed.

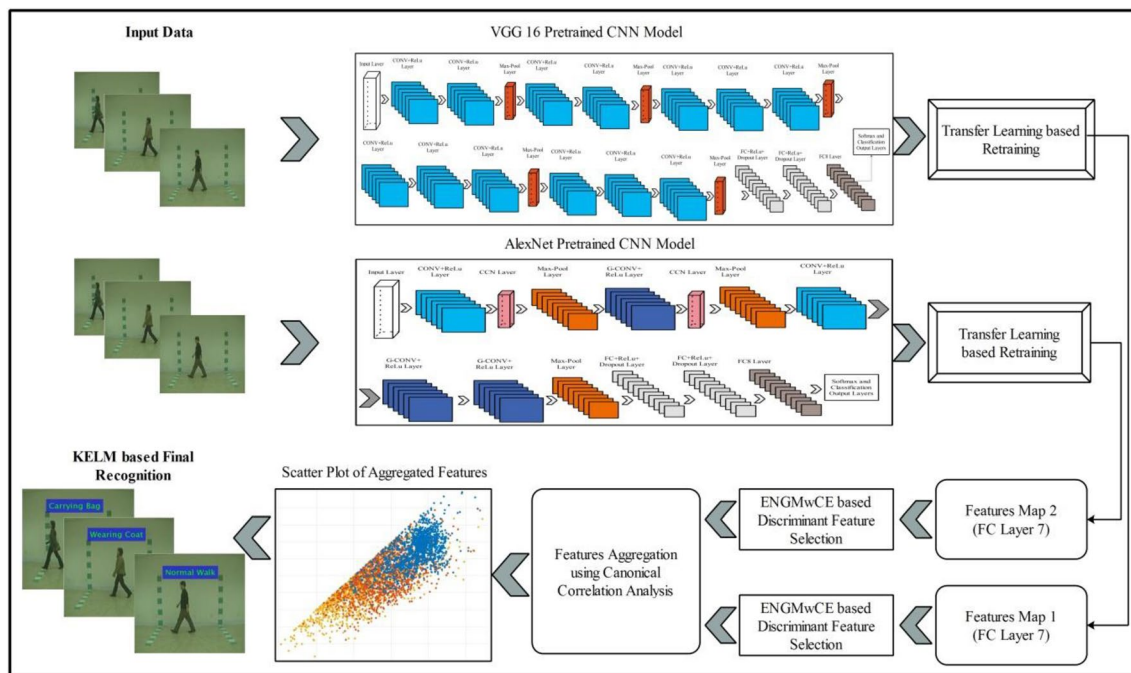


Fig. 3 The proposed architecture of deep aggregated features in KELM for Human Gait Recognition

Convolutional neural network (CNN)

Deep learning shows a huge interest in computer vision research due to improved classification performance [52]. CNN is a deep learning architecture, which consists of several layers (input, hidden, and output). It was functional in many industrial applications like visual surveillance and biometrics [53–55]. CNN builds the graphical view of the mechanism of the individual, which can perform supervised learning as well as unsupervised learning. The kernel parameters of a convolutional layer are connected with hidden layers that later enables the CNN into a smaller weight for learning. Features are extracted automatically from these layers without using any preprocessing step. A simple CNN design consists of various layers and the first layer is the input layer. After the input layer, the convolutional layer is added to perform the convolutional operation. The convolutional operation is based on the convolutional kernel defined by $k_1 \times a$ and stride $s_1 \times s_2$. Mathematically, the convolution operation is defined by Eq. (1):

$$\psi_l^n = \phi(W_l^i \otimes h_{(x,y)} + \beta_l^i), \tag{1}$$

where, ψ_l^n is the size of the convolutional layer of dimension $k \times M - k_1 - N - a$ and \otimes denotes convolutional operation, $\phi(\cdot)$ denotes activation function, $h_{(x,y)}$ represent input, W_l^i represent weights matrix, and β^i denotes the bias matrix, respectively. The weights and bias are updated after the

addition of another convolutional layer. Mathematically, the update in weights and bias are defined by Eqs. (2) and (3):

$$W_l^{i+1} = \frac{-r}{q} W_l^i - \frac{r}{n} \left(\frac{\partial F}{\partial W_l^i} \right) + m W_l^i, \tag{2}$$

$$\beta_l^{i+1} = \frac{r}{n} \left(\frac{\partial F}{\partial \beta_l^i} \right) + \beta_l^i, \tag{3}$$

where, W_l^{i+1} represent updated weight matrix, β_l^{i+1} represent updated bias matrix, r represent learning rate, and F represents fitness function. In this layer, initially defined a filter size of dimension $n \times k$ and normally its value is 3. The kernel size is 3×3 and number of channels are 32. The other parameters which are involves in this layer are learning factor and learning rate.

Next, ReLu layer is added to improve the problem of sigmoid partial gradient fit and gradient loss. This layer is mostly followed by the convolutional layer. In CNN architecture, the problem of overfitting is resolved through max-pooling layer. Using this layer, reduce the length of the weight matrix. Mathematically, it is formulated by Eqs. (4) and (5):

$$M_l^i = \text{Max}(W_l^i), \tag{4}$$

$$M_l^{i+1} = \text{Max}(W_l^{i+1}), \tag{5}$$

where the output matrix size after max-pooling operation is $k \times \frac{M-k_1}{c} \times \frac{(N-a)}{d}$. Another important layer name fully connected layer is including in a CNN used to extract the high-level features of an image. The formulation of FC layer is defined by Eq. (6):

$$Y_l^i = \sum_i W_l^i \cdot \psi_l^n + \beta_l^i, \tag{6}$$

where, Y_l^i denotes the output of FC layer and \tilde{l} denotes the layer before FC layer. In this layer, features are extracted. A Softmax layer is added after FC layer. This layer is known as the classification layer. The cross-entropy function is used in this layer to calculate the loss of classification output. A mostly sigmoidal function is used to train a CNN model.

Deep learning features extraction

Transfer learning (TL) In machine learning, TL is a concept of knowledge sharing from one domain to another domain within minimum time and less energy [56]. The main purpose of TL is train an existing CNN model on selected datasets with same parameters. Consider a domain D and a task w can be well-defined by a space of label X and a prediction function $f(\cdot)$ that is learned from the vector of attribute and pairs of lable $\{s_i, x_i\}$ where $s_i \in S$ and $x_i \in X$. By considering the defect classification application of software module, X is a label set that contains true and false in this case. The value of x_i is either true or false and the learner is considered as $f(x)$ that is used for prediction of module of software s . From the above definition, a domain $D = \{S', P(S)\}$ and a task $w = \{X, f(\cdot)\}$. The D_a can be defined as data of source domain where $D_a = \{(s_{a1}, x_{a1}) (s_{an}, x_{an})\}$ where $s_{ai} \in S'_a$ is

the i -th instance of data of D_a and label of the corresponding class s_{ai} is $x_{ai} \in X_a$. By considering this D_b can be as data of target domain where $D_b = \{(s_{b1}, x_{b1}) (s_{bn}, x_{bn})\}$ where $s_{bi} \in S'_w$ is i th instance of data D_b and class label for corresponding s_{bi} is $x_{bi} \in X_w$. Furthermore, the task of source can be demonstrated as w_a and task of target as w_b . The prediction function of source can be defined as $f_a(\cdot)$ and for the target as $f_b(\cdot)$. Visually, it is also presenting from Fig. 4.

Training data explanation For CASIA B dataset, we utilize the first 74 subjects for training the model and the remaining 50 subjects for testing the proposed scheme. This division means that we employ a 60:40 approach along with cross-validation value of 10. For CAISA A dataset, we utilize the first 12 subjects for training the model and the remaining 8 subjects for evaluation of the proposed system, where the cross-validation is 10.

Feature vector l In this work, we use two pre-trained CNN structures named AlexNet [57] and VGG 16 Net [58] for feature extraction. The visual structure of AlexNet is shown in Fig. 6. In this CNN structure, two convolutional layers (CONV), three grand convolutional layers (G-CONV), seven ReLu layers, two normalization layers, 3 max-pooling, 3 FC layers, one Softmax, two dropout layers, and one classification layer are used. A sigmoid function is used to train this model. For training the model, the original RGB frames are passed to the network that is later resized according to the size of the first layer named the input layer. In AlexNet, the input layer size is $227 \times 227 \times 3$. After retraining this model using TL, we extract features from FC Layer 7 which is $l - 1$ layer of FC8 (from Fig. 5). The nature of FC layer is in the form of numeric features and features vector must be

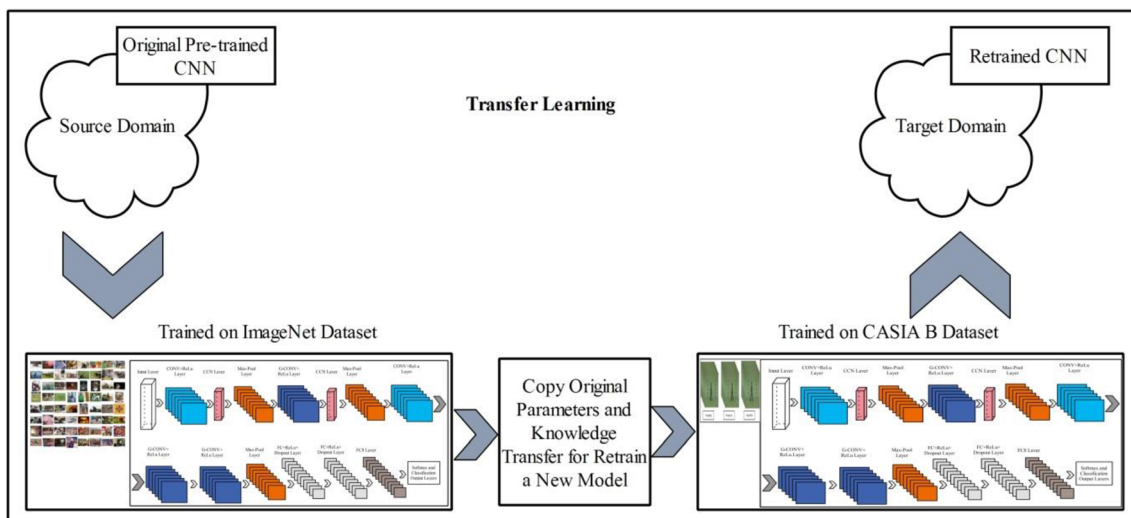


Fig. 4 Visual concept of Transfer Learning for knowledge sharing to new model

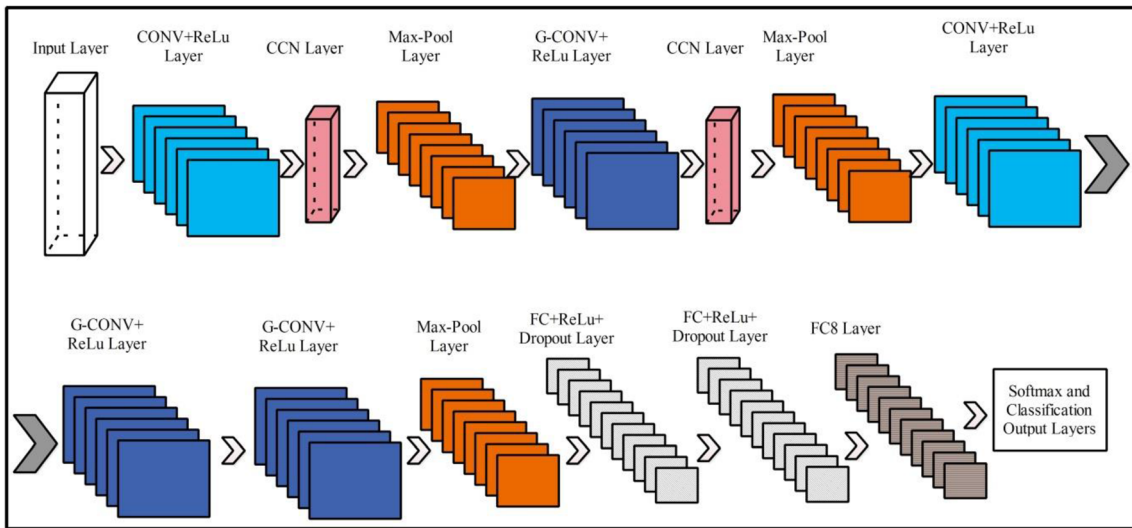


Fig. 5 A general architecture of AlexNet pre-trained CNN

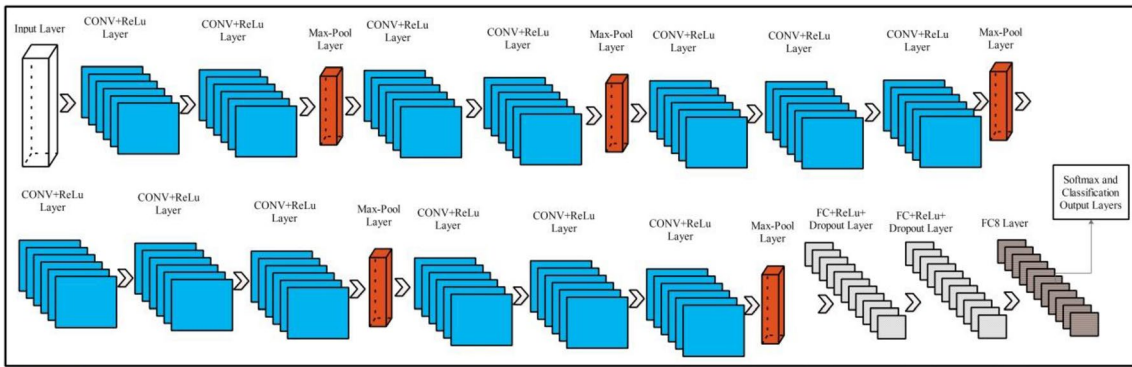


Fig. 6 A general architecture of VGG16 pre-trained CNN

1D. The length of resultant 1D vector is $N \times 4096$, where N represent number of frames used for training and testing. Mathematically, this vector is denoted by ξ_{V1} .

Feature Vector 2 For extraction of the second feature vector, we use a pre-trained VGG16 Net CNN structure. Like AlexNet, this structure is also originally trained on ImageNet, where the training function is sigmoid. This CNN structure consists of thirteen convolutional layers, thirteen ReLu activation, five max pooling, two dropouts, three FC layers, and one Softmax layer. For training the model on the selected datasets, the original RGB frames are passed to the network that is later resized according to the size of the first layer named the input layer. In VGG16, the input layer size is $224 \times 224 \times 3$, so all frames are resized according to the input layer size. Visually, the VGG16 Net architecture is shown in Fig. 6. The FC layer seven is considered in this work for extraction of the deep learning features. The sizes

of convolutional filters are fixed in this structure. The resultant feature vector dimension is $N \times 4096$, denoted by ξ_{V2} .

Discriminant feature selection

The performance of a classification system depends on the number of input features. From previous studies, it is shown that the removal of redundant information increases the recognition accuracy and minimizes the execution time. The feature selection methods select the best subset of features from the original vector instead of generating new features.

We have two feature vectors denoted by ξ_{V1} and ξ_{V2} of $N \times 4096$. Suppose that ξ is a vector of a subset of features of ξ_{V1} and ξ_1 is a vector of a subset of features ξ_{V2} . First, we consider vector $\xi = \{\xi_1, \dots, \xi_M\}$, $M \leq 4096$, is a vector of input features and $Y_i = \{y_i, \dots, y_{i_q}\}$ represents corresponding labels for each feature ξ_i , $i = 1, \dots, M$, extracted from an

image, respectively. For the selection of the most discriminant features, we propose a new technique named Euclidean Norm and Geometric Mean Maximization along with Conditional Entropy (ENGMwCE). Using this technique, we initially select the most discriminant features based on the maximization property of Euclidean norm (EN) and Geometric mean norm (GMN). Then, we combine the information of both techniques using a serial approach. Later, conditional entropy is applied to refine the negative features and passed them to a threshold function. The features that meet the condition of the threshold function are examined through fitness function (FKNN). Based on FKNN error, the condition is terminated where the target error is 0.08. If FKNN error is below the target error, then the selection process is terminated and obtained a selected feature vector. Similarly, this process is performed for feature vector ξ_{V2} . Mathematically, the selection process is defined as follows:

First, the EN is calculated from vector ξ and selects only those features whose are greater L^2 norm. The formulation is defined by Eq. (7):

$$ENM = \delta = \delta_k \cup \operatorname{argmax}_{\xi_i \in \xi - \delta} \left[\left| \Psi^3(\xi_i, Y) \right| \right], \tag{7}$$

where δ_k denotes feature subset, Ψ denotes mutual information function that is utilized to compute the mutual information among features, and it is defined as:

$$\Psi(\xi_i, Y) = \sum_{a \in \xi_i} \sum_{b \in Y} p(a, b) \log_2 \frac{p(a, b)}{p(a)p(b)}, \tag{8}$$

$p(a, b)$ is probability in that a and b occur together, and Y denotes label set, respectively. The formulation of $\Psi^3(\xi_i, Y)$ is given by Eqs. (9) and (10):

$$\Psi^3(\xi_i, Y) = \{ \Psi(\xi_i, y) \mid y \in Y \}, \tag{9}$$

$$\left| \Psi^3(\xi_i, Y) \right| = \sqrt[q]{\sum_{j=1}^q \Psi(\xi_i, y_j)}. \tag{10}$$

Next, we implement a GM maximization approach on ξ . The GM maximization selects the largest GM values. The main difference among features through GM is a scaling factor that is mathematically defined by Eq. (11):

$$G = \delta_1 = \delta_{1k} \cup \operatorname{argmax}_{\xi_i \in \xi - \delta_1} [G(\Psi(\xi_i, Y))], \tag{11}$$

$$G(\Psi(\xi_i, Y)) = \left(\prod_{j=1}^q \Psi(\xi_i, y_j) \right)^{1/q}, \Psi(\xi_i, y_j) > 0, 1 \leq j \leq q. \tag{12}$$

Based on this formulation, the selected features of both δ and δ_1 are simply concatenated using a serial approach. After applying the serial approach, the conditional entropy is implemented to remove the uncertainty among features. The formulation of CE is formulated by Eqs. (13) and (14):

$$H(\xi_{i+1}, \xi_i) = - \sum_{\xi_k \in \xi} p(\xi_k) \sum_{\xi_{k+1} \in \xi} p(\xi_{k+1} | \xi_k) \log p(\xi_{k+1} | \xi_k), \tag{13}$$

$$H(\xi_{i+1}, \xi_i) = - \sum_{\xi_k \in \xi} \sum_{\xi_{k+1} \in \xi} p(\xi_k, \xi_{k+1}) \log_2 p(\xi_{k+1} | \xi_k). \tag{14}$$

The entropy vector is sorted into descending order and defines a threshold function to select the best features. A threshold function is formulated by Eq. (15):

$$\xi_{sl}(i, y) = \begin{cases} \xi_{xi}, & \text{if } \tilde{H} \geq \mu \\ \text{Remove} & \text{otherwise} \end{cases}, \tag{15}$$

where ξ_{xi} represents the selected features in each iteration, and the total number of iterations depends on the target error rate or 100 iterations, and μ denotes the mean value of entropy vector \tilde{H} . This function ensures that only those entropy features are selected that are greater than the mean value of \tilde{H} and the rest of them are removed. These selected features are embedded into a fitness function which is Fine KNN, and the error rate is computed, where the target error rate is 0.08. After meeting this error rate, the selected vector is obtained, denoted by ξ_s of dimension $N \times K_1$. Similarly, this formulation is performed for feature vector 2 and a vector is obtained of dimension $N \times K_2$ and denoted by ξ_{s1} .

Feature aggregation

Feature aggregation combines features in one matrix to increase the salient information for improved recognition accuracy. Moreover, the aggregation of features represents to reduce the overall vector length [59, 60]. In this work, for aggregation of deep learning features, we implemented a canonical correlation analysis (CCA) approach [61]. In this approach, the correlation is computed among two sets of features and find the higher correlated transformed features. Finally, the transformed features are combined as a resultant vector.

Consider, $\xi_s \in \mathbb{R}^{N \times K_1}$ and $\xi_{s1} \in \mathbb{R}^{N \times K_2}$ are two selected feature vectors, where N denotes training samples and K_1, K_2 represent the dimension of feature vectors ξ_s and ξ_{s1} , respectively. Let $\Delta_{ss} \in \mathbb{R}^{K_1 \times K_1}$ and $\Delta_{s1s1} \in \mathbb{R}^{K_2 \times K_2}$ denotes the covariance matrix of ξ_s and ξ_{s1} , respectively, and $\Delta_{s1s} \in \mathbb{R}^{K_1 \times K_2}$ is the between-sets covariance matrix, where $\Delta_{s1s} = \Delta_{ss1}^T$. The overall covariance matrix is defined as $\Delta \in \mathbb{R}^{(K_1+K_2) \times (K_1+K_2)}$ and computed by Eq. (16):

$$\Delta = \begin{pmatrix} \mathbb{C}(\xi_s) & \mathbb{C}(s, s_1) \\ \mathbb{C}(s_1, \xi_s) & \mathbb{C}(s_1) \end{pmatrix} = \begin{pmatrix} \Delta_{ss} & \Delta_{ss_1} \\ \Delta_{s_1s} & \Delta_{s_1s_1} \end{pmatrix}, \tag{16}$$

where \mathbb{C} represents covariance function and Δ denotes covariance matrix. The covariance is computed by $\mathbb{C} = \sum \frac{(s_i - \bar{s})(s_{i_1} - \bar{s}_{i_1})}{N}$. The key objective of CCA is to define a linear combination between $\xi_s^* = \omega_s^T \xi_s$ and $\xi_{s_1}^* = \omega_{s_1}^T \xi_{s_1}$ which maximize the pair-wise correlation through both feature sets as following Eqs. (17)–(19):

$$\text{Cor}(\xi_s^*, \xi_{s_1}^*) = \frac{\mathbb{C}(\xi_s^*, \xi_{s_1}^*)}{\sigma^2(\xi_s^*), \sigma^2(\xi_{s_1}^*)}, \tag{17}$$

$$\sigma^2(\xi_s^*) = \omega_s^T \Delta_{ss} \omega_s, \tag{18}$$

$$\sigma^2(\xi_{s_1}^*) = \omega_{s_1}^T \Delta_{s_1s_1} \omega_{s_1}. \tag{19}$$

Next, the problem of maximization is resolved through Lagrange Multipliers to satisfy the equality constraint. Finally, both transformed features are combined through the simple concatenation method is defined by Eq. (20):

$$\text{Fin}(V) = \xi_s^* + \xi_{s_1}^* = \begin{pmatrix} \omega_{\xi_s} \\ \omega_{\xi_{s_1}} \end{pmatrix}^T \begin{pmatrix} \xi_s \\ \xi_{s_1} \end{pmatrix}, \tag{20}$$

where $\text{Fin}(V)$ is the final aggregated feature vector embed into a Kernel ELM for final classification. The features in final aggregated feature vector $\text{Fin}(V)$ are also plotted in Fig. 7.

Kernel ELM

In this section, we explain the classification method which we are using for final recognition. The Extreme Learning Machine (ELM) is a Feed Forward Neural Network (FWNN) that consists of a single hidden layer [62]. As compared to NN, in ELM, a few parameters are required to train a model. In ELM, the weights and bias are not adjusted, whereas only hidden layers are needed to be attuned. Based on these characteristics of ELM, it has a faster convergence rate and learns better.

Given training features $\xi^* \in FV(V)$ and $\xi^* \in \{(f_i, y_j), i, j = 1, 2, 3, \dots, N\}$ where $f_i \in [f_{i1}, f_{i2}, \dots, f_{iN}]$ represent input feature vector and $y_j \in [y_{j1}, y_{j2}, \dots, y_{jN}]$ represent the corresponding labels. Then, the output function of KELM is defined by Eq. (21):

$$\Phi(f) = h(f)b = h(f)H^T \left(\frac{\Pi}{C} + HH^T \right)^{-1} \hat{O}, \tag{21}$$

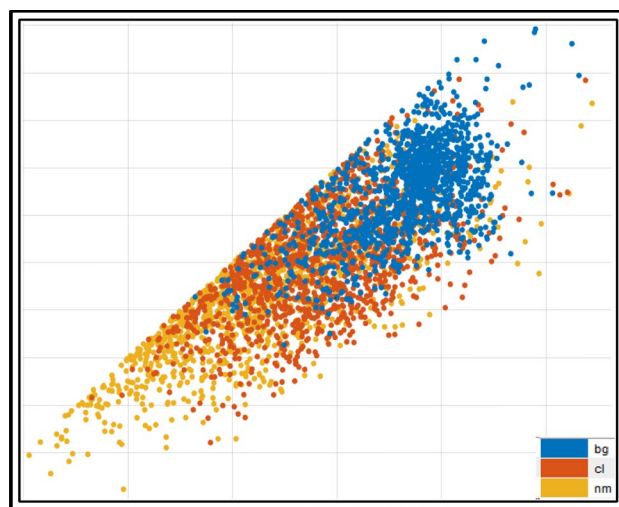


Fig. 7 Visualization of aggregated features in terms of scatter plots

where \hat{O} is target output, $\frac{\Pi}{C}$ is kernel parameter, Π is an identity matrix, C is the penalty parameter, and H represents an output matrix, respectively. Then, the kernel function of ELM is defined by Eqs. (22)–(24):

$$\widetilde{\text{KELM}} = HH^T, \tag{22}$$

$$\widetilde{\text{KELM}} = h(f_i)h(f_j) = \mathbb{K}(f_i, f_j), \tag{23}$$

$$g(f) = \begin{bmatrix} K(f, f_1) \\ \vdots \\ K(f, f_N) \end{bmatrix} \left(\left(\frac{\Pi}{C} + \widetilde{\text{KELM}} \right)^{-1} \hat{O} \right), \tag{24}$$

where $g(f)$ is a model function of ELM and $\mathbb{K}(f, f_i)$ is the kernel function of KELM. The kernel function is defined by $\mathbb{K}(f, f_i) = f \cdot f_i + \bar{b}$. As in this work, we are using the polynomial kernel function. Finally, the error between the output \hat{O} and target labels $Y \in y$ is calculated by Eq. (25) for final recognition.

$$\sum_{j=1}^N \left\| (\hat{O}_j - y_j) \right\| = 0. \tag{25}$$

The prediction results of the proposed scheme are shown in Fig. 8. In this figure, the testing videos are passed to the proposed scheme and in an output labeled results are generated (a normal walk, wearing a coat, and carrying a bag). We get these images in the testing phase, where each label is assigned based on the train model.

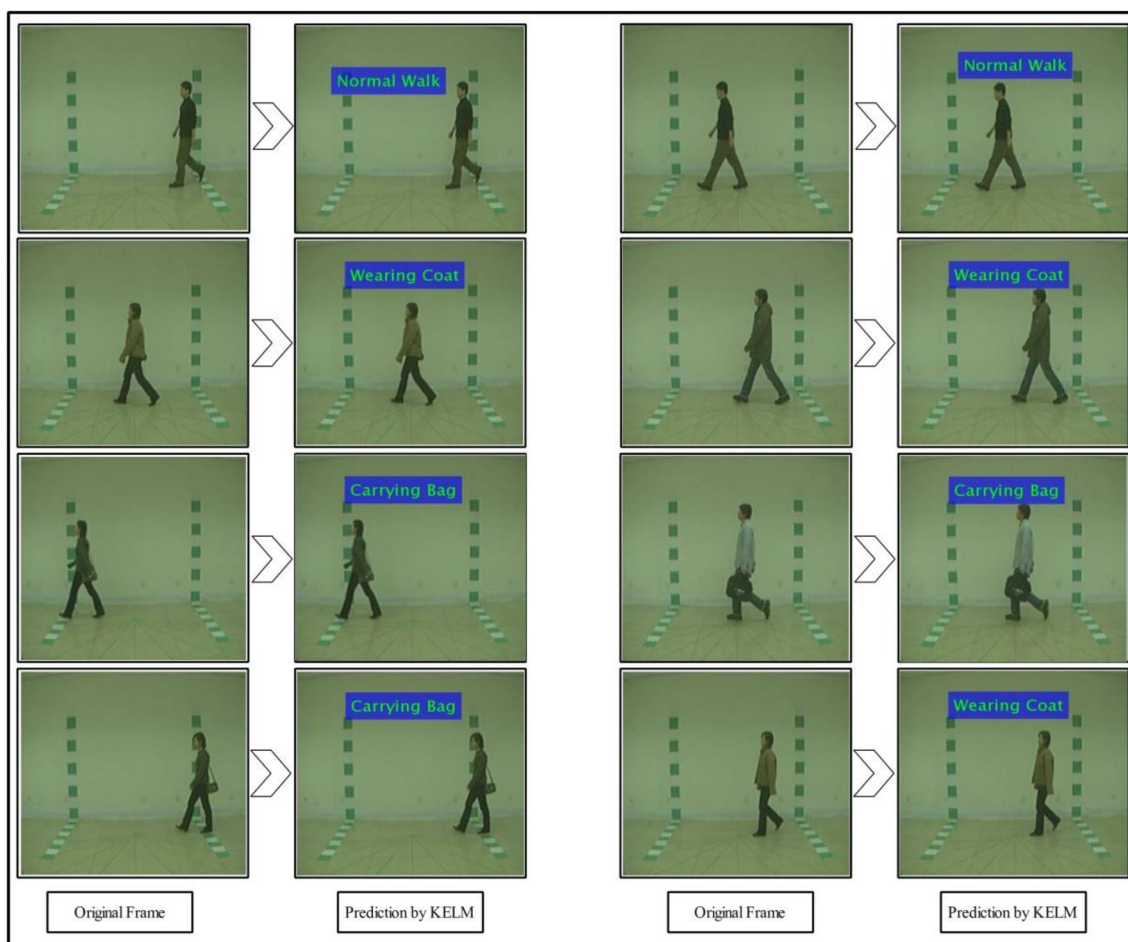


Fig. 8 Proposed recognition results using CASIA B (90°) for KELM Classifier

Results and analysis

We present the evaluation results of the proposed method using accuracy, figures, and visual plots for ROC curves, confusion matrix, and box plots. The results are computed from three different features sets as follows: (1) The features of AlexNet are computed from FC7, and the best amongst them are selected using the proposed ENGMwCE method. (2) The features of VGG16 Net are computed from FC7, and the robust features are selected using the proposed ENGMwCE selection method. (3) Here, aggregation of the robust features is performed. The selected features, using each feature set, are subjected to the KELM classifier for final recognition. The performance metrics used for the quantitative comparison include accuracy, error rate, and computational time. Moreover, the defined feature sets are also tested on a few other classifiers, such as ELM, Multi-class Support Vector Machine (MSVM), Fine Tree, Naïve Bayes, and Ensemble tree.

Implementation detail

The proposed method is implemented in a series of steps. Initially, we configure Matconvenet deep learning library and split the training data and testing (60, 40). After that, we re-train the existing deep learning networks using TL. For re-training, we set mini-batch size of 64 and learning rate of 0.006 for both. Later, we select the most discriminant information from both networks using Euclidean Norm and Geometric Mean Maximization along with Conditional Entropy (ENGMwCE). The Fine KNN is utilized for fitness evaluation, where the method is Euclidean distance and the number of neighbors selected is 10. The best-selected information is fused in the next step and passed the resultant information in the KELM. The target labels are provided to KELM for final output. Thereafter, we test our method on videos, and in the output; we get labeled results as few of them are shown in Fig. 8. A Personal Desktop Computer is used in the implementation where the specification of the system

is- Corei7, 256 SSD, 16 GB of RAM, and 8 GB Nvidia Graphics Card. Moreover, MATLAB2019b is employed as an implementation tool.

Results of CASIA B dataset (54°)

In Table 1, the results of 54° angle are presented for each of the three features sets using several widely adopted classifiers. The results must be interpreted as follows: the first row, for example, shows that the features set 1, when augmented with KELM classifier, manages to achieve an accuracy of 88.10%, against an error rate of 11.90%, for 14.296 s of computational time. While most of the table is self-explanatory in the same way, two important observations about the results are discussed next. Before proceeding, however, note that the entries in bold font highlight the best accuracies, minimum error, and computational time of the proposed features aggregation methodology, to assist in understanding the forthcoming description.

It is noted that the proposed aggregation methodology outperforms the other two feature sets for accuracy and error, irrespective of the classifier used. This is evident in the last row of each group of three rows—where the latter corresponds to a different classifier. This gives the proposed methodology an immense advantage over the existing ones. On the other hand, it is noted that the aggregation of features increases the execution time. For instance, observe the jump in the last column between rows 2 and 3 for KELM classifier. This trend continues for each classifier as well. The reason behind this is that after aggregation, the feature

dimension is increased, and the newly obtained vector adds more relevant information for correct recognition. Similarly, for other classifiers like ELM, Fine Tree, Naïve Bayes, MSVM, and Ensemble Tree, the accuracy achieved by the proposed method is significantly higher and the error rate is significantly lower than the other two approaches. However, the computational time increases in each case.

The KELM classifier outperforms all others for accuracy, error, and computational time. It is recommended that this classifier be augmented with the proposed aggregation technique for the best result. The performance of KELM for the feature set 1 is verified through Fig. 9, which presents the confusion matrix for the three walking statuses. Likewise, Fig. 10 presents the confusion matrix for the accuracy of KELM on the feature set 2. Finally, for the proposed approach with the KELM classifier, the accuracy may be confirmed by Fig. 11. Figure 12 also confirms the same result in the form of Receiver-Operating Characteristic (ROC) curves. From the latter, the true positive rate and area under curve (AUC) of each gait class is calculated, this later provides an accuracy value.

Results of CASIA B dataset (90°)

Table 2 presents the results of 90° angle for the used feature sets. All the observations made for the 90° angle hold equally true in this case as well. While the feature set 2 performs better than the features set 1 in terms of both accuracy and error rate. The proposed aggregation method, on the other hand, outperforms each competitor in the achieved

Table 1 Proposed recognition results using 54° angle of CASIA B dataset

Method	Feature set			Performance evaluations		
	Set 1	Set 2	Proposed	Accuracy (%)	Error (%)	Time (s)
KELM	✓			88.10	11.90	14.296
		✓		90.04	9.96	17.467
			✓	96.90	3.10	23.171
ELM	✓			85.42	14.58	21.979
		✓		85.96	14.04	19.117
			✓	93.70	6.30	27.528
Fine tree	✓			83.40	16.60	20.579
		✓		84.26	15.74	19.917
			✓	89.20	10.80	31.284
Naïve Bayes	✓			86.10	13.90	20.336
		✓		84.94	15.06	25.629
			✓	90.20	9.80	35.661
MSVM	✓			83.90	16.10	26.521
		✓		86.82	13.18	17.830
			✓	90.30	9.70	34.458
Ensemble baggage tree	✓			85.96	14.04	21.446
		✓		87.34	12.66	27.925
			✓	92.50	7.50	38.961

Fig. 9 Confusion matrix of KELM for feature set 1 (CASIA B of 54°)

Class	Gait Class		
	Carrying Bag	Wearing Coat	Normal Walk
Carrying Bag	86.5%	11.5%	2.0%
Wearing Coat	5.2%	89.7%	5.1%
Normal Walk	8.0%	4.0%	88.0%

Fig. 10 Confusion matrix of KELM for feature set 2 (CASIA B of 54°)

Class	Gait Class		
	Carrying Bag	Wearing Coat	Normal Walk
Carrying Bag	87.4%	10.1%	2.5%
Wearing Coat	4.4%	92.7%	2.9%
Normal Walk	6.0%	3.9%	90.1%

accuracy and error rate irrespective of the used classifier. Once again, the proposed method costs more computational time, due to the increase of the feature dimension.

The KELM classifier outperforms the other options in this case as well, where the achieved accuracy is confirmed by the confusion matrices illustrated in Figs. 13, 14, 15. In the first, the accuracy of the KELM classifier on the feature set 1 is verified by analyzing the diagonal values. The latter represents the correct prediction rate. In Fig. 14, the efficiency of the KELM classifier is verified for the feature set 2. Similarly, the accuracy of the KELM on the proposed features set is verified through Fig. 15. Moreover, the ROC curves are also plotted for the accuracy of the proposed scheme on the KELM classifier in Fig. 16.

Different view angles results (CASIA-B dataset)

Table 3 shows the recognition results of different view angles using CASIA-B dataset. The results are presented for three variations such as carrying a bag, normal walk, and walk with wearing a coat. The best accuracy for 36° is 91.46% on KELM classifier. Accuracy of other classifiers is 89.77%, 89.93%, 88.15%, 90.41%, and 89.80%, respectively. For 54°, the KELM classifier gives better performance. The achieved accuracy of KELM is 96.90%, whereas the second-best performance is 93.70%. Similarly, the KELM classifier gives better accuracy for the rest of the selected view angles such as 72°, 90°, 108°, and 144°, respectively. Their achieved accuracy are 94.20%, 96.50%, 91.33%, and

Fig. 11 Confusion matrix of KELM using the proposed method (CASIA B of 54°)

Class	Gait Class		
	Carrying Bag	Wearing Coat	Normal Walk
Carrying Bag	97.0%	3.0%	-
Wearing Coat	3.0%	95.0%	2.0%
Normal Walk	-	2.0%	98.0%

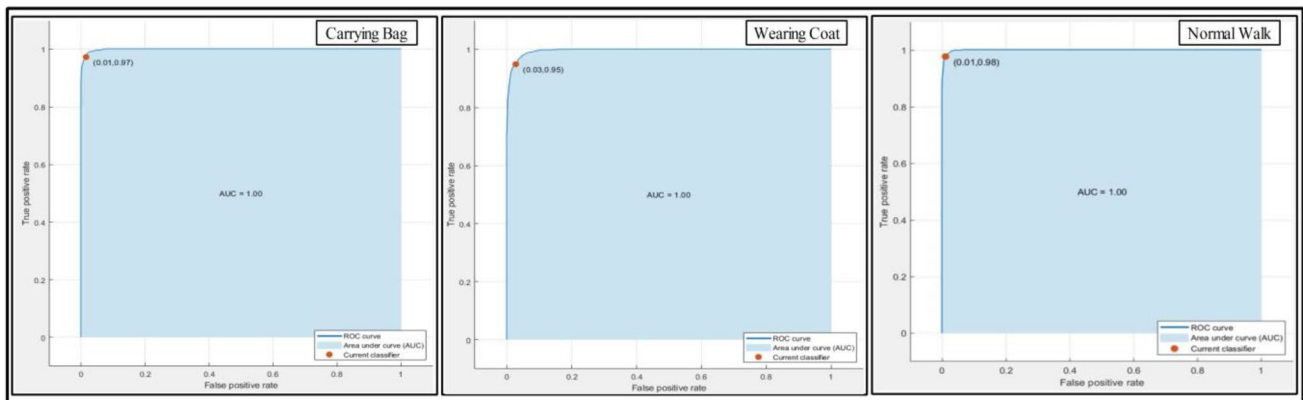


Fig. 12 ROC plots of CASIA B dataset using 54° angle

92.49%, respectively. It is also shown that the MSVM classifier also gives consistent accuracy.

Results on CASIA A dataset

The results are given in Table 4 for CASIA-A dataset using a testing ratio of 40% and CV = 10. In this table, it is presented that KELM gives better accuracy of 98.76% along with FNR is 1.24% and testing time 71.772 (s). The performance of KELM is also verified in Table 5. In this table, it is illustrated that the correct recognition rate of Oblique 45 class is 98.96%, whereas the Frontal 90 and Lateral 0 have 99.06% and 98.24% correct recognition rate. For other classifiers such as ELM, Fine Tree, Naïve Bayes, MSVM, and Baggage tree accuracies are 96.21%, 91.77%, 93.06%, 97.96%, 95.53%, respectively. In the last, we compare the proposed accuracy with a recent method, as given in Table 6. From this table, it is illustrated that the proposed method achieves

improved accuracy for Oblique 45 and Frontal 90, whereas on Lateral 0, the method presented in [40] gives better performance. Overall, the proposed method shows improved recognition accuracy on this dataset.

Discussion

For a fair evaluation of the feature aggregation approach, we have also performed the same simulations on two widely adopted feature sets, in conjunction with several renowned classifiers. We have demonstrated that the proposed features aggregation methodology, once augmented with the Kernel Extreme Learning Machine (KELM) classifier, achieves the best Human Gait Recognition (HGR) accuracy, with minimum error rate; thereby outperforming the existing equivalents. Figures 17 and 18 shows scatter plots of the KELM best accuracy on the proposed method for 54° and

Table 2 Proposed recognition results using 90° angle of CASIA B dataset

Method	Feature set			Performance evaluations		
	Set 1	Set 2	Proposed	Accuracy (%)	Error (%)	Time (s)
KELM	✓			90.53	9.47	19.830
		✓		91.60	8.40	13.372
			✓	96.50	3.50	27.491
ELM	✓			87.23	12.77	17.883
		✓		86.18	13.82	19.047
			✓	91.36	8.64	28.004
Fine tree	✓			87.36	12.64	17.016
		✓		86.02	13.98	15.184
			✓	91.16	8.84	24.716
Naïve Bayes	✓			84.67	15.33	21.073
		✓		88.92	11.08	19.736
			✓	92.50	7.50	33.163
MSVM	✓			86.94	13.06	21.847
		✓		87.48	12.52	23.994
			✓	92.76	7.24	36.940
Ensemble Baggage Tree	✓			86.68	12.47	23.552
		✓		90.24	8.76	21.040
			✓	93.30	3.40	31.553

Fig. 13 Confusion matrix of KELM for feature set 1 (CASIA B of 90°)

Class	Gait Class		
	Carrying Bag	Wearing Coat	Normal Walk
Carrying Bag	86.3%	9.6%	4.1%
Wearing Coat	4.5%	95.0%	2.0%
Normal Walk	4.5%	2.1%	98.0%

90°. Based on these figures, the false predicted points are represented by the cross sign.

A detailed comparison is also conducted with recent techniques in Table 7. In this table, the techniques are mentioned that also use 54° and 90° of CASIA B dataset for human gait analysis. Recently, Asif et al. [42], presented a deep learning framework for gait recognition. For experimental analysis, they used 54° angle of CASIA B dataset and achieved an accuracy of 94.70%. Arshad et al. [40] presented a binomial distribution based approach for gait recognition and achieved 87.70% accuracy on 90° angle of CASIA B

dataset. Castro et al. [63] improved the previous accuracy up to 90.6%. Later, in [43], the authors improved the current accuracy of gait recognition using deep learning. They used CASIA B dataset and selected 90°, where the achieved accuracy was 93.30%. In the proposed work, we have achieved an improved accuracy of 96.50% and 96.90% on 54° and 90° using the proposed selected features aggregation in the KELM classifier, giving this work a substantial edge over the existing equivalents. Anusha and Jaidhar [49] suggested a novel binary descriptor and used feature dimensionality reduction to achieve 91.90% accuracy on 54° CASIA B

Fig. 14 Confusion matrix of KELM for feature set 2 (CASIA B of 90°)

Class	Gait Class		
	Carrying Bag	Wearing Coat	Normal Walk
Carrying Bag	87.5%	8.2%	4.3%
Wearing Coat	3.8%	92.8%	3.4%
Normal Walk	4.2%	1.5%	94.3%

Fig. 15 Confusion matrix of KELM for the proposed method (CASIA B of 90°)

Class	Gait Class		
	Carrying Bag	Wearing Coat	Normal Walk
Carrying Bag	95.6%	4.0%	0.4%
Wearing Coat	3.0%	95.0%	2.0%
Normal Walk	0.4%	2.0%	97.6%

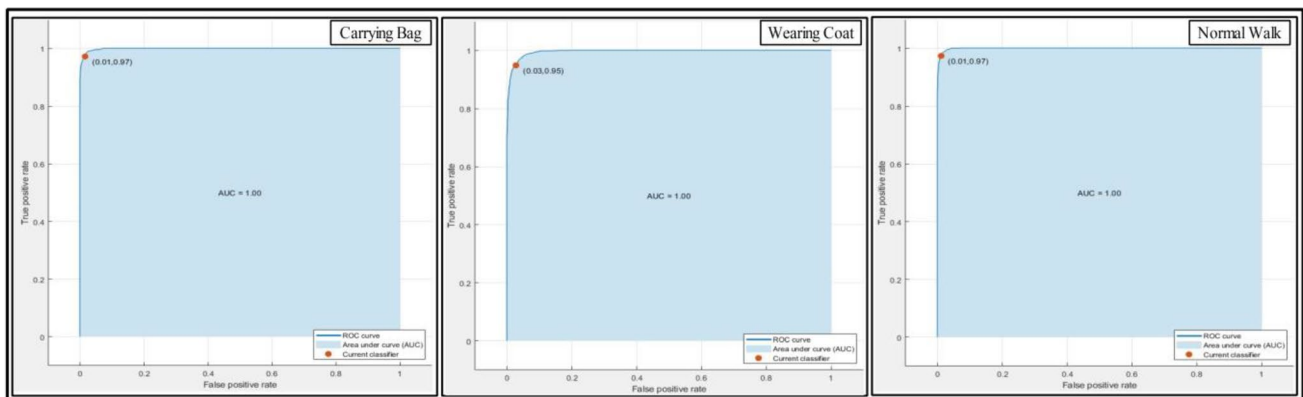


Fig. 16 ROC plots of CASIA B dataset using 90° angle

Table 3 Proposed recognition results using different view angles of CASIA B dataset

Classifier	View angles					
	36°	54°	72°	90°	108°	144°
KELM	91.46	96.90	94.20	96.50	91.33	92.49
ELM	89.77	93.70	92.55	91.36	87.54	90.55
Fine Tree	89.93	89.20	90.63	91.16	81.42	86.43
Naïve Bayes	88.15	90.20	88.04	92.50	88.90	87.17
MSVM	90.41	90.30	90.72	92.76	90.72	90.67
Ensemble Baggage Tree	89.80	92.50	91.14	93.30	88.53	89.91

Table 4 Proposed recognition results of the proposed method on CASIA-A dataset

Classifier	Validation measures		
	Accuracy (%)	FNR (%)	Time (s)
KELM	98.76	1.24	71.772
ELM	96.21	3.79	98.516
Fine Tree	91.77	8.23	117.442
Naïve Bayes	93.06	6.94	97.512
MSVM	97.96	2.04	90.153
Ensemble baggage tree	95.53	4.47	105.421

Table 5 Confusion matrix of KELM using CASIA-A dataset

Gait Class	Gait class		
	Oblique 45 (%) (%)	Frontal 90 (%)	Lateral 0 (%)
Oblique 45	98.96	1	< 1
Frontal 90	< 1	99.06	< 1
Lateral 0	1	< 1	98.24

Table 6 Comparison with existing techniques for CASIA-A dataset

Method	Year	Gait angles of CASIA-A dataset		
		Oblique 45 (%)	Frontal 90 (%)	Lateral 0 (%)
Habiba et al. [40]	2019	97	99	99
Proposed	2020	98.96	99.06	98.24

images. Leyva and Sanchez [50] suggested a spatio-temporal binary descriptor and combined it with Fisher Vectors to obtain 84.90% accuracy on 54° CASIA B images.

To analyze the proposed method, we have performed the statistical analysis. In the analysis, we perform simulation up to 500 iterations, and in the output, three values are achieved: minimum, average, and maximum. Based on these

values, we also determine the confidence interval. Table 8 shows the values for the selected view angles such as 36°, 54°, 72°, 90°, 108°, and 144°, respectively. We only find these values for KELM to check the consistency of the proposed scheme. From this table, it is illustrated that the values of the proposed method are consistent after the selected iterations. A very minor change occurs due to an update in the feature values. In this table, we also calculate the CI at confidence level 95%, $1.960\sigma_{\bar{x}}$. Also, a standard error mean has been calculated for each view angle and it is shown that the performance of the proposed method on 108° is much consistent as compared to other view angles.

Conclusions

In an attempt to improve the efficiency of human gait analysis (HGA) using deep learning, we have proposed an entire framework, mainly based on an aggregation of deep features, augmented with the Kernel Extreme Learning Machine (KELM) classifier. In this regard, a novel mechanism called Euclidean Norm and Geometric Mean Maximization along with Conditional Entropy has been proposed for selecting the most relevant and robust deep features. Canonical Correlation Analysis (CCA) has been employed for feature aggregation. The evaluation of the proposed scheme is performed on a publicly available gait database named CASIA B.

From the above discussion, we conclude that the selection of discriminant features from the original extracted feature sets improves the classification performance of the selected patient's gait. The original features include several redundant and extraneous features that affect the recognition performance. However, during the selection process, a few important features are also lost. Therefore, the aggregation of deep learning features that are extracted from two different CNN models (VGG16 and AlexNet), improve the features information, and fill the gap of the features removed during the selection process. It has been observed that following the aggregation process, the recognition accuracy improves, but the execution time increases, which is a limitation in this

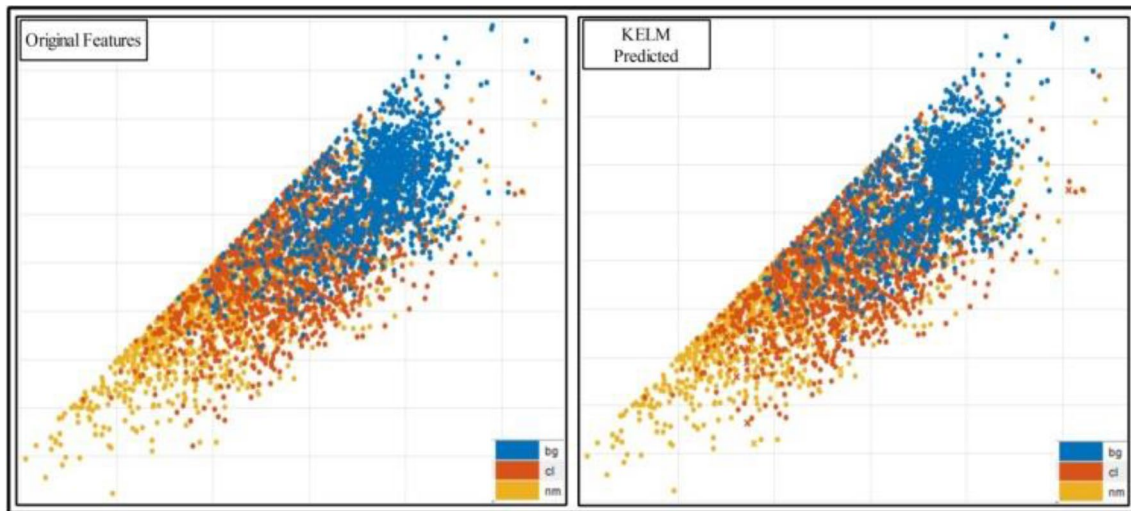


Fig. 17 Visualization of scatter plots for CASIA B dataset (54°)

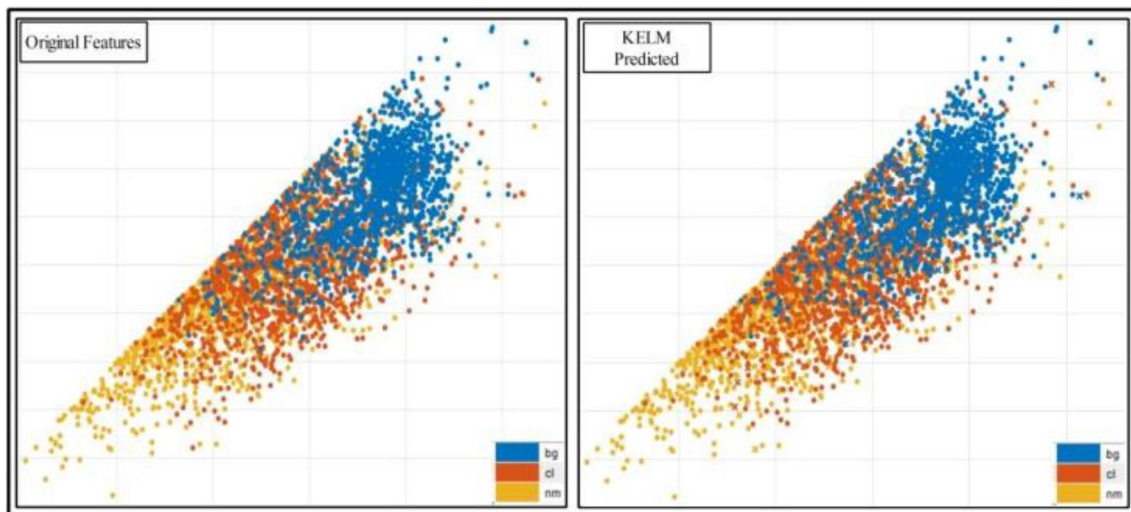


Fig. 18 Visualization of scatter plots for CASIA B dataset (90°)

Table 7 Comparison of proposed feature aggregation scheme with existing methods

Method	Year	Dataset	Angle	Accuracy (%)
Mehmood et al. [42]	2020	CASIA B	54°	94.70
Zeng et al. [44]	2014	CASIA B	90°	90.43
Yao et al. [64]	2019	CASIA B	54°	71.57
Arshad et al. [40]	2019	CASIA B	90°	87.70
Arshad et al. [43]	2020	CASIA B	90°	93.30
Zhang et al. [45]	2019	CASIA B	54°	91.00
Anusha and Jaidhar [49]	2020	CASIA B	54°	91.90
Leyva and Sanchez [50]	2019	CASIA B	54°	84.90
Proposed	2020	CASIA B	54°	96.50
Proposed	2020	CASIA B	90°	96.90

work. Moreover, we also conclude that the kernel selection of ELM is a major issue, which is yet to be addressed. Based on the kernel selection, we can easily analyze the performance of KELM. In the future, we aim to explore more optimistic approaches and evaluate more angles of CASIA B dataset. The developed method can contribute to ensuring security in smart cities by supporting intelligent video analytics of video surveillance material.

Table 8 Statistical analysis of the proposed method

Measures	View angle					
	36°	54°	72°	90°	108°	144°
Minimum	89.97	95.16	92.77	95.06	90.01	90.48
Average	90.71	96.03	93.48	95.78	90.67	91.485
Maximum	91.46	96.90	94.20	96.50	91.33	92.49
SEM	0.5267	0.6151	0.5055	0.5091	0.4666	0.7106
CI at (95%, $1.960\sigma_{\bar{x}}$)	90.715 ± 1.033 (± 1.14%)	96.03 ± 1.206 (± 1.26%)	93.485 ± 0.991 (± 1.06%)	95.78 ± 0.998 (± 1.04%)	90.67 ± 0.915 (± 1.01%)	91.485 ± 1.393 (± 1.52%)

Funding This research received no external funding.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Khan MA, Ashraf I, Alhaisoni M, Damaševičius R, Scherer R, Rehman A et al (2020) Multimodal brain tumor classification using deep learning and robust feature selection: a machine learning application for radiologists. *Diagnostics* 10:565
- Khan MA, Kadry S, Alhaisoni M, Nam Y, Zhang Y, Rajinikanth V et al (2020) Computer-aided gastrointestinal diseases analysis from wireless capsule endoscopy: a framework of best features selection. *IEEE Access* 8:132850–132859
- Hussain UN, Khan MA, Lali IU, Javed K, Ashraf I, Tariq J et al (2020) A unified design of ACO and skewness based brain tumor segmentation and classification from MRI scans. *J Control Eng Appl Informat* 22:43–55
- Majid A, Khan MA, Yasmin M, Rehman A, Yousafzai A, Tariq U (2020) Classification of stomach infections: a paradigm of convolutional neural network along with classical features fusion and selection. *Microsc Res Tech* 83:562–576
- Ismail ET, Abbas T, Javad S, Reza S (2020) Gait analysis of patients with piriformis muscle syndrome compared to healthy controls. *Musculoskelet Sci Pract*:102165
- Zhang Y, Jordan JM (2010) Epidemiology of osteoarthritis. *Clin Geriatr Med* 26:355–369
- Shull P, Lurie K, Shin M, Besier T, Cutkosky M (2010) Haptic gait retraining for knee osteoarthritis treatment. In: 2010 IEEE haptics symposium, 2010, pp 409–416
- Mabrouk AB, Zagrouba E (2018) Abnormal behavior recognition for intelligent video surveillance systems: a review. *Expert Syst Appl* 91:480–491
- COVID TC (2020) Characteristics of Health Care Personnel with COVID-19-United States, February 12–April 9, 2020. <https://www.cdc.gov/mmwr/volumes/69/wr/pdfs/mm6915e6-H.pdf>
- Condell J, Chaurasia P, Connolly J, Yogarajah P, Prasad G, Monaghan R (2018) Automatic gait recognition and its potential role in counterterrorism. *Stud Conflict Terror* 41:151–168
- Barria P, Aguilar R, Delgado DS, Moris A, Andrade A, Azorin JM (2020) Instrumented gait analysis of stroke patients after FES-cycling therapy
- Choudhury SD, Tjahjadi T (2015) Robust view-invariant multi-scale gait recognition. *Pattern Recogn* 48:798–811
- Sharif M, Attique M, Tahir MZ, Yasmim M, Saba T, Tanik UJ (2020) A machine learning method with threshold based parallel feature fusion and feature selection for automated gait recognition. *JOEUC* 32:67–92
- Damaševičius R, Vasiljevas M, Šalkevičius J, Woźniak M (2016) Human activity recognition in AAL environments using random projections. *Comput Math Methods Med* 2016
- Khan MA, Javed K, Khan SA, Saba T, Habib U, Khan JA et al (2020) Human action recognition using fusion of multiview and deep features: an application to video surveillance. *Multimedia Tools Appl*:1–27
- Damaševičius R, Maskeliūnas R, Venčkauskas A, Woźniak M (2016) Smartphone user identity verification using gait characteristics. *Symmetry* 8:100
- Bouchrika I (2018) A survey of using biometrics for smart visual surveillance: gait recognition. *Surveillance in action*. Springer, Berlin, pp 3–23
- Khan MH, Li F, Farid MS, Grzegorzec M (2017) Gait recognition using motion trajectory analysis. In: International conference on computer recognition systems, pp 73–82
- Li X, Makihara Y, Xu C, Yagi Y, Ren M (2019) Joint intensity transformer network for gait recognition robust against clothing and carrying status. *IEEE Trans Inf Forens Secur* 14:3102–3115
- Kusakunniran W, Wu Q, Zhang J, Li H (2012) Gait recognition under various viewing angles based on correlated motion regression. *IEEE Trans Circuits Syst Video Technol* 22:966–980
- Deng M, Wang C (2018) Human gait recognition based on deterministic learning and data stream of Microsoft Kinect. *IEEE Trans Circuits Syst Video Technol* 29:3636–3645
- Tafazzoli F, Safabakhsh R (2010) Model-based human gait recognition using leg and arm movements. *Eng Appl Artif Intell* 23:1237–1246
- Shirke S, Pawar S, Shah K (2014) Literature review: Model free human gait recognition. In: 2014 Fourth international conference on communication systems and network technologies, pp 891–895
- Zeng W, Wang C, Li Y (2014) Model-based human gait recognition via deterministic learning. *Cogn Comput* 6:218–229
- Wu Z, Huang Y, Wang L, Wang X, Tan T (2016) A comprehensive study on cross-view gait based human identification with deep cnns. *IEEE Trans Pattern Anal Mach Intell* 39:209–226

26. Song C, Huang Y, Huang Y, Jia N, Wang L (2019) GaitNet: an end-to-end network for gait based human identification. *Pattern Recogn* 96:106988
27. Kovač J, Štruc V, Peer P (2019) Frame-based classification for cross-speed gait recognition. *Multimedia Tools Appl* 78:5621–5643
28. Gabryel M, Damaševičius R (2017) The image classification with different types of image features. In: *International conference on artificial intelligence and soft computing*, pp 497–506
29. Hussain N, Khan MA, Sharif M, Khan SA, Albeshar AA, Saba T et al (2020) A deep neural network and classical features based scheme for objects recognition: an application for machine inspection. *Multimedia Tools Appl*. <https://doi.org/10.1007/s11042-020-08852-3>
30. Elboushaki A, Hannane R, Afdel K, Koutti L (2020) MultiD-CNN: a multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences. *Expert Syst Appl* 139:112829
31. Khan MA, Sharif M, Javed MY, Akram T, Yasmin M, Saba T (2017) License number plate recognition system using entropy-based features selection approach with SVM. *IET Image Proc* 12:200–209
32. Sharif M, Khan MA, Zahid F, Shah JH, Akram T (2020) Human action recognition: a framework of statistical weighted segmentation and rank correlation-based selection. *Pattern Anal Appl* 23:281–294
33. Połap D, Woźniak M (2017) The use of wavelet transformation in conjunction with a heuristic algorithm as a tool for feature extraction from signals. *Inf Technol Control* 46:372–381
34. Sharif M, Khan MA, Faisal M, Yasmin M, Fernandes SL (2018) A framework for offline signature verification system: Best features selection approach. *Pattern Recogn Lett*
35. Woźniak M, Połap D, Napoli C, Tramontana E (2016) Graphic object feature extraction system based on cuckoo search algorithm. *Expert Syst Appl* 66:20–31
36. Khan MA, Sharif M, Akram T, Raza M, Saba T, Rehman A (2020) Hand-crafted and deep convolutional neural network features fusion and selection strategy: an application to intelligent human action recognition. *Appl Soft Comput* 87:105986
37. Saba T, Khan MA, Rehman A, Marie-Sainte SL (2019) Region extraction and classification of skin cancer: a heterogeneous framework of deep CNN features fusion and reduction. *J Med Syst* 43:289
38. Zheng S, Zhang J, Huang K, He R, Tan T (2011) Robust view transformation model for gait recognition. In: *2011 18th IEEE international conference on image processing*, pp 2073–2076
39. Castro FM, Marín-Jiménez MJ, Mata NG, Muñoz-Salinas R (2017) Fisher motion descriptor for multiview gait recognition. *Int J Pattern Recognit Artif Intell* 31:1756002
40. Arshad H, Khan MA, Sharif M, Yasmin M, Javed MY (2019) Multi-level features fusion and selection for human gait recognition: an optimized framework of Bayesian model and binomial distribution. *Int J Mach Learn Cybern* 10:3601–3618
41. Li C, Min X, Sun S, Lin W, Tang Z (2017) DeepGait: a learning deep convolutional representation for view-invariant gait recognition using joint Bayesian. *Appl Sci* 7:210
42. Mehmood A, Khan MA, Sharif M, Khan SA, Shaheen M, Saba T et al Prosperous Human Gait Recognition: an end-to-end system based on pre-trained CNN features selection
43. Arshad H, Khan MA, Sharif MI, Yasmin M, Tavares JMR, Zhang YD et al (2020) A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition. *Expert Syst*:e12541
44. Alotaibi M, Mahmood A (2017) Improved gait recognition based on specialized deep convolutional neural network. *Comput Vis Image Underst* 164:103–110
45. Zhang Z, Tran L, Yin X, Atoum Y, Liu X, Wan J et al (2019) Gait recognition via disentangled representation learning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4710–4719
46. Yu S, Chen H, Wang Q, Shen L, Huang Y (2017) Invariant feature extraction for gait recognition using only one uniform model. *Neurocomputing* 239:81–93
47. Marcin D (2017) Human gait recognition based on ground reaction forces in case of sport shoes and high heels. In: *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 2017, pp 247–252
48. Khan MH, Li F, Farid MS, Grzegorzec M (2018) Gait recognition using motion trajectory analysis. Springer, Cham, pp 73–82
49. Anusha R, Jaidhar C (2020) Clothing invariant human gait recognition using modified local optimal oriented pattern binary descriptor. *Multimedia Tools Appl* 79:2873–2896
50. Leyva R, Sanchez V, Li C-T (2019) Compact and low-complexity binary feature descriptor and fisher vectors for video analytics. *IEEE Trans Image Process* 28:6169–6184
51. Zheng S, Huang K, Tan T (2011) Evaluation framework on translation-invariant representation for cumulative foot pressure image. In: *2011 18th IEEE international conference on image processing*, pp 201–204
52. Voulodimos A, Doulami N, Doulami A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. *Comput Intell Neurosci* 2018
53. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. *Commun ACM* 60:84–90
54. Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW (2017) Action recognition in video sequences using deep bi-directional LSTM with CNN features. *IEEE Access* 6:1155–1166
55. Muhammad K, Ahmad J, Mehmood I, Rho S, Baik SW (2018) Convolutional neural networks based fire detection in surveillance videos. *IEEE Access* 6:18174–18183
56. Pan SJ, Yang Q (2009) A survey on transfer learning. *IEEE Trans Knowl Data Eng* 22:1345–1359
57. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 1097–1105
58. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*
59. Khan MA, Sarfraz MS, Alhaisoni M, Albeshar AA, Wang S, Ashraf I (2020) StomachNet: optimal deep learning features fusion for stomach abnormalities classification. *IEEE Access* 8:197969–197981
60. Wang C, Elazab A, Wu J, Hu Q (2017) Lung nodule classification using deep feature fusion in chest radiography. *Comput Med Imaging Graph* 57:10–18
61. Thompson B (2005) Canonical correlation analysis. *Encyclopedia of statistics in behavioral science*
62. Lv L, Wang W, Zhang Z, Liu X (2020) A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine. *Knowl Based Syst*:105648
63. Castro FM, Marín-Jiménez MJ, Guil N (2016) Multimodal features fusion for gait, gender and shoes recognition. *Mach Vis Appl* 27:1213–1228
64. Yao L, Kusakunniran W, Wu Q, Zhang J, Tang Z, Yang W (2019) Robust gait recognition using hybrid descriptors based on Skeleton Gait Energy Image. *Pattern Recogn Lett*