A Multilevel Paradigm for Deep Convolutional Neural Network Features Selection with an Application to Human Gait Recognition

Habiba Arshad[1], Muhammad Attique Khan[2], Muhammad Irfan Sharif[3], Mussarat Yasmin[1*], João Manuel R. S. Tavares[4], Yu-Dong Zhang[5], Suresh Chandra Satapathy[6]

[1]Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan

[2]Department of Computer Science, HITEC University, Taxila, Pakistan

[3]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

[4]Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Porto, PORTUGAL

[5] Department of Informatics, University of Leicester, Leicester LE17RH, UK

[6] School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, 751024 Odisha, India

*Corresponding author: mussaratabdullah@gmail.com, attique.khan440@gmail.com

Abstract- Human gait recognition (HGR) shows high importance in the area of video surveillance due to remote access and security threats. HGR is a technique commonly used for the identification of human style in daily life. However, many typical situations like change of clothes condition and variation in view angles degrade the system performance. Lately, different machine learning (ML) techniques have been introduced for video surveillance which gives promising results among which deep learning (DL) shows best performance in complex scenarios. In this article, an integrated framework is proposed for HGR using deep neural network and Fuzzy Entropy controlled Skewness (FEcS) approach. The proposed technique works in two phases: In the first phase, Deep Convolutional Neural Network (DCNN) features are extracted by pre-trained CNN models (VGG19 and AlexNet) and their information is mixed by parallel fusion approach. In the second phase, entropy and skewness vectors are calculated from fused feature vector (FV) to select best subsets of features by suggested FEcS approach. The best subsets of picked features are finally fed to multiple classifiers and finest one is chosen on the basis of accuracy value. The experiments were done on four well-known datasets namely AVAMVG gait, CASIA A, B and C. The achieved accuracy of each dataset was 99.8%, 99.7%, 93.3% and 92.2%, respectively. Therefore, the obtained overall recognition results lead to conclude that the proposed system is very promising.

Keywords: Gait recognition, CNN features, parallel fusion, features selection, recognition

1. Introduction

Human identification using biometric technology has become essential and most important topic for the past few decades in the field of Computer Vision (CV) and bioinformatics (Shah, Chen, Sharif, Yasmin, & Fernandes, 2017; Sharif, Akram, Raza, Saba, & Rehman, 2019). Gait is a natural and common behaviour of all human beings but from viewpoint of analysis, it is a very complicated phenomenon as it works with the collaboration of nerves, brain and muscles (Sharif et al., 2020). Human moving gestures have been groping by orthopaedists, physiotherapists and neurologists for long to examine and appraise the condition of patients, rehabilitation and treatment plans. Usually, human gait has been examined intuitively via pictorial interpretations but nowadays with the improvement in technology, human gait exploration can be represented empirically and quantitatively (Prakash, Kumar, & Mittal, 2018). Gait is a walking behaviour of human body that is used to distinctively recognize individuals at a distance from a camera even in less illuminating and dense environmental areas. Unlike former biometric techniques such as fingerprints, iris, hand veins and face (Shiraga, Makihara, Muramatsu, Echigo, & Yagi, 2016), it does not require subjects co-operation. Because of these discriminating characteristics, it has received a lot of attention from researchers and hence used in various applications like visual surveillance (A. Sharif et al., 2019), suspect identification and robot vision (Xu, Zhu, & Wang, 2018). This article has used gait recognition (S. L. Fernandes & Bala, 2016a), face detection (S. L. Fernandes & Bala, 2015), face recognition (S. Fernandes & Bala, 2013) and composite sketch matching (S. L. Fernandes & Bala, 2016b; L Fernandes & G Bala, 2017) approaches and found them innovative. The results provided by authors of these approaches are interesting when tested on real-world datasets.

In spite of the distinctive characteristics of gait features, there are various factors that influence gait recognition, for instance, camera viewpoints, load carrying, lighting condition, variation in clothing, walking speed and shadow under feet (Arshad, Khan, Sharif, Yasmin, & Javed, 2019). Therefore, for correct classification of gait, it is essential to construct a system that is robust enough to overcome these challenges. Many DL algorithms are specifically designed to deal with clothing variations, carrying objects, walking speed and single viewing angle while some are specific for larger appearance variants due to change in viewing angle. In addition, some of them are utilized to deal with performance of specific types of variations but the performance assessment for one particular view, multiple view, interactive and non-interactive situations are still needed (Siddiqui et al., 2018). These existing methods for gait recognition process can be divided into two groups: model and appearance based methodologies. Usually, appearance based approach deals with the moving gestures of human body and works on silhouettes by extracting gait features (Jain, Kumar, & Fernandes, 2017). Human silhouette extraction, detection of time interval, representation and recognition are common context that fall in the category of appearance based approach. Model based method is used to extract the individual's foot step parameters and it uses human body silhouette for describing gait. Although gait recognition frequently requires realtime and valuable results at low resolution, this model needs images with high resolution and its computation is really very expensive (Li, Min, Sun, Lin, & Tang, 2017).

Many computerized methods are proposed for video surveillance using classical feature transforms (Raza et al., 2018; Sharif et al., 2017; Sharif, Khan, Zahid, Shah, & Akram, 2019) but lately DL has been gathering attention in ML applications, for instance, biometrics, video surveillance and medical imaging (Alotaibi & Mahmood, 2017; M. A. Khan et al., 2020; Raza et al., 2018). As compared to handcrafted descriptors, a DL model tries to simulate the activities of human brain that come naturally. In the last few years, this kind of ML based approaches has achieved high level of classification accuracy that is not being obtained before. It can use a hierarchical method for learning of high level facts generalization. HGR using deep CNN models is commonly considered as an effective method as it can accurately represent training data and extract larger feature sets with complex patterns. A common deep network comprises of input layer, at least one hidden layer and an output layer such that every single layer has to do particular type of ordering and sorting process (Hershey et al., 2017). Deep networks contain as much as 150 hidden layers as oppose to traditional neural networks which usually contain only 2-3 hidden layers. DL models are also used to handle large sets of unlabelled and unstructured datasets (Ansari, Shah, Yasmin, Sharif, & Fernandes, 2018; Raza et al., 2018). These models are trained by NN architectures and larger sets of labelled data hence enabling them to acquire features directly from data without using handcrafted feature extraction. Nonlinear multiple layer processing and supervised or unsupervised feature learning are two main factors in DL (Vargas, Mosavi, & Ruiz). In nonlinear multi-layer processing, the current layer is used to take the product of former layer as input. Then, these layers are arranged in a hierarchical manner to extract useful data whereas supervised and unsupervised feature learning mainly deal with targeting class labels. Availability of these labels results in a supervised classification problem otherwise unsupervised one.

1.1. Motivation

The increasing number of crime rates and terror attacks in numerous locals generates alarming conditions to security organizations for the sake of citizens' security (Sharif, Raza, Shah, Yasmin, & Fernandes, 2019). Moreover, a large number of CCTV cameras commonly mounted in many roads, buildings and traffic signals require constant checking. The manual checking and identification of humans is considerably a difficult task; additionally, the biometric information extracted visually can be affected very deficiently because of poor biometric techniques, crime locations and image quality (M. Sharif, M. Raza, et al., 2019). This research has used the approaches introduced in (Martis, Gurupur, Lin, Islam, & Fernandes, 2018; Raza et al., 2018; M. Sharif, M. Raza, et al., 2019) and implemented the state-of-art technique presented in (S. L. Fernandes & Bala, 2016a). To overcome the limitations discussed in these articles, an automated HGR method has been proposed using advanced DL technique while taking into account the unique characteristics of humans under both controlled and uncontrolled environments.

1.2. Problem Statement

Enormous research work for human identification by gait has been conducted which comprises tasks of pre-processing, silhouette based feature extraction and classification. There are many problematic factors that affect HGR and also degrade its performance such as object carrying posture, dressing, illumination effects, walking speed and shadow under feet (M. Khan et al., 2019). However, in this research work, following problems were taken into consideration: i) Covariate factors such as dressing and carrying variations and view angles that can degrade HGR performance, ii) Difficulty to discriminate unidentified inter-group dissimilarities and intra-group variants of human gait in real world scenarios, iii) Viewing angles variations that can adversely affect the recognition process, iv) Many ML methods did not give better results for selection of handcrafted gait features, v) Changes in human posture that typically occur while using smart phones and carrying objects hence bringing intra-class variations that cause difficulty in gait recognition, vi) Incomplete gait cycle acquisition because of occlusions that can degrade recognition performance, and vii) View-invariant gait features also bring difficulty in gait recognition.

1.3. Contribution

Major contributions of this work are listed below.

I. To extract deep features, two pre-trained CNN models are applied and their information is fused via a parallel approach. In the proposed parallel fusion approach, both feature vectors are compared with each other and a strongly correlated feature is added into a fused matrix.

II. Entropy and Skewness vectors are computed from the fused matrix and a new approach FEcS is proposed to select the best subsets of features.

III. The best subsets of features are supplied to various classifiers to select a strong classifier based on their higher accuracy value.

IV. A feature based comparison was conducted and the performance of fused FV, 70% selected features and 50% selected features accuracy was compared for evaluation.

This article is organized as: In section 2, the related work for HGR is presented; section 3 contains a description of CNN model while the proposed methodology is discussed in section 4. In section 5, the experimental results, discussion and comparison of results with existing methods are described. Finally, section 6 concludes the article.

2. Related Work

Nowadays, DL is showing improved performance in the field of CV (Bokhari, Syedia, Sharif, Yasmin, & Fernandes; Muhammad Attique Khan et al., 2018; Rashid et al., 2018). Recently, many DL approaches are applied for human gait recognition that outperforms well-known handcrafted descriptors (Ben et al., 2019; Gadaleta & Rossi, 2018; Wolf, Babaee, & Rigoll, 2016). Ozen et al. (Ozen, Boulgouris, & Swash, 2017) introduced a Holoscopic based HGR where one simple array of lenses is used to attach with the traditional cameras and Holoscopic gait energy image (HGEI) is built instead of traditional imaging process. This method gives better performance in complex and crowded scenarios as well as it enhances the competency of background subtraction. Chen et al. (Q. Chen, Wang, Liu, Liu, & Huang, 2017) presented a cross view HGR method based on combination of CNN and Siamese net. In this approach, CNN is used to take the image of random length as input to extract similar attributes. Feature vector pooling method is applied to get cumulative of extracted features. Then a system alike to Siamese is applied for classification. A large sample dataset OU-ISIR is utilized for the validation of the network so that it can efficiently extract and aggregates features from silhouette of gait image sequences. Sahak et al. (Sahak, Tahir, Yassin, & Kamaruzaman, 2017) presented a frontal and oblique view gait recognition method where Kinect is used for feature extraction, orthogonal least square for attribute selection and multi-layer perceptron for classification. Then the optimized Multi-Layer Perceptron (MLP) with two feature sets (orthogonal least square gait features and direct gait features) is used for the recognition of gait and evaluated its effectiveness by using neural network classifier that provides best classification results. Tang et al. (Tang, Luo, Tjahjadi, & Guo, 2017) presented a view invariant HGR technique using partial likeness between the individual's correspondences. A 3D framework is utilized in this method and a gait model is built with 2D gait patterns using individual's body pattern that contains posture and silhouette distortion. A GPSM model is presented for extraction of multiple views partial likeness features and then multiple linear subspace classifiers are combined with high weighted selection to get correct detection of individuals. The trial outcomes for the proposed method have proved that it is strong enough for various covariant factors. Shiqi et al. (Yu, Chen, Wang, Shen, & Huang, 2017) introduced a uniform deep model for gait recognition based on an auto encoder to extract covariant features. In this approach, multi view gait sequences are converted into any single view. It also eliminates the effect of wearing and carrying objects by individuals that affect the gait identification. The effectiveness of proposed model was assessed on the gait databases SZU RGB-D and CASIA B and obtained good recognition results. Chen et al. (C. Wu, Zhang, & Song, 2018) presented a multiple view gait recognition technique. In this approach, a two dimensional enhanced GEI technique is used for feature extraction by preserving the structural information of human silhouette through remodelling of 2DPCA. After this, non-negative pattern decomposition is applied to acquire organized local features for precision loss during view variation. At the end, 2DLDA is taken for feature estimation in a discriminant space for improvement of classification. Introduced method was applied on the CASIA B multi view gait dataset and compared it with Stack Aggressive AutoEncoder (SPAE) and deep CNN approaches to validate its effectiveness. Huimin et al. (H. Wu, Weng, Chen, & Lu, 2018) presented deep convolutional location weight descriptor technique for merging deep and simple features into a generally unified pre-trained architecture for classification. From the input layer, deep feature vectors are extracted first and then the filtered function is trained with guiding weights in accordance with fully connected (FC) layer for input layer. The handcrafted features are extracted through these guiding weights. Two publicly available datasets OU-ISIR and CASIA B were used to validate the effectiveness of presented method. Battistone and Petrosino (Battistone & Petrosino, 2018) introduced a deep learning model for gait recognition named as time based chart for long term memory. This model learns time changes in chart and collaboratively utilizes temporal information and structured data based on deep network. This network then learns long and short term dependences collectively using graph pattern. Chen et al. (X. Chen, Weng, Lu, & Xu, 2018) presented a multi-gait recognition method for recognizing gait of the individual walking alone or with other people. LCRF method is presented to determine unchanged and unknown gait features of single and multi-gait. Features created using dense trails are taken out to enhance the effectiveness of acquired features for recognition.

2.1. Real-Time Processing

The real time HGR using video processing is an active research area due to the famous application intelligent video surveillance (Lai, Yang, & Chen, 2007) performing real time security monitoring with multiple cameras in the defence systems. But due to high complexities in this area like video analysis or understanding, a better system is still not available. The real time video processing is also used in sports especially in football and cricket to take decisive action upon violence (Fenil et al., 2019). Although more recent real time gait recognition techniques are introduced but still they have many limitations to achieve required performance (Do, Nguyen, & Kim, 2019; Sien, Lim, & Au, 2019).

2.2. Low quality Video Processing

HGR under low quality video sequences is a difficult task due to cluttered background and hidden information. Image compression is an important aspect in video processing because the information in video is compressed which may lead to the loss of important features (Kamble, Thakur, & Bajaj, 2018). Therefore, it is essential to perform enhancement techniques to get better video frames leading to a significant recognition rate (Damahe & Thakur, 2019).

3. Convolutional Neural Network (CNN)

A common CNN model is divided into two parts containing different layers where each layer performs distinct role. Feature learning part contains convolutional layers, activation or ReLU layers and pooling layer whereas classification part contains FC and softmax layers. Each CNN layer converts an input size to an output size of neuron initiation that is finally given to FC layer by mapping the input data in to 1D feature vector (Voulodimos, Doulamis, Doulamis, & Protopapadakis, 2018). The general architecture of a CNN is presented in Figure 1.

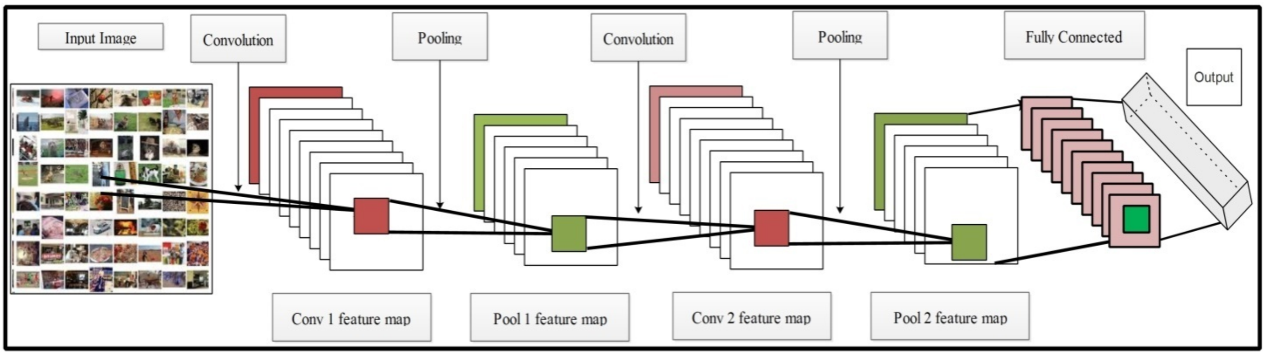

Figure 1: General architecture of a DCNN

4. Proposed Methodology

A new integrated system for HGR is proposed in this article by using DCNN features fusion and best subsets of features selection. The features are extracted from two well-known pre-trained CNN models known asAlexNet and VGG. The selection of these models was based on their excellent performance in previous studies. The main flow of the proposed method is presented in Figure 2.

4.1. AlexNet

AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) is a deep CNN model that has significantly improved the classification performance as compared with other models thus provoking the researchers interest in various CV applications. The input size of this network is 227×227×3. It is considered as a deeper network because it contains many filters at each layer. It has a total of 5 convolutional layers for feature extraction, max pooling layers, drop out, data augmentation, ReLU or activation layers, stochastic gradient descent (SGD) algorithm and 3 FC layers in order to classify with almost 60 million free parameters. The size of the first convolutional layer is 11×11 with 96 kernels and the second convolutional layer of size 5×5 with kernel size 256 came after max pooling layer. The third and fourth convolutional layers of size 3×3 with kernel size 384 are directly connected. The fifth convolutional layer with size 3×3 and kernel size 256 is followed by max pooling layer. The FC-6 and FC-7 layers contain 4096 neurons and FC-8 layer contains 1000 neurons for classification, therefore, a total of 1000 channels are possible for each class. Moreover, a drop out layer is used after each FC layer to decrease the problem of over fitting. In this work, FC layer 7 is used to perform activation for feature extraction. The size of the extracted FV is N×4096 which is utilized for the next step.

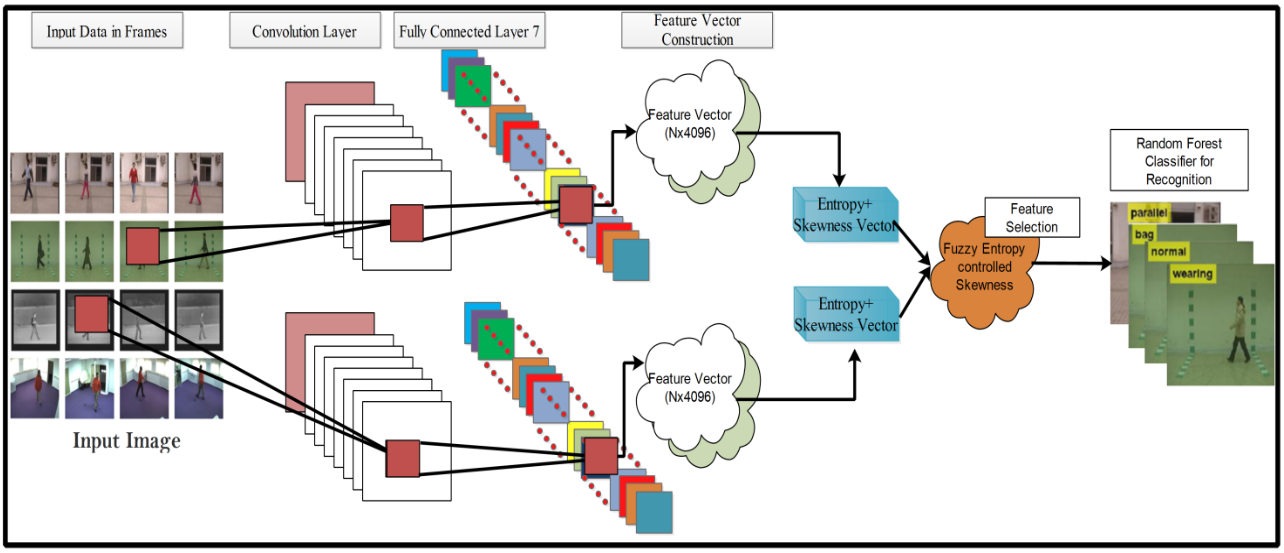

Figure 2: Proposed flow of HGR using Deep NN features fusion and FEcS based selection

4.2. VGG19

VGG (Visual Geometry Group) (Simonyan & Zisserman, 2014) is a simple deep network of 16 and 19 learnable weight layers with 16 convolutional layers of 64, 128, 256, 512 filter sizes and 3 FC layers of which two FC layers contain 4096 features and a third FC layer comprises of 1000 features. Final layer is softmax classification layer. The convolutional layers of size 3×3 and pooling layers of size 2×2 are used all over the net. FC layer 7 is utilized to extract deep features through the activation function. The resultant FV of size N×4096 is obtained which is used in the next step.

4.3. Higher Index Features Fusion

Features fusion is a traditional research area in the field of CV (Majid et al., 2020; Sharif, Tanvir, Munir, Khan, & Yasmin, 2018). It means to combine the patterns information of multi-properties features in one matrix. There are several advantages and limitations of features fusion. Major advantage is that the system accuracy is enhanced and redundant information is removed from the features (Akram, Khan, Sharif, & Yasmin, 2018; Liaqat et al., 2018). But the main limitation is efficiency of the system which means that overall execution time of system is increased while combining multiple features. In this article, a novel parallel features fusion method named as high index value selection (HIVS) is presented. In this approach, entropy and skewness vectors are calculated by utilizing original deep FVs like AlexNet and

VGG19. After vector construction, each index of both vectors is fused one by one by applying higher value index. The key condition of this technique is equal length of both vectors. The mathematical formulation of entropy and skewness is given as follows:

Entropy- In terms of mathematics, entropy is applied to determine the variability of issues. In information learning, entropy is applied to compute the average variability of data. It is used to evaluate non-uniform positions, for instance, variations and irregularity in vibrant parts of human gait. Entropy using a Markov process is termed as:

$$H(W) = -\sum p_k \log_2 p_k \tag{6}$$

Where, p_k is probability of k. In first order, the probability of selecting features of a Markov model is directly dependent on the prior entropy feature selection rate given as:

$$H(W) = -\sum_k p_k \sum_j p_k(j) \log_2 p_j(j) \tag{7}$$

Where, k is the state of certain prior features and $p_n(j)$ is the probability for j given k as prior features. In second order of Markov model, the entropy rate is defined:

$$H(W) = -\sum_k p_k \sum_j p_k(j) \sum_u p_{k,j}(u) \log_2 p_{k,j}(u) \tag{8}$$

Simply the t binary entropy of an individual is derived as $E = (E, P)$ with derived character object $E = \{i_1, \dots, i_k\}$ and probability distribution $P = \{j_1, \dots, j_k\}$. Here, j_i is probability of i_n that is termed as:

$$H_t(W) = -\sum_{k=1}^{l} p_k \log_t p_k \tag{9}$$

Where, binary entropy t is the total number of features that are used as a standard feature measure for source features. The length of the resultant entropy vector is $K \times 4096$.

Skewness- Skewness measure defined by Pearson gives information about the aggregate and direction for departure from symmetry. The values for symmetrical distribution can be positive and negative or may be undefined. If the absolute value for the measure of skewness is high, it gives more asymmetric distribution. This asymmetric distribution is used to discriminate humans who walk slowly or fast.

Central or k^{th} raw moments for the random variable G is $E(G^k)$, where E is the estimated value of random variable with k^{th} power. The first raw moment for the random variable Y is mean and it is symbolised by μ.

$$\mu = E(G) \tag{10}$$

Where, G denotes total pixels and μ is the mean value of those pixels. The second raw moment is named as variance and it is symbolised by σ^2 and its square root is standard deviation (SD) denoted by σ.

$$\sigma^2 = \frac{E[G-\mu]^2}{\sigma^2} \tag{11}$$

$$= \frac{E[G^2]-\mu^2}{\sigma^2} \tag{12}$$

$$\sigma = \sqrt{\frac{E[G^2]-\mu^2}{\sigma^2}} \tag{13}$$

The measuring ratio between SD and mean $\frac{\sigma}{\mu}$ is also known as coefficient of variation. The ratio for the 3rd central moment with the cube of SD is known as Pearson's coefficient of skewness and it is usually denoted by β_1:

$$\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{E\left[(G-\mu)^3\right]}{\sigma^3} \tag{14}$$

$$\beta_1 = \frac{E\left(G^3\right)-3\mu\sigma^2-\mu^3}{(\sigma^2)^{3}/_{2}} \tag{15}$$

Hence, the resultant feature vector is of dimension $L \times 4096$. Now for fusion of both Entropy and Skewness vectors $K \times 4096$ and $L \times 4096$, the index of each one is compared and higher feature is put into new matrix which denotes HIVS fusion. The fusion process is illustrated in Figure 3 where it is shown that originally features are computed from VGG19 and AlexNet for matrices of features. Later, from the original vectors, entropy and skewness vectors are calculated which are fused based on parallel high index value feature to obtain a new vector of dimension $\tilde{L} \times 4096$.

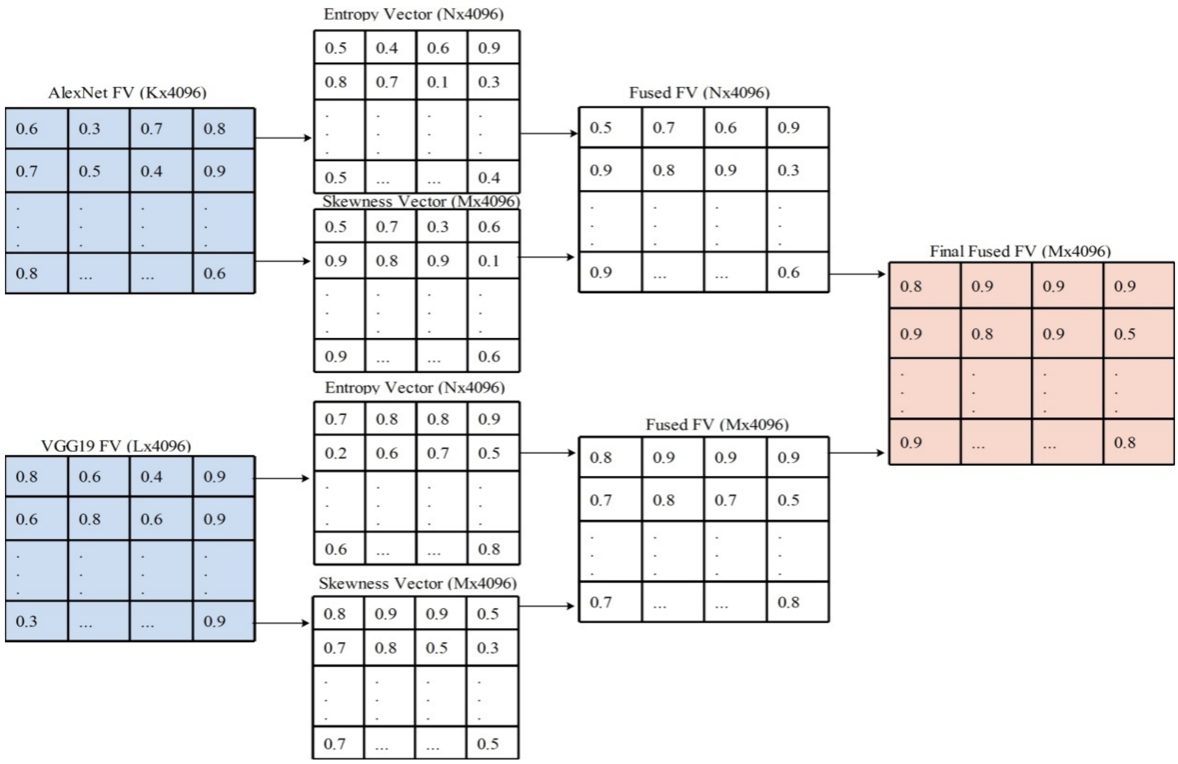

Figure 3: Proposed parallel features fusion working

4.4. *Fuzzy Entropy Controlled Skewness*

Several features selection techniques are implemented in literature by using the concept of entropy, skewness and Genetic Algorithm (GA) (M Attique Khan et al., 2018; Nasir et al., 2018; Sharif, Khan, Faisal, Yasmin, & Fernandes, 2018; Sharif, Khan, Iqbal, et al., 2018) but the fuzzy approach is not utilized which performs efficiently for video sequences. In this article, a novel technique is introduced named as FEcS to select best subsets for features selection. In information theory, Shannon entropy (Rajinikanth, Thanaraj, Satapathy, Fernandes, & Dey, 2019; Shannon, 1948) is broadly used to distinguish the impurities from sample spaces and compute the unpredictability related with an arbitrary variable. Fuzzy entropy is an extended idea of Shannon entropy with fuzzy sets which is used for entropy estimation. The concept of fuzzy entropy is entirely diverse from traditional Shannon entropy as fuzzy entropy is comprised of fuzzy unpredictability whereas Shannon entropy consists of random unpredictability (Kosko, 1986). In statistic, the conventional method to measure entropy uses the idea of probabilities created by histograms. The fuzzy logic or fuzzy entropy based entropy is measured through minimum time than it is passed through statistical methods with more suitable evaluations.

Suppose a random variable R with limited set of m components, $R = \{r_1, r, \dots, r_m\}$, where R describes the fused FV. If component r_i occurs with probability $p(r_i)$ then the information set $Y(r_i)$ related with r_i is termed as:

$$Y(r_i) = -\log_2 p(r_i) \tag{16}$$

Entropy E(R) of R is defined as:

$$E(R) = -\sum_{i=1}^{m} p(r_i)\log_2 p(r_i) \tag{17}$$

Where, m is the number of components and $p(r_i)$ represents the resultant probability of component r_i. Zadeh (Zadeh, 1976) introduced fuzzy entropy over a fuzzy set $\tilde{L}$ for a finite set $R = \{r_1, r, \dots, r_m\}$ using approximate probability distribution $P = \{p_1, p_2, \dots, p_m\}$ represented as:

$$E = -\sum_{i=1}^{m} \mu_{\tilde{L}}(r_i) p_i \log p_i \tag{18}$$

Where, $\mu_{\tilde{L}}$ represents the membership function of $\tilde{L}$, $\mu_{\tilde{L}}(r_i)$ means the category of association of r_i related to the fuzzy set $\tilde{L}$, p_i represents the probability r_i and $1 \leq i \leq m$. The difference in Equation (17) and Equation (18) is the expression $\mu_{\tilde{L}}(r_i)$ that can be used as a weighted multiplier.

Assume R is a fuzzy set on discourse $X = \{x_1, x_2, \dots, x_m\}$, the association vector of R is $R = (r_1, r_2, \dots, r_m)^s$, where $r_i = \mu_R(x_i) \in [0,1]$, then the fuzzy entropy for distance of R is given as:

$$E_d(R) = \frac{2}{m}\sum_{i=1}^{m} |r_i - \mu_{R^0}(x_i)| \tag{19}$$

Where, R^0 is a crisp group with minimum distance to R, the feature function for R^0 is represented as:

$$\mu_{R^0}(x_i) = \begin{cases} 0 & r_i < 0.5 \\ 1 & r_i \geq 0.5 \end{cases} \tag{20}$$

The features with index greater or equal to 0.5 are selected and multiplied by Skewness vector for ultimate classification. The selected FV is given to random forest (RF) classifier {Pal, 2005 #146} for final recognition. The proposed labelled outcomes are shown in Figures 4 and 5 for all selected datasets.

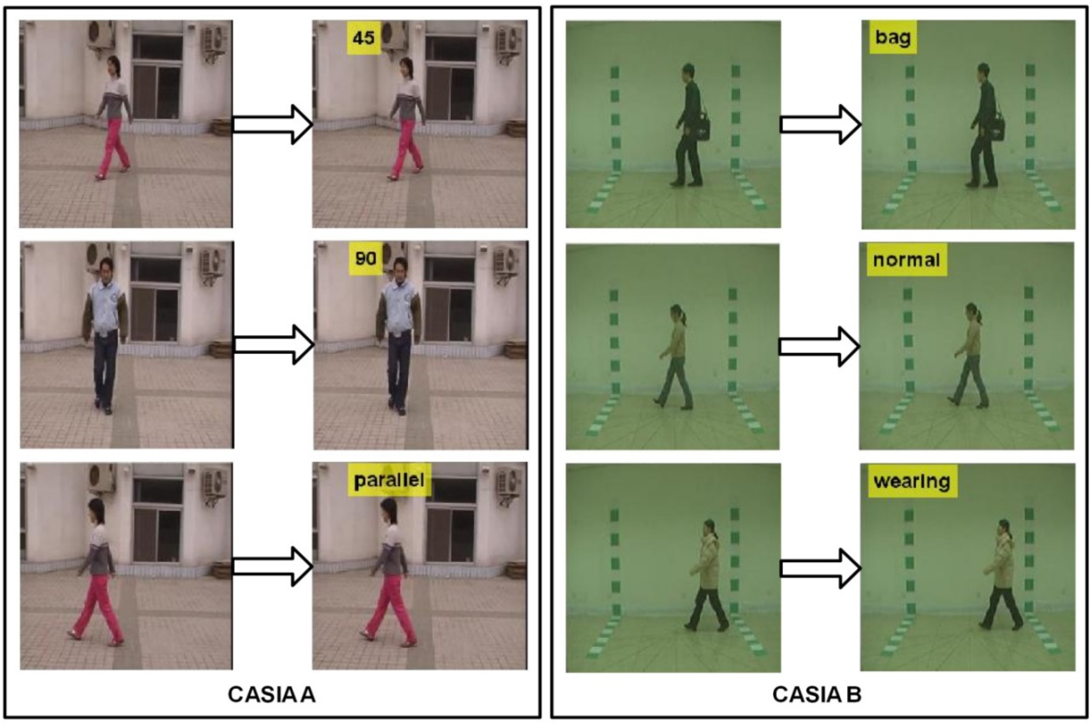

Figure 4: Proposed labelled outcomes for CASIA A and CASIA B datasets

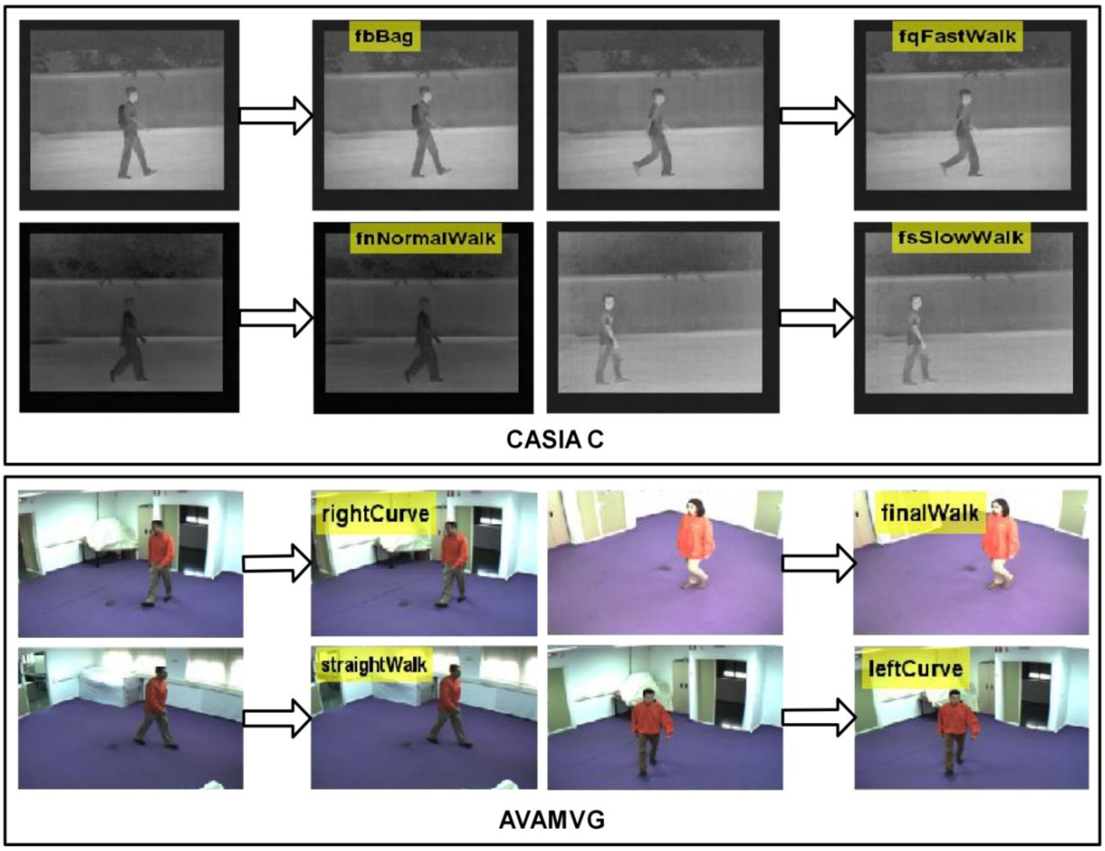

Figure 5: Proposed labelled outcomes for CASIA C and AVAMVG datasets

5. Experimental Results and Analysis

5.1. Experimental Setup

The validation of proposed system is presented in this section by using different datasets and performance metrics. Four publicly available datasets were utilized for this purpose and a 70:30 approach was applied for training and testing. After splitting the data, pre-trained model was loaded and cross entropy function was performed for activation. A mini batch size was initialized as 64 and learning rate was 0.001. This process was implemented on MATLAB2018a using Matconvnet deep learning toolbox (Vedaldi & Lenc, 2015). The measures used for system performance were sensitivity, precision rate, FNR, FPR, Area under Curve (AUC), F1-score and accuracy. Further, overall classification time of the proposed system was also calculated.

5.2. Performance Validation

The performance of proposed approach was compared in different steps. In the first step, extracted VGG and AlexNet features were fused and classification was performed. In the second step, the results of introduced FEcS based 70% features were selected and classification was performed. Likewise, in the third step, FEcS based 50% features were chosen for classification. The description of all steps is given below.

5.3. Datasets

Four publicly available datasets were utilized in this work: CASIA A gait dataset (Wang, Tan, Ning, & Hu, 2003), CASIA B gait dataset (Yu, Tan, & Tan, 2006), CASIA C gait dataset (Tan, Huang, Yu, & Tan, 2006) and AVAMVG multi-view gait dataset (López-Fernández, Madrid-Cuevas, Carmona-Poyato, Marín-Jiménez, & Muñoz-Salinas, 2014). The sample frames of each dataset are shown in Figure 6. The brief description for related datasets is as follows:

CASIA A gait dataset was recorded in an outdoor environment and in two alternate days. In this dataset, 20 subjects were involved to perform four gait sequences in three distinct views as laterally (0°), obliquely (45°) and frontal (90°). Hence the dataset contains $20\times4\times3=240$ number of gait sequences recorded in 25 fps (frames per second) with 352×240 image resolution and average length of each gait sequence is about 90 frames. In this work, 168 video sequences were utilized for training the system and others for testing.

CASIA B gait dataset is broadly used as a multi-view gait recognition dataset. Gait videos were recorded in an indoor environment through 11 different views using USB cameras and 124 subjects are included in recording that includes a total of 93 males and 31 females. The difference between each view angle direction is 18° arranged as 0º, 18º, 36º, 54º, 72º, 90º, 108º, 126º, 144º, 162º and 180º. Gait sequence for multi-view were recorded with three variations that include 6 video sequences for a normal walk (NW), 2 video sequences by wearing a coat (WC) and 2 video sequences by carrying bag (CB). The videos were recorded with a 320×240 frame size at the rate of 25 fps. Hence, the entire video sequences for the dataset is 10 ×11×124=13640. In this article, only 90º view was considered which includes a total of 1240 video sequences. Similar to CASIA A dataset, 70:30 approach was performed for validation.

CASIA C gait dataset is Infrared Night Gait Dataset recorded with a thermal infrared camera. Gait sequences for the dataset were recorded in an outdoor environment and 153 subjects were included (130 males and 23 females). In each gait sequence, four types of variations were recorded that includes walk with carrying-bag (CB), slow walk (SW), normal walk (NW) and quick walk (QW). In recording session, each individual had to walk in the sequence of 4 times normal walk, 2 times with carrying bag, 2 times walk slowly and 2 times walk quickly. Hence 1530 number of gait sequences was recorded with 320×240 resolution at the rate of 25 fps. Further, 1071 video sequences were used for training the proposed system and others for testing.

AVA multi-view gait dataset was recorded in an indoor environment and from six multiple views that gives a view of 360°. In recording session, 10 walking sequences were performed by 20 subjects that include 4 females and 16 males. Gait sequences were recorded in 4:3 formats with 640×480 image resolution at a rate of 25 fps. Each walking sequence includes 3 straight walking sequences, six curved sequences and finally walks in a straight path. In addition, 20,000 total numbers of extracted frames with image resolution 256×141 were considered for evaluation.

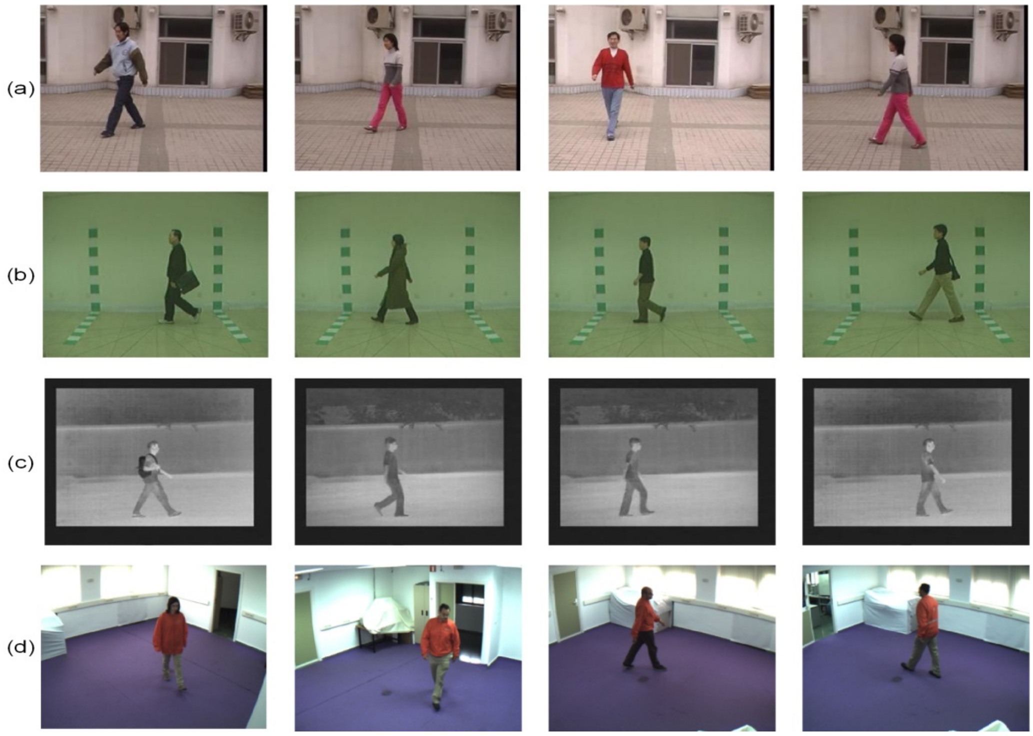

Figure 6: Sample frames of the selected gait recognition datasets

5.4. Results

The effectiveness of the proposed deep learning method is addressed in this section by applying 10-fold cross validation (10CV) on selected FV. The selected FV was randomly divided into 10 subgroups from which one group was utilized for testing and remaining nine groups for training. This method was repeated 10 times and single average value was calculated after 10 repetitions. Eight classification methods including Fine Tree (FT), Linear Discriminant Analysis (LDA), Linear SVM (LSVM), Quadratic SVM (QSVM), Cubic SVM (CSVM), Fine KNN (FKNN), Weighted KNN (WKNN) and RF are utilized for classification.

The recognition outcomes for the CASIA A dataset are presented in Table 1. The results are shown in three distinct feature sets such as fused FV, selection of only 70% features by proposed FEcS approach and selection of top 50% features. The best achieved accuracy of CASIA A dataset was 99.7% on RF classifier whereas other classification methods also achieved better accuracy on 50% selected features. From results, the accuracy of 70% selected features and fused FV was 99.3% which is also verified by the confusion matrices (CM) in Figure 7. The selection of features had effects on the efficiency of introduced framework which can be observed in Figure 8. In Figure 8, the best Execution Time (ET) was 34.29 seconds for RF.

Table 1: Classification outcomes of the proposed technique for CASIA A gait dataset. The words Sen denotes sensitivity rate, Pre denotes precision rate, FNR is false negative rate, Acc denote accuracy and C* denotes the classification method.

C*	**Features**			**Performance Metrics**					
	Fused	**FEcS 70%**	**FEcS 50%**	**Sen (%)**	**Pre (%)**	**FNR (%)**	**AUC**	**F_1 Score (%)**	**Acc (%)**
FT	✓			90.67	90.67	9.5	0.94	90.67	90.5
		✓		91.00	91.0	9.0	0.94	91.00	91.0
			✓	90.00	90.67	9.7	0.94	90.33	90.3
LSVM	✓			97.67	98.0	1.9	0.99	97.83	98.1
		✓		97.67	97.67	2.1	0.99	97.67	97.9
			✓	98.00	98.33	1.7	0.99	98.16	98.3
QSVM	✓			99.00	99.33	0.7	1.00	99.16	99.3
		✓		99.00	99.33	0.7	1.00	99.16	99.3
			✓	99.00	99.33	0.7	1.00	99.16	99.3
CSVM	✓			99.00	99.00	0.7	1.00	99.00	99.3
		✓		99.00	99.00	0.7	0.99	99.00	99.3
			✓	99.00	99.00	0.7	0.99	99.00	99.3
FKNN	✓			98.33	98.00	1.4	0.99	98.16	98.6
		✓		98.33	98.33	1.3	0.99	98.33	98.7
			✓	99.00	100.0	0.8	0.99	99.50	99.2
WKNN	✓			96.67	96.67	3.6	1.00	96.67	96.4
		✓		97.00	97.00	3.1	1.00	97.00	96.9
			✓	97.00	97.00	3.0	1.00	97.00	97.0
EBT	✓			98.00	98.00	2.0	1.00	98.00	98.0
		✓		97.00	97.33	2.5	1.00	97.16	97.5
			✓	98.00	98.33	1.5	1.00	98.16	98.5
RF	✓			99.33	99.00	0.7	0.99	99.16	99.3
		✓		99.33	99.00	0.7	0.99	99.16	99.3
			✓	**99.47**	**99.33**	0.3	**1.00**	99.40	**99.7**

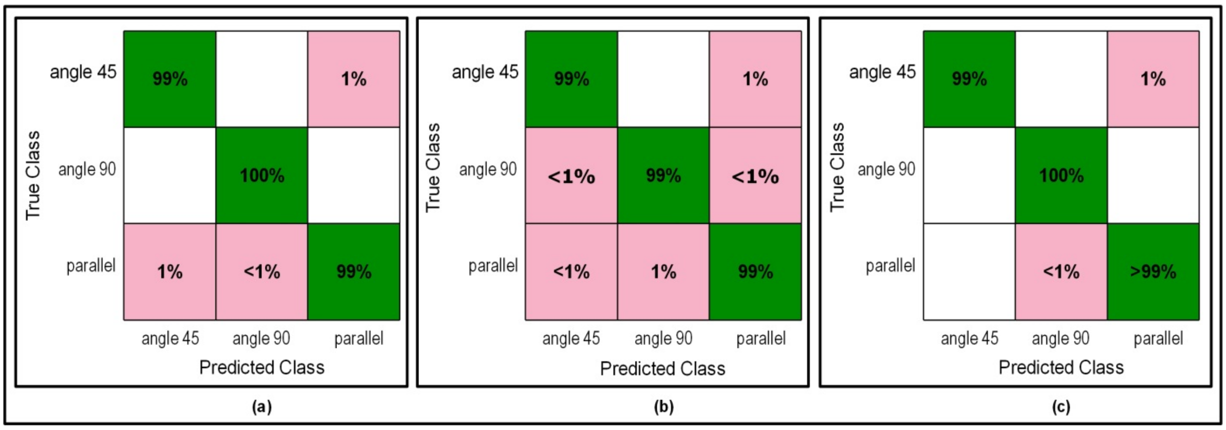

Figure 7: Confusion matrices obtained for CASIA A dataset (a) Fused FV, (b) Selection of top 70% features and
(c) Selection of top 50% features

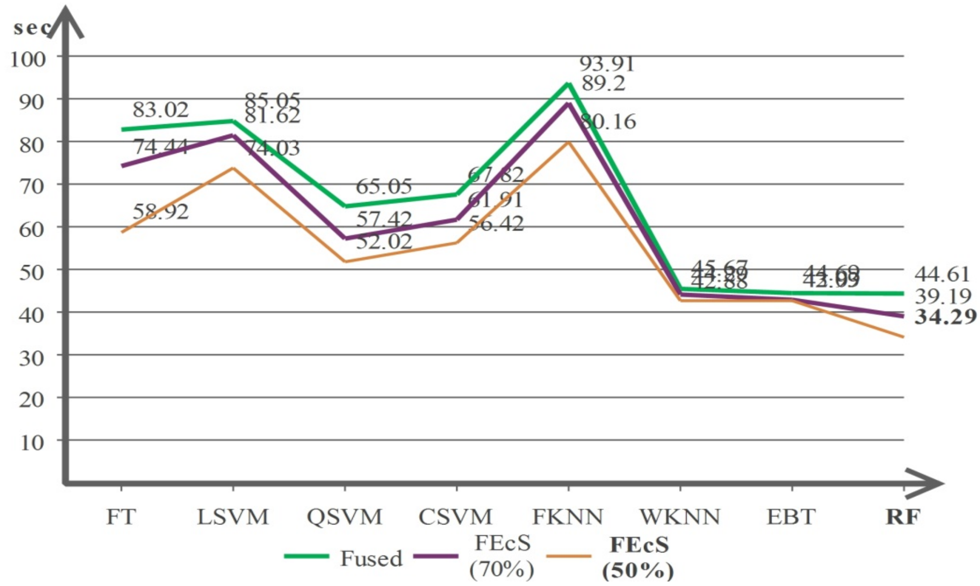

Figure 8: Classification time of the proposed technique for CASIA A dataset

The classification outcomes of the proposed technique using the CASIA B dataset are presented in Table 2. Similar to CASIA A, the results were obtained using three different feature sets. From Table 2, the accuracy of Fused FV was 93.4% on RF classifier which is verified by CM shown in Figure 9 (a). Then 70% and 50% proposed FEcS approach accuracy was 93.2% and 93.4% respectively which is also verified by Figure 9 (b) and Figure 9 (c). The change in results after selection of best subsets of features did not occur but the efficiency of proposed method was significantly improved as shown in Figure 10 where minimum ET was 98.7 seconds on RF whereas worst ET was 337.71 seconds on FT using fused FV.

Table 2: Classification outcomes of the proposed technique for CASIA B gait dataset

C*	Features	Performance Metrics

	Fused	**FEcS (70%)**	**FEcS (50%)**	**Sen (%)**	**Pre (%)**	**FDR (%)**	**FNR (%)**	**AUC**	F_1 **Score (%)**	**Acc (%)**
FT	✓			60.33	60.33	39.67	39.7	0.76	60.33	60.3
		✓		61.67	62.0	38.0	38.6	0.76	61.83	61.4
			✓	60.33	61.0	39.0	39.5	0.76	60.66	60.5
LSVM	✓			82.33	82.33	17.67	17.6	0.93	82.33	82.4
		✓		82.33	82.33	17.67	17.6	0.93	82.33	82.4
			✓	82.00	82.67	17.33	17.7	0.93	82.33	82.3
QSVM	✓			86.00	86.33	13.67	14.0	0.96	86.16	86.0
		✓		86.33	86.67	13.33	13.9	0.96	86.50	86.1
			✓	86.00	86.0	14.0	14.1	0.96	86.00	85.9
CSVM	✓			92.00	92.0	8.0	8.0	0.99	92.00	92.0
		✓		92.00	91.33	8.67	8.4	0.99	91.83	91.6
			✓	92.00	91.67	8.33	8.2	0.99	91.83	91.8
FKNN	✓			83.33	83.33	16.67	17.0	0.87	83.33	83.0
		✓		83.33	83.67	16.33	16.7	0.88	83.50	83.3
			✓	83.33	83.33	16.67	16.9	0.87	83.33	83.1
WKNN	✓			81.33	82.0	18.0	18.6	0.94	81.66	81.4
		✓		81.00	82.33	17.67	18.6	0.94	81.66	81.4
			✓	81.33	82.0	18.0	18.6	0.94	81.66	81.4
EBT	✓			77.0	77.33	22.67	22.8	0.92	77.16	77.2
		✓		78.67	79.0	21.0	21.1	0.92	78.83	78.9
			✓	77.67	78.0	22.0	22.2	0.92	77.83	77.8
RF	✓			93.33	93.33	6.67	6.6	0.99	93.33	93.4
		✓		93.33	93.0	7.00	6.8	0.99	93.16	93.2
			✓	**93.33**	**93.33**	**6.67**	**6.6**	**0.99**	**93.33**	**93.4**

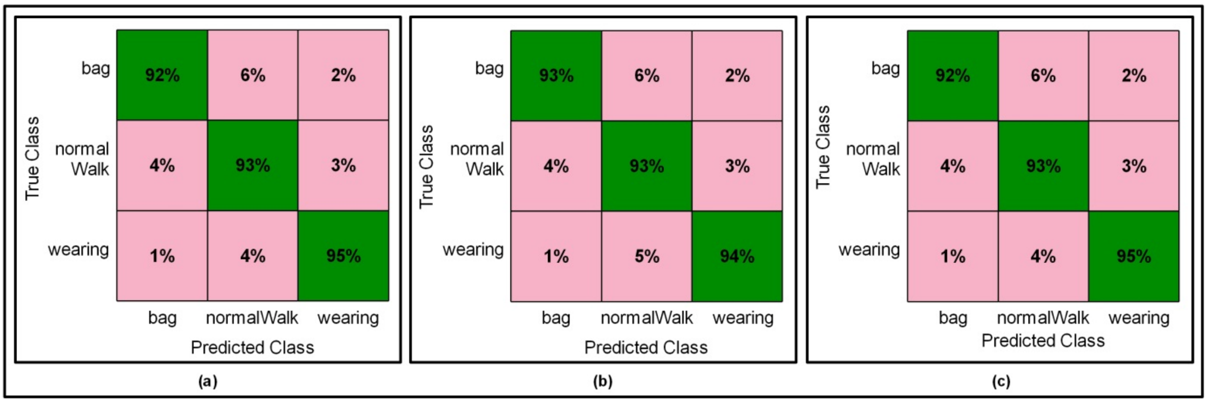

Figure 9: Confusion matrices obtained for CASIA B dataset (a) Fused FV (b) Selection of top 70% features and (c) Selection of top 50% features

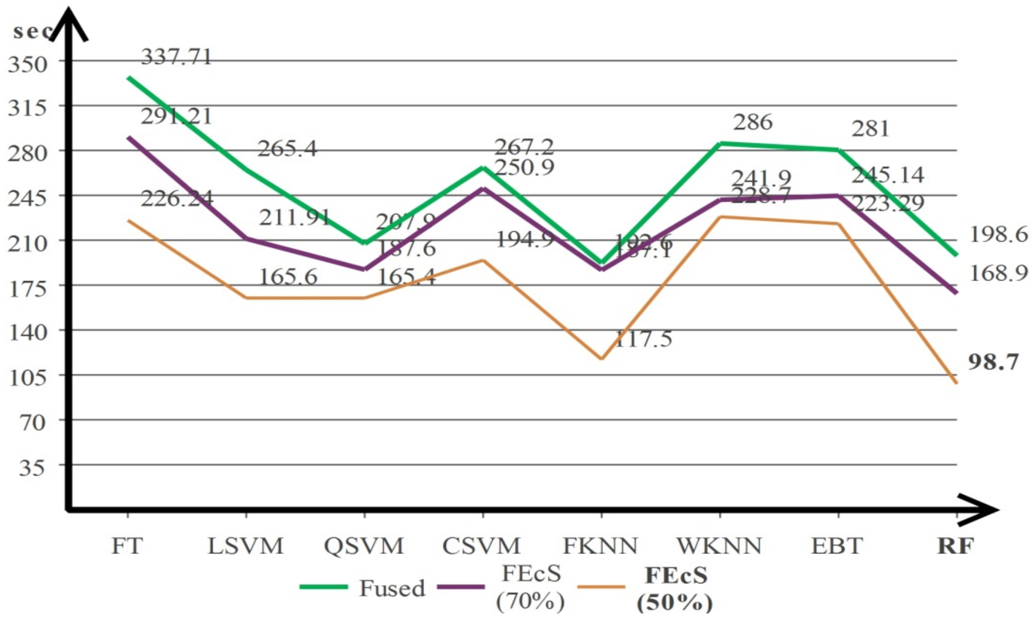

Figure 10: Classification time of the proposed technique for CASIA B dataset

The classification outcomes of the proposed technique using a CASIA C dataset are shown in Table 3. Fused FV, top 70% selected features and top 50% features were selected for evaluation and analysis of the performance of proposed solution. The results shown in Table 3 indicate that proposed model performed well on top 50% selected features and achieved high accuracy of 92.2% using RF classification method which can be observed by Figure 11 (c). Whereas the accuracy of Fused FV and top 70% selected features was 88.8% and 91.9% respectively, this is also verified by CM in Figure 11 (a) and Figure 11 (b). The results illustrate that the selection process improved the overall recognition accuracy on this dataset using all classification methods. In addition, the selection of best subsets of features rallied the efficiency of whole system as demonstrated in Figure 12 where the best reported ET is 23.21 seconds with RF classifier using top 50% of features obtained by proposed features selection approach.

Table 3: Classification outcomes of proposed technique for CASIA C gait dataset using different feature sets

C*	Features			Performance Metrics						
	Fused	**FEcS (70%)**	**FEcS (50%)**	**Sen (%)**	**Pre (%)**	**FDR (%)**	**FNR (%)**	**AUC**	**f_1 Score (%)**	**Acc (%)**
FT	✓			53.25	53.00	47.00	46.8	0.710	53.12	53.2
		✓		55.25	55.75	44.25	44.7	0.730	55.50	55.3
			✓	61.51	61.75	38.25	38.6	0.773	61.63	61.4
LSVM	✓			71.75	73.00	27.00	28.2	0.900	72.37	71.8
		✓		74.52	75.00	25.00	25.7	0.910	74.76	74.3
			✓	74.75	75.25	24.75	25.3	0.915	74.99	74.7
QSVM	✓			86.25	86.75	13.25	13.6	0.970	86.50	86.4
		✓		86.00	86.5	13.50	13.7	0.970	86.25	86.3
			✓	87.00	87.00	13.00	13.1	0.965	87.00	86.9
CSVM	✓			88.00	88.50	11.50	11.9	0.970	88.25	88.1
		✓		86.75	87.00	13.00	13.3	0.970	86.87	86.7
			✓	87.5	87.50	12.50	12.5	0.965	87.50	87.5
FKNN	✓			80.00	80.25	19.75	19.9	0.870	80.12	80.1
		✓		78.75	79.00	21.0	21.2	0.860	78.87	78.8
			✓	78.75	79.25	20.75	21.0	0.860	78.99	78.8
WKNN	✓			66.25	66.75	33.25	33.7	0.865	66.50	66.3
		✓		65.5	66.25	33.75	34.4	0.860	65.87	65.6
			✓	67.00	68.00	32.00	32.9	0.870	67.50	67.1
EBT	✓			70.50	70.25	29.75	29.4	0.885	70.37	70.6
		✓		69.25	69.75	30.25	30.8	0.890	69.50	69.2
			✓	71.50	72.00	28.00	28.5	0.895	71.75	71.5
RF	✓			88.75	88.75	11.25	11.2	0.925	88.75	88.8
		✓		91.75	92.00	8.00	8.1	0.950	91.87	91.9
			✓	**92.00**	**92.40**	**7.60**	**7.8**	**0.970**	**92.2**	**92.2**

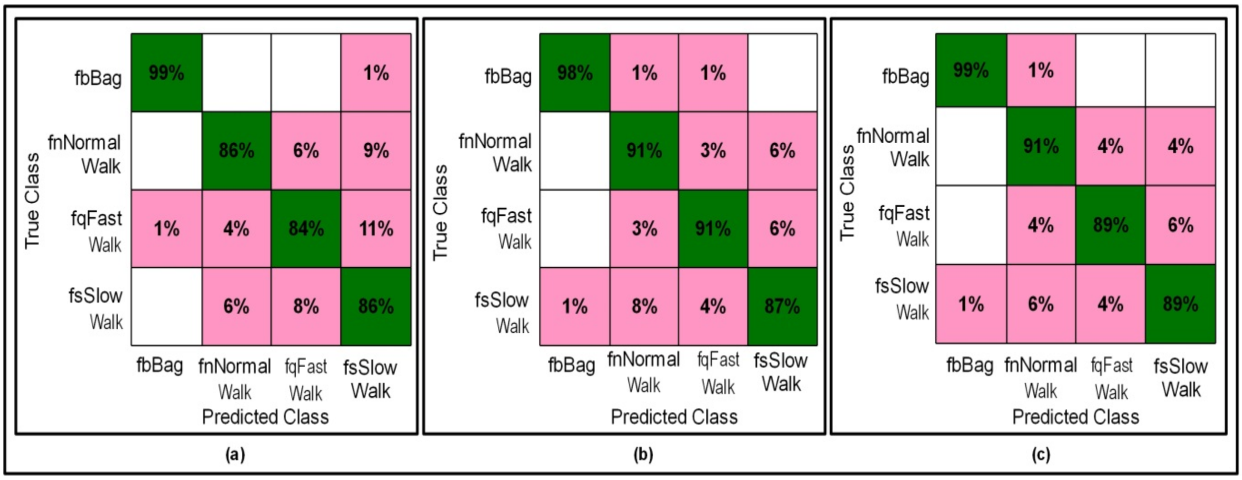

Figure 11: Confusion matrices obtained for CASIA C dataset (a) Fused FV (b) Selection of top 70% features and
(c) Selection of top 50% features

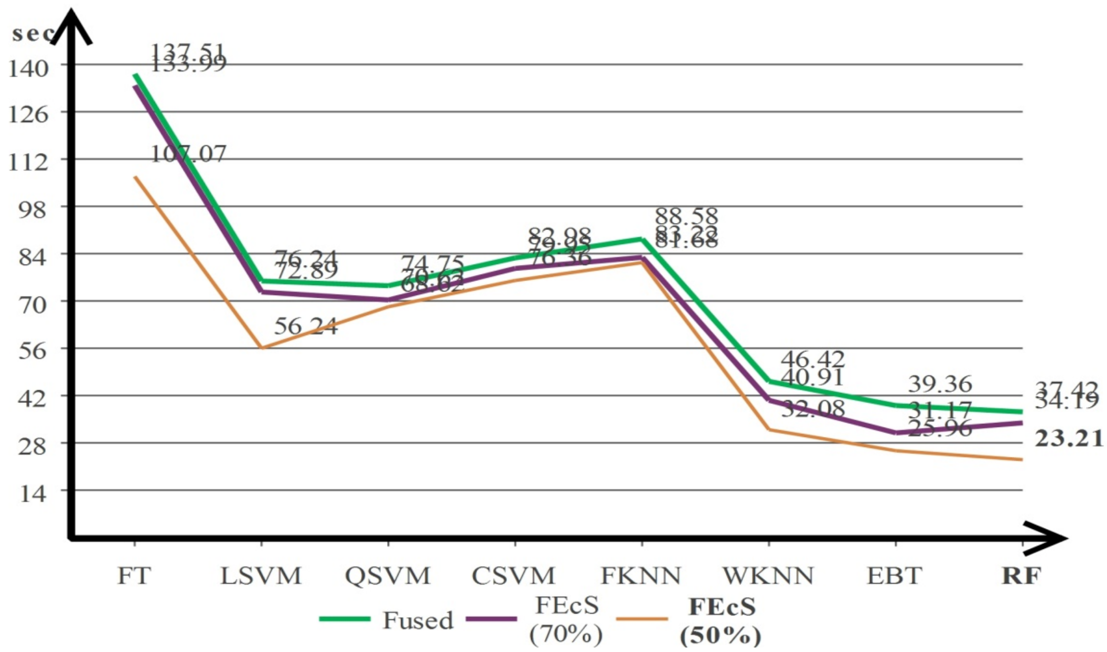

Figure 12: Classification time of the proposed technique for CASIA C dataset

The classification outcomes of the proposed approach using a AVAMVG gait dataset are indicated in Table 4. The outcomes were evaluated in distinct phases. In the first phase, features of both AlexNet and VGG19 were fused which were extracted from FC layer 7. The best obtained accuracy of this experiment was 99.8% which is verified by CM demonstrated in Figure 13(a). Moreover, the best ET of this phase was 139.5 seconds on CSVM. In second and third phases, top 70% and 50% features were selected and achieved an accuracy of 99.8% which is verified through Figure 13 (b) and Figure 13 (c). The major difference among fused and selected features was done through the efficiency of system illustrated in Figure 14 where high and best execution time of Fused FV was 190.1 seconds and 141.8 seconds whereas top

70% selected features executed in 182.62 seconds and 112.9 seconds. After the selection of 50% features, the accuracy of proposed system was still consistent and best ET was 91.96 seconds which is most excellent as compared to both previous feature experiments.

Table 4: Classification outcomes of proposed technique for AVAMVG gait dataset

C*	Features			Performance Metrics						
	Fused	**FEcS (70%)**	**FEcS (50%)**	**Sen (%)**	**Pre (%)**	**FDR (%)**	**FNR (%)**	**AUC**	**f_1 Score (%)**	**Acc (%)**
FT	✓			64.75	65.25	34.75	35.3	0.84	64.99	64.7
		✓		64.00	65.50	35.50	35.8	0.83	64.74	64.2
			✓	64.00	65.50	34.50	35.9	0.83	64.74	64.1
LSVM	✓			99.00	99.00	1.00	0.8	1.00	99.00	99.2
		✓		99.00	99.00	1.00	0.9	1.00	99.00	99.1
			✓	99.00	99.00	1.00	0.9	1.00	99.00	99.1
QSVM	✓			92.50	92.50	7.50	7.6	0.99	92.50	92.4
		✓		92.50	92.50	7.50	7.5	0.99	92.50	92.5
			✓	92.50	92.50	7.50	7.5	0.99	92.50	92.5
CSVM	✓			98.80	98.80	1.10	1.0	1.0	98.80	99.0
		✓		98.75	98.75	1.25	0.9	1.0	98.75	99.1
			✓	98.75	98.75	1.25	0.9	1.00	98.75	99.1
FKNN	✓			99.00	99.00	1.00	0.3	1.00	99.00	99.7
		✓		99.00	99.00	1.00	0.3	1.00	99.00	99.7
			✓	99.00	99.00	1.00	0.3	1.00	99.00	99.7
WKNN	✓			98.80	99.00	1.00	1.1	1.00	98.90	98.9
		✓		98.80	99.00	1.00	1.1	1.00	98.80	98.9
			✓	98.75	99.00	1.00	1.1	1.00	98.87	98.9
EBT	✓			97.80	98.00	2.00	2.1	1.00	97.50	97.9

		✓		98.00	98.00	2.00	2.1	1.00	98.00	97.9
			✓	97.75	98.00	2.00	1.9	1.00	97.87	98.1
RF	✓			99.0	99.00	1.00	0.2	1.00	**99.00**	**99.8**
		✓		99.0	99.00	1.00	0.2	1.00	**99.00**	**99.8**
			✓	99.0	99.00	1.00	0.2	1.00	**99.00**	**99.8**

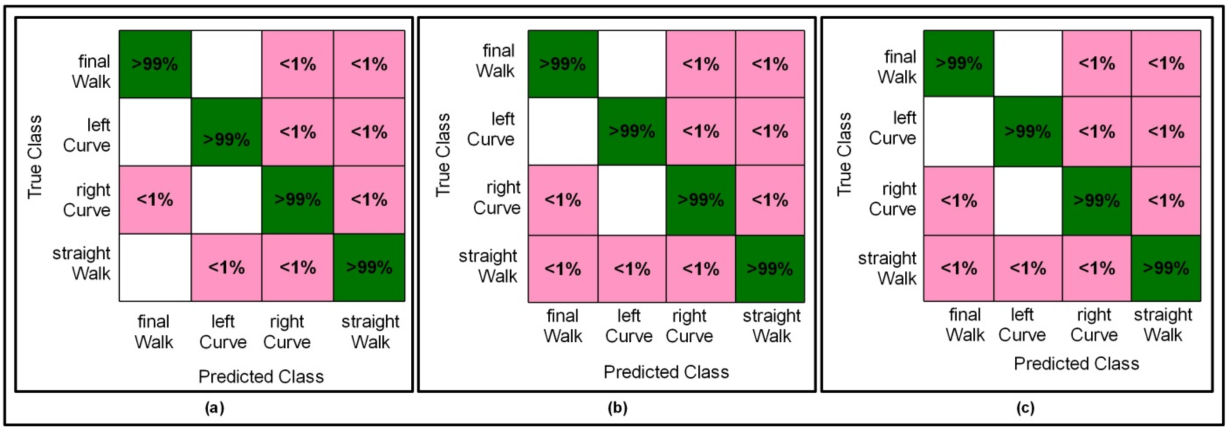

Figure 13: Confusion matrices obtained for AVAMVG dataset (a) Fused FV (b) Selection of top 70% features and
(c) Selection of top 50% features

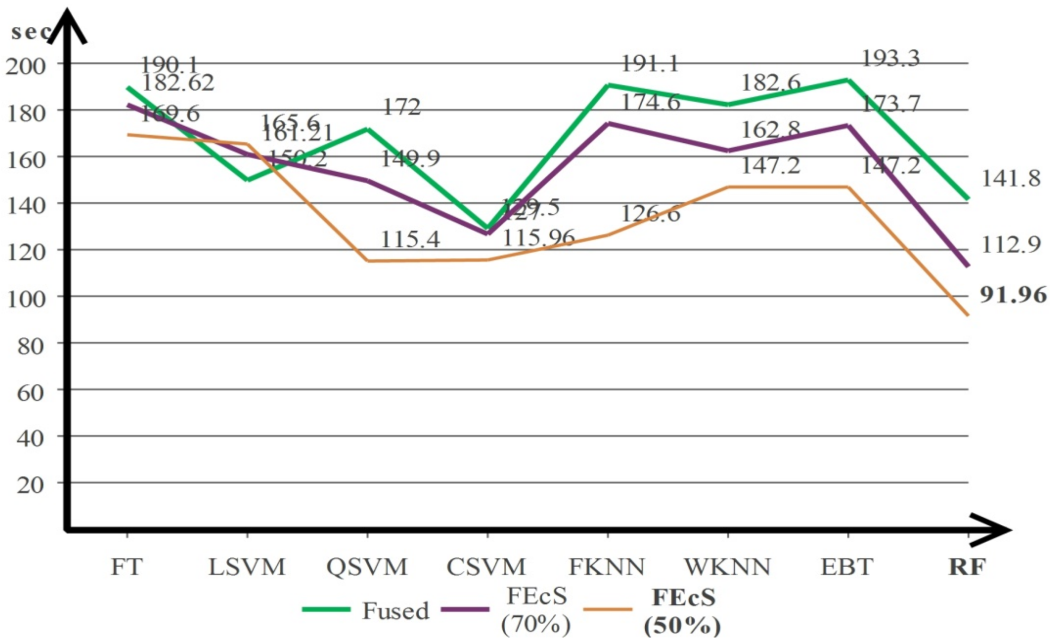

Figure 14: Classification time of proposed method for AVAMVG dataset

5.3. Discussion

Deep CNN features were extracted using two pre-trained models. In general, CNN model comprises of different layers as Convolutional, ReLu, Pooling and FC as illustrated in Figure 1. The proposed HGR flow architecture is given in Figure 2 which describes two major steps as features fusion and features selection. The experiments were performed on four datasets and outcomes are presented in Tables 1, 2, 3 and 4 and their confusion matrices are presented in Figures 7, 9, 11 and 13. Additionally, execution time of each dataset was calculated for all three features experiments (Figures 8, 10, 12 and 14). In the last, evaluation is also performed on individual features such as AlexNet and VGG19 as given in Table 5. The results indicated that the introduced system outperformed on selected datasets.

Table 5: Recognition outcomes for VGG16 and VGG19 models on selected datasets

Classifiers	CASIA A		CASIA B		CASIA C		AVAMVG	
	VGG16	VGG19	VGG16	VGG19	VGG16	VGG19	VGG16	VGG19
FT	80.4%	88.3%	54.6%	53.2%	48.8%	43.5%	62.1%	59.7%
LSVM	95.6%	98.8%	78.4%	75.5%	69.3%	64.9%	82.8%	83.0%
QSVM	**96.9%**	**99.3%**	82.6%	80.1%	81.9%	75.6%	97.3%	96.8%
CSVM	91.3%	97.0%	67.9%	65.5%	64.4%	57.8%	95.3%	94.9%
RF	96.5%	99.2%	**83.1%**	**81.3%**	**83.3%**	**76.9%**	**98.8%**	**98.5%**

A general datasets based comparison was also conducted with recent HGR techniques. A comparison of recognition accuracy rate of the proposed deep learning method for CASIA A dataset with other existing techniques is given in Table 6 by considering various variations in viewing angle such as 0°, 45° and 90°. The results indicate the superiority of the proposed approach. In Table 7, the comparison of recognition accuracy results for CASIA B dataset is given with other existing approaches by considering various cofactors under 90°. Rida et al. (Rida, Jiang, & Marcialis, 2016), Castro et al. (Castro, Marín-Jiménez, & Guil, 2016), Arora et al. (Arora, Hanmandlu, & Srivastava, 2015), Zhang et al. (Zhang, Zhao, & Xiong, 2010) and Yu et al. (Yu et al., 2017) attained recognition accuracies of 88.7%, 90.6%, 86.3%, 87.5% and 68.01% whereas the recognition results of proposed deep learning model obtained an improved performance of 93.4%, 93.2% and 93.4% using fused FV, top 70% selected features and top 50% selected features. Similarly in Table 8 and Table 9, a comparison is given for CASIA C and AVAMVG gait analysis datasets. The results indicate the dominance of the proposed technique.

Table 6: Comparison of recognition outcomes for CASIA A gait dataset

References		Recognition Rate (%)			
		0°	45°	90°	**Accuracy**
Goffredo et al. (Goffredo, Carter, & Nixon, 2008)		**100.0**	97.50	91.00	96.17
Chen and Gao (S. Chen & Gao, 2007)		92.5	85.00	65.00	80.83
Geng et al. (Geng, Wang, Li, Wu, & Smith-Miles, 2007)		90.0	95.0	90.0	91.67
Proposed	Fused	99.0	**99.0**	**100**	**99.3**
	Top 70% Features	99.0	**99.0**	**99.0**	**99.3**
	Top 50% Features	99.0	**99.0**	**100**	**99.7**

Table 7: Comparison of recognition outcomes for CASIA B gait dataset

References		Recognition Rate (%)			
		NW	**CB**	**WC**	**Accuracy**
Rida et al. (Rida et al., 2016)		98.39	75.89	91.96	88.70
Castro et al. (Castro et al., 2016)		**100.0**	**99.20**	72.60	90.60
Arora et al. (Arora et al., 2015)		**100.0**	90.0	69.0	86.30
Zhang et al. (Zhang et al., 2010)		98.39	91.94	72.18	87.50
Yu et al. (Yu et al., 2017)		95.97	65.32	42.74	68.01
Proposed	Fused	93.0	92.0	**95**	**93.40**
	Top 70% Features	93.0	93.0	**94**	**93.20**
	Top 50% Features	93.0	92.0	**95**	**93.40**

Table 8: Comparison of recognition outcomes for CASIA C gait dataset

References	Recognition Rate (%)				
	NW	**SW**	**FW**	**CB**	**Accuracy**

Kusakunniran et al. (Kusakunniran, Wu, Li, & Zhang, 2009)		**99.02**	86.39	89.56	80.72	88.92
Tan et al. (Tan, Huang, Yu, & Tan, 2007)		98.4	**91.3**	**93.7**	24.7	77.03
Marin et al. (Marín-Jiménez, Castro, Carmona-Poyato, & Guil, 2015)		96.9	67.7	79.9	79.3	80.95
Proposed	Fused	86.0	86.0	84.0	**99.0**	88.80
	Top 70% Features	91.0	87.0	91.0	**98.0**	**91.90**
	Top 50% Features	91.0	89.0	89.0	**99.0**	**92.20**

Table 9: Comparison of recognition outcomes for AVAMVG gait dataset

References		**Recognition Rate (%)**
Castro et al. (Castro, Marín-Jiménez, Mata, & Muñoz-Salinas, 2017)		95.00
Fernandez et al. (López-Fernández et al., 2014)		98.10
Castro et al. (Castro, Marín-Jimenez, & Medina-Carnicer, 2014)		95.00
Fernandez et al. (López-Fernández, Madrid-Cuevas, Carmona-Poyato, Muñoz-Salinas, & Medina-Carnicer, 2015)		98.01
Fernandez et al. (López-Fernández et al., 2016)		96.10
Proposed	Fused	**99.80**
	Top 70% Features	**99.80**
	Top 50% Features	**99.80**

6. Conclusion

HGR is an active research application of video surveillance in the last two decades in the domain of CV and ML. In this article, a new automated HGR system was proposed using two primary steps: DCNN features fusion and best features selection through FEcS approach. The results were conducted on four gait analysis datasets: CASIA A, B, C and AVAMAG and the best accuracy of 99.8%, 99.7%, 93.3% and 92.2% respectively was achieved using the RF classifier. It can be concluded from the results that the fusion of multiple CNN frameworks improved the recognition accuracy. It was also observed that the selection of best features not only enhances the system accuracy but even minimizes the execution time.

The performance of proposed approach is relied on the number of selected features but sometimes it is possible that useful features are neglected. Moreover, low resolution video sequences also affect the system accuracy, therefore in future; focus will be on the use of better resolution of video sequences and computing best features from all directions which can be recognized through reinforcement learning.

Conflicts of interest: None

References

Akram, T., Khan, M. A., Sharif, M., & Yasmin, M. (2018). Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features. *Journal of Ambient Intelligence and Humanized Computing*, 1-20.

Alotaibi, M., & Mahmood, A. (2017). Improved gait recognition based on specialized deep convolutional neural network. *Computer Vision and Image Understanding, 164*, 103-110.

Ansari, G. J., Shah, J. H., Yasmin, M., Sharif, M., & Fernandes, S. L. (2018). A novel machine learning approach for scene text extraction. *Future Generation Computer Systems, 87*, 328-340.

Arora, P., Hanmandlu, M., & Srivastava, S. (2015). Gait based authentication using gait information image features. *Pattern Recognition Letters, 68*, 336-342.

Arshad, H., Khan, M. A., Sharif, M., Yasmin, M., & Javed, M. Y. (2019). Multi-level features fusion and selection for human gait recognition: an optimized framework of Bayesian model and binomial distribution. *International Journal of Machine Learning and Cybernetics*, 1-18.

Battistone, F., & Petrosino, A. (2018). TGLSTM: A time based graph deep learning approach to gait recognition. *Pattern Recognition Letters*.

Ben, X., Zhang, P., Lai, Z., Yan, R., Zhai, X., & Meng, W. (2019). A general tensor representation framework for cross-view gait recognition. *Pattern Recognition*.

Bokhari, F., Syedia, T., Sharif, M., Yasmin, M., & Fernandes, S. L. (2018). Fundus image segmentation and feature extraction for the detection of glaucoma: A new approach. *Current Medical Imaging Reviews, 14*(1), 77-87.

Castro, F. M., Marín-Jiménez, M. J., & Guil, N. (2016). Multimodal features fusion for gait, gender and shoes recognition. *Machine Vision and Applications, 27*(8), 1213-1228.

Castro, F. M., Marín-Jiménez, M. J., Mata, N. G., & Muñoz-Salinas, R. (2017). Fisher motion descriptor for multiview gait recognition. *International Journal of Pattern Recognition and Artificial Intelligence, 31*(01), 1756002.

Castro, F. M., Marín-Jimenez, M. J., & Medina-Carnicer, R. (2014). *Pyramidal fisher motion for multiview gait recognition.* Paper presented at the Pattern Recognition (ICPR), 2014 22nd International Conference on.

Chen, Q., Wang, Y., Liu, Z., Liu, Q., & Huang, D. (2017). *Feature map pooling for cross-view gait recognition based on silhouette sequence images.* Paper presented at the Biometrics (IJCB), 2017 IEEE International Joint Conference on.

Chen, S., & Gao, Y. (2007). *An invariant appearance model for gait recognition.* Paper presented at the Multimedia and Expo, 2007 IEEE International Conference on.

Chen, X., Weng, J., Lu, W., & Xu, J. (2018). Multi-Gait Recognition Based on Attribute Discovery. *IEEE transactions on pattern analysis and machine intelligence, 40*(7), 1697-1710.

Damahe, L. B., & Thakur, N. V. (2019). Review on Image Representation Compression and Retrieval Approaches *Technological Innovations in Knowledge Management and Decision Support* (pp. 203-231): IGI Global.

Do, T. D., Nguyen, V. H., & Kim, H. (2019). Real-time and robust multiple-view gender classification using gait features in video surveillance. *Pattern Analysis and Applications*, 1-15.

Fenil, E., Manogaran, G., Vivekananda, G., Thanjaivadivel, T., Jeeva, S., & Ahilan, A. (2019). Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM. *Computer Networks, 151*, 191-200.

Fernandes, S., & Bala, J. (2013). Performance Analysis of PCA-based and LDA-based Algorithms for Face Recognition. *International Journal of Signal Processing Systems, 1*(1), 1-6.

Fernandes, S. L., & Bala, G. J. (2016a). Fusion of sparse representation and dictionary matching for identification of humans in uncontrolled environment. *Computers in biology and medicine, 76*, 215-237.

Fernandes, S. L., & Bala, G. J. (2016b). ODROID XU4 based implementation of decision level fusion approach for matching computer generated sketches. *Journal of Computational Science, 16*, 217-224.

Fernandes, S. L., & Bala, J. G. (2015). *Low power affordable and efficient face detection in the presence of various noises and blurring effects on a single-board computer.* Paper presented at the Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1.

Gadaleta, M., & Rossi, M. (2018). Idnet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognition, 74*, 25-37.

Geng, X., Wang, L., Li, M., Wu, Q., & Smith-Miles, K. (2007). *Distance-driven fusion of gait and face for human identification in video.* Paper presented at the Image and Vision Computing Conference.

Goffredo, M., Carter, J. N., & Nixon, M. S. (2008). *Front-view gait recognition.* Paper presented at the Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on.

Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., . . . Seybold, B. (2017). *CNN architectures for large-scale audio classification.* Paper presented at the 2017 ieee international conference on acoustics, speech and signal processing (icassp).

Jain, V. K., Kumar, S., & Fernandes, S. L. (2017). Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *Journal of Computational Science, 21*, 316-326.

Kamble, S. D., Thakur, N. V., & Bajaj, P. R. (2018). Fractal Coding Based Video Compression Using Weighted Finite Automata. *International Journal of Ambient Computing and Intelligence (IJACI), 9*(1), 115-133.

Khan, M., Akram, T., Sharif, M., Muhammad, N., Javed, M., & Naqvi, S. (2019). An Improved Strategy for Human Action Recognition; Experiencing a Cascaded Design. *IET Image Processing*.

Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., & Saba, T. (2018). CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features. *Computers and Electronics in Agriculture, 155*, 220-236.

Khan, M. A., Akram, T., Sharif, M., Shahzad, A., Aurangzeb, K., Alhussein, M., . . . Altamrah, A. (2018). An implementation of normal distribution based segmentation and entropy controlled features selection for skin lesion detection and classification. *BMC cancer, 18*(1), 638.

Khan, M. A., Khan, M. A., Ahmed, F., Mittal, M., Goyal, L. M., Hemanth, D. J., & Satapathy, S. C. (2020). Gastrointestinal diseases segmentation and classification based on duo-deep architectures. *Pattern Recognition Letters, 131*, 193-204.

Kosko, B. (1986). Fuzzy entropy and conditioning. *Information Sciences, 40*(2), 165-174.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks.* Paper presented at the Advances in neural information processing systems.

Kusakunniran, W., Wu, Q., Li, H., & Zhang, J. (2009). *Automatic gait recognition using weighted binary pattern on video.* Paper presented at the Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on.

L Fernandes, S., & G Bala, J. (2017). A novel decision support for composite sketch matching using fusion of probabilistic neural network and dictionary matching. *Current Medical Imaging Reviews, 13*(2), 176-184.

Lai, C., Yang, J., & Chen, Y. (2007). *A real time video processing based surveillance system for early fire and flood detection.* Paper presented at the 2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007.

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. *Applied Sciences, 7*(3), 210.

Liaqat, A., Khan, M. A., Shah, J. H., Sharif, M., Yasmin, M., & Fernandes, S. L. (2018). Automated ulcer and bleeding classification from wce images using multiple features fusion and selection. *Journal of Mechanics in Medicine and Biology, 18*(04), 1850038.

López-Fernández, D., Madrid-Cuevas, F. J., Carmona-Poyato, Á., Marín-Jiménez, M. J., & Muñoz-Salinas, R. (2014). *The AVA multi-view dataset for gait recognition.* Paper presented at the International Workshop on Activity Monitoring by Multiple Distributed Sensing.

López-Fernández, D., Madrid-Cuevas, F. J., Carmona-Poyato, A., Marín-Jiménez, M. J., Muñoz-Salinas, R., & Medina-Carnicer, R. (2016). independent gait recognition through morphological descriptions of 3D human reconstructions. *Image and Vision Computing, 48*, 1-13.

López-Fernández, D., Madrid-Cuevas, F. J., Carmona-Poyato, A., Muñoz-Salinas, R., & Medina-Carnicer, R. (2015). Entropy volumes for viewpoint-independent gait recognition. *Machine Vision and Applications, 26*(7-8), 1079-1094.

Majid, A., Khan, M. A., 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. *Microscopy research and technique*.

Marín-Jiménez, M. J., Castro, F. M., Carmona-Poyato, Á., & Guil, N. (2015). On how to improve tracklet-based gait recognition systems. *Pattern Recognition Letters, 68*, 103-110.

Martis, R. J., Gurupur, V. P., Lin, H., Islam, A., & Fernandes, S. L. (2018). Recent advances in big data analytics, internet of things and machine learning: Elsevier.

Nasir, M., Attique Khan, M., Sharif, M., Lali, I. U., Saba, T., & Iqbal, T. (2018). An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. *Microscopy research and technique, 81*(6), 528-543.

Ozen, H., Boulgouris, N. V., & Swash, R. (2017). *Gait recognition based on 3D holoscopic gait energy image.* Paper presented at the 3D Immersion (IC3D), 2017 International Conference on.

Prakash, C., Kumar, R., & Mittal, N. (2018). Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artificial Intelligence Review, 49*(1), 1-40.

Rajinikanth, V., Thanaraj, K. P., Satapathy, S. C., Fernandes, S. L., & Dey, N. (2019). Shannon's Entropy and Watershed Algorithm Based Technique to Inspect Ischemic Stroke Wound *Smart Intelligent Computing and Applications* (pp. 23-31): Springer.

Rashid, M., Khan, M. A., Sharif, M., Raza, M., Sarfraz, M. M., & Afza, F. (2018). Object detection and classification: a joint selection and fusion strategy of deep convolutional neural network and SIFT point features. *Multimedia Tools and Applications*, 1-27.

Raza, M., Sharif, M., Yasmin, M., Khan, M. A., Saba, T., & Fernandes, S. L. (2018). Appearance based pedestrians' gender recognition by employing stacked auto encoders in deep learning. *Future Generation Computer Systems, 88*, 28-39.

Rida, I., Jiang, X., & Marcialis, G. L. (2016). Human body part selection by group lasso of motion for model-free gait recognition. *IEEE Signal Processing Letters, 23*(1), 154-158.

Sahak, R., Tahir, N. M., Yassin, A. I., & Kamaruzaman, F. (2017). *Human gait recognition based on frontal view using kinect features and orthogonal least square selection.* Paper presented at the 2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT).

Shah, J. H., Chen, Z., Sharif, M., Yasmin, M., & Fernandes, S. L. (2017). A novel biomechanics-based approach for person re-identification by generating dense color sift salience features. *Journal of Mechanics in Medicine and Biology, 17*(07), 1740011.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal, 27*(3), 379-423.

Sharif, A., Khan, M. A., Javed, K., Gulfam, H., Iqbal, T., Saba, T., . . . Nisar, W. (2019). Intelligent Human Action Recognition: A Framework of Optimal Features Selection based on Euclidean Distance and Strong Correlation. *Journal of Control Engineering and Applied Informatics, 21*(3), 3-11.

Sharif, M., Akram, T., Raza, M., Saba, T., & Rehman, A. (2019). Hand-crafted and deep convolutional neural network features fusion and selection strategy: an application to intelligent human action recognition. *Applied Soft Computing*, 105986.

Sharif, M., Attique, M., Tahir, M. Z., Yasmim, M., Saba, T., & Tanik, U. J. (2020). A Machine Learning Method with Threshold Based Parallel Feature Fusion and Feature Selection for Automated Gait Recognition. *Journal of Organizational and End User Computing (JOEUC), 32*(2), 67-92.

Sharif, M., Khan, M. A., Akram, T., Javed, M. Y., Saba, T., & Rehman, A. (2017). A framework of human detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropy-based features selection. *EURASIP Journal on Image and Video Processing, 2017*(1), 89.

Sharif, M., Khan, M. A., Faisal, M., Yasmin, M., & Fernandes, S. L. (2018). A framework for offline signature verification system: Best features selection approach. *Pattern Recognition Letters*.

Sharif, M., Khan, M. A., Iqbal, Z., Azam, M. F., Lali, M. I. U., & Javed, M. Y. (2018). Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. *Computers and Electronics in Agriculture, 150*, 220-234.

Sharif, M., Khan, M. A., Zahid, F., Shah, J. H., & Akram, T. (2019). Human action recognition: a framework of statistical weighted segmentation and rank correlation-based selection. *Pattern Analysis and Applications*, 1-14.

Sharif, M., Raza, M., Shah, J. H., Yasmin, M., & Fernandes, S. L. (2019). An Overview of Biometrics Methods *Handbook of Multimedia Information Security: Techniques and Applications* (pp. 15-35): Springer.

Sharif, M., Tanvir, U., Munir, E. U., Khan, M. A., & Yasmin, M. (2018). Brain tumor segmentation and classification by improved binomial thresholding and multi-features selection. *Journal of Ambient Intelligence and Humanized Computing*, 1-20.

Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2016). *Geinet: View-invariant gait recognition using a convolutional neural network.* Paper presented at the Biometrics (ICB), 2016 International Conference on.

Siddiqui, S., Khan, M. A., Bashir, K., Sharif, M., Azam, F., & Javed, M. Y. (2018). Human action recognition: a construction of codebook by discriminative features selection approach. *International Journal of Applied Pattern Recognition, 5*(3), 206-228.

Sien, J. P. T., Lim, K. H., & Au, P.-I. (2019). *Deep Learning in Gait Recognition for Drone Surveillance System.* Paper presented at the IOP Conference Series: Materials Science and Engineering.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tan, D., Huang, K., Yu, S., & Tan, T. (2006). *Efficient night gait recognition based on template matching.* Paper presented at the Pattern Recognition, 2006. ICPR 2006. 18th International Conference on.

Tan, D., Huang, K., Yu, S., & Tan, T. (2007). *Recognizing night walkers based on one pseudoshape representation of gait.* Paper presented at the Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on.

Tang, J., Luo, J., Tjahjadi, T., & Guo, F. (2017). Robust arbitrary-view gait recognition based on 3d partial similarity matching. *IEEE Transactions on Image Processing, 26*(1), 7-22.

Vargas, R., Mosavi, A., & Ruiz, L. DEEP LEARNING: A REVIEW.

Vedaldi, A., & Lenc, K. (2015). *Matconvnet: Convolutional neural networks for matlab.* Paper presented at the Proceedings of the 23rd ACM international conference on Multimedia.

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience, 2018*.

Wang, L., Tan, T., Ning, H., & Hu, W. (2003). Silhouette analysis-based gait recognition for human identification. *IEEE transactions on pattern analysis and machine intelligence, 25*(12), 1505-1518.

Wolf, T., Babaee, M., & Rigoll, G. (2016). *Multi-view gait recognition using 3D convolutional neural networks.* Paper presented at the 2016 IEEE International Conference on Image Processing (ICIP).

Wu, C., Zhang, Y., & Song, Y. (2018). *Multi-view gait recognition using 2D-EGEI and NMF.* Paper presented at the Identity, Security, and Behavior Analysis (ISBA), 2018 IEEE 4th International Conference on.

Wu, H., Weng, J., Chen, X., & Lu, W. (2018). Feedback weight convolutional neural network for gait recognition. *Journal of Visual Communication and Image Representation, 55*, 424-432.

Xu, W., Zhu, C., & Wang, Z. (2018). Multiview max-margin subspace learning for cross-view gait recognition. *Pattern Recognition Letters, 107*, 75-82.

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.

Yu, S., Tan, D., & Tan, T. (2006). *A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition.* Paper presented at the Pattern Recognition, 2006. ICPR 2006. 18th International Conference on.

Zadeh, L. A. (1976). A fuzzy-algorithmic approach to the definition of complex or imprecise concepts *Systems Theory in the Social Sciences* (pp. 202-282): Springer.

Zhang, E., Zhao, Y., & Xiong, W. (2010). Active energy image plus 2DLPP for gait recognition. *Signal Processing, 90*(7), 2295-2302.