Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-08851-4

Offline signature verification system: a novel technique of fusion of GLCM and geometric features using SVM

Faiza Eba Batool[1] · Muhammad Attique[2] · Muhammad Sharif[1] · Kashif Javed[3] · Muhammad Nazir[2] · Aaqif Afzaal Abbasi[4] · Zeshan Iqbal[5] · Naveed Riaz[6]

Received: 4 December 2019 / Revised: 5 February 2020 / Accepted: 13 March 2020
Published online: 04 April 2020

Abstract
In the area of digital biometric systems, the handwritten signature plays a key role in the authentication of a person based on their original samples. In offline signature verification (OSV), several problems exist that are challenging for verification of authentic or forgery signature by the digital system. Correct signature verification improves the security of people, systems, and services. It is applied to uniquely identify an individual based on the motion of pen as up and down, signature speed, and shape of a loop. In this work, the multi-level features fusion and optimal features selection based automatic technique is proposed for OSV. For this purpose, twenty-two Gray Level Co-occurrences Matrix (GLCM) and eight geometric features are calculated from pre-processing signature samples. These features are fused by a new parallel approach which is based on a high-priority index feature (HPFI). A skewness-kurtosis based features selection approach is also proposed name skewness-kurtosis controlled PCA (SKcPCA) and selects the optimal features for final classification into forged and genuine signatures. MCYT, GPDS synthetic, and CEDAR datasets are utilized for validation of the proposed system and show enhancement in terms of Far and FRR as compared to existing methods.

Keywords Biometric system · Segmentation · Features extraction · Features fusion · Classification

✉ Aaqif Afzaal Abbasi
attique@ciitwah.edu.pk; aaqif.afzaal@fui.edu.pk

[1] Department of CS, COMSATS University Islamabad, Wah campus, Pakistan

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

[3] Department of Robotics, SMME Nust, Islamabad, Pakistan

[4] Department of Software Engineering, Foundation University Islamabad, Islamabad, Pakistan

[5] Department of Computer Engineering, UET Taxila, Taxila, Pakistan

[6] SEECS, National University of Science and Technology, Islamabad, Pakistan

1 Introduction

Biometric is an important application that applied in several fields such as banks [2], schools, universities, passport authentication [25], and names a few more [47–50]. These applications are extensively applied to the behavioral and physiological traits of humans like fingerprints, DNA, iris, palm, face, voice, signature, and handwriting [19]. The consistent verification of a person is essential in this digital environment and shows an important role in refining the security of individuals [28, 40]. Handwritten offline signatures are most generally applicable in the identification of banking applications like transactions, legal documents, and debit cards that are an essential part of the current economy [34]. Two main types of signature verifications are termed as online and offline [26]. The online verification system uses different smart devices such as pressure-sensitive phone, stylus or digitizer [30]. In the other type, signatures are signed on paper, bank cheques, and vouchers which are scanned and given as an input to the system for verification [11, 36].

OSV is an exciting domain due to the shortage of powerful features. Each individual has its personal signature and distinct to other people [39]. For example, signatures are little bit changes after several repetitions [37]. In addition, few other challenges exist that make the verification process difficult such as age factor, sickness, and physiological factors. This leads to the inaccuracy of the offline system [31]. Three types of forgeries take place, a) a random forgery occurs when the signer adapts his particular manner to make a forgery as he does not have knowledge about the shape of the signature and uses the name of the victim. This forgery can be detected easily b) the signer has an unclear idea of the signature but does not have much practice so he adapts his own style and forged the signature and c) skilled forgery: this sort of forgery is tricky to identify because it is forged by an expert person who has a sound understanding about the signature of the victim and also has much practice to replicate the signature [4].

Thus different OSV techniques are considered for the detection of forgeries to save the accounts and individualities to being misused by the offenders [20]. For OSV, several feature-based techniques are implemented. Among them, the GLCM [10], geometric, texture, and DCT features are mostly extracted [15, 42]. The GLCM features are mostly utilized for the extraction of texture information. In [44, 45] GLCM are extracted for texture information of an image. In [16] GLCM features are extracted from second-order statistical texture information of an input image. These extracted features are classified by supervised learning techniques like SVM and name a few more [9, 39].

1.1 Motivation

Most of the recent OSV systems are based on texture and shape information of an individual; however, the more recent the entrance of deep learning shows much enhancement in machine learning applications. The major drawback of these deep learning techniques is to require a handsome amount of training data. But in this, the selected OSV datasets are not enough in amount to train a good CNN model. For this purpose, we choose the gray level co-occurrences matrix (GLCM) and shape tendency features like geometric for OSV. The purpose of geometric points is to retrieve the shape of signature like edges etc. while GLCM gives the textural information like a change of each signature.

1.2 Problem statement and contribution

The following problems are considered in the article. a) differences that appear with the pictorial resemblance in skilled forgeries b) variances in genuine signatures and c) prominent features selection from raw data. Moreover, several individual signatures are same in shape which is another major challenge of this work. To tackle with these challenges, a new automated system is proposed. Our contributions in this article are listed below.

a) In features extraction, twenty-two GLCM features and eight geometric features are calculated. The extracted features are fused by a new parallel approach name high-priority index feature (HPIF). Through proposed HPIF approach, select the high indexed features from both GLCM and geometric vector.
b) A SKcPCA approach is also proposed in this work for optimal features selection. The PCA based features score values are computed and then individually skewness and kurtosis are calculated from fused vector. Finally, a weight function is employed which selects the top optimal features. The selected features are finally verified by SVM.
c) Experimentation is done using four classification methods on three publically available datasets. Cross validation and hold out are used. Proposed results are compared with proved existing techniques.

2 Related work

OSV is a challenging research area in the field of computer vision (CV) from last few decades due to very important applications like Banks etc. [26, 46]. Several systems have been designed for the verification and classification of signatures [14, 41]. Each method has its own importance based on the performance of different techniques. In the literature, optimum results are claimed on the basis of selected classifier, extracted features and selection choice of benchmark datasets [12]. This section briefly covers review of highly correlated and enhanced techniques. Neamah et al. [29] discussed a method for offline signature verification incorporating pattern features. The developed features are based on the fusion of direction of skeleton and gravity centre point. Gravity centre points are extracted by Centre of Gravity (COG) method. Miskhat et al. [27] described an OSV technique based on ANN and Gaussian SVM (GSVM). The implemented classifier is trained on the gradient histogram, dot density, and slices feature sets for verification. This method achieved accuracy of 96.5% which is improved as compared to previous techniques. Abbas et al. [1]. In their work, they use one-class SVM classifier based on writer-independent parameters. Features are generated through DCT and CT using only genuine signatures. They use a combination of two WI offline handwritten signature verification systems. Chandra et al. [7] introduced geometric features for offline signature verification method. For recognition and verification phase ANN model is developed to produce useful results. Paul et al. [24] presented an OSV system based on graph edit distance (GED) and deep NN. The both GED and DNN models are utilized as a features extraction and later on combine them for verification. MCYT and GPDS datasets are utilized for evaluation and lead to significant improvement. Sharif et al. [37] introduced an optimal features selection based approach for signature verification. Global and local features are extracted based on geometric information and horizontal and vertical points. Later, the genetic algorithm is performed based on MSER fitness function to selects best features. Experiments

are conducted on three datasets including CEDAR, MCYT, and GPDS and shows significant performance. Luiz et al. [13] presented a verification system using static signature samples. A spatial pyramid pooling is performed for fixed size samples and further investigates to system evaluation on low and high-resolution signature samples. The GPDS dataset is utilized for experimental process and improves the performance on high-resolution skilled forgeries as compare to existing methods.

The literature clearly explains that the texture representation of signature samples is very important for accurate recognition [35]. In addition, geometric features also play an important role due to shape, width, and height of signature. Therefore, in this work, we focused on the best texture and geometric representation for OSV.

3 Proposed work

In this section, a novel OSV method is proposed based on parallel multiple features fusion and optimal features selection. A structure flow of proposed system is shown in Fig. 1.

3.1 Preprocessing

The preprocessing increases the contrast level of input signature image that is received from different sources like publicly available datasets or by scanning signature image. The purpose is to get maximum information. This step is a conjunction of two primary sunblock's: Firstly image acquisition and secondly normalization. The detail of each sub-block is described below.

3.1.1 Image acquisition

Acquires input signature image which is further utilized for normalization, segmentation, features extraction, and classification. The grayscale image has been denoted by ϕ^I which is utilized for further process.

3.1.2 Normalization

In this step, normalize the range of intensity values of each image in order to enhance the quality of query image. Let ϕ^I is an input gray image, the 0... $\phi^L - 1$ are range of gray levels of query image, where ϕ^L denotes the gray levels and range from 0 to 255 and defined as 0... $\phi^L - 1$. Let $\phi^{P_k}(x)$ be the normalized histogram of ϕ^{f_τ}, described as follows:

$$\phi^{P_k}(x) = \frac{\phi^{n_i}}{\phi^N} \tag{1}$$

The refine image histogram ϕ^{f_τ} is computed as:

$$\phi^f(x,y) = \varphi^\eta\left(\left(\phi^L - 1\right)\sum_{k=0}^{j}\phi^{P_k}(x)\right), \tag{2}$$

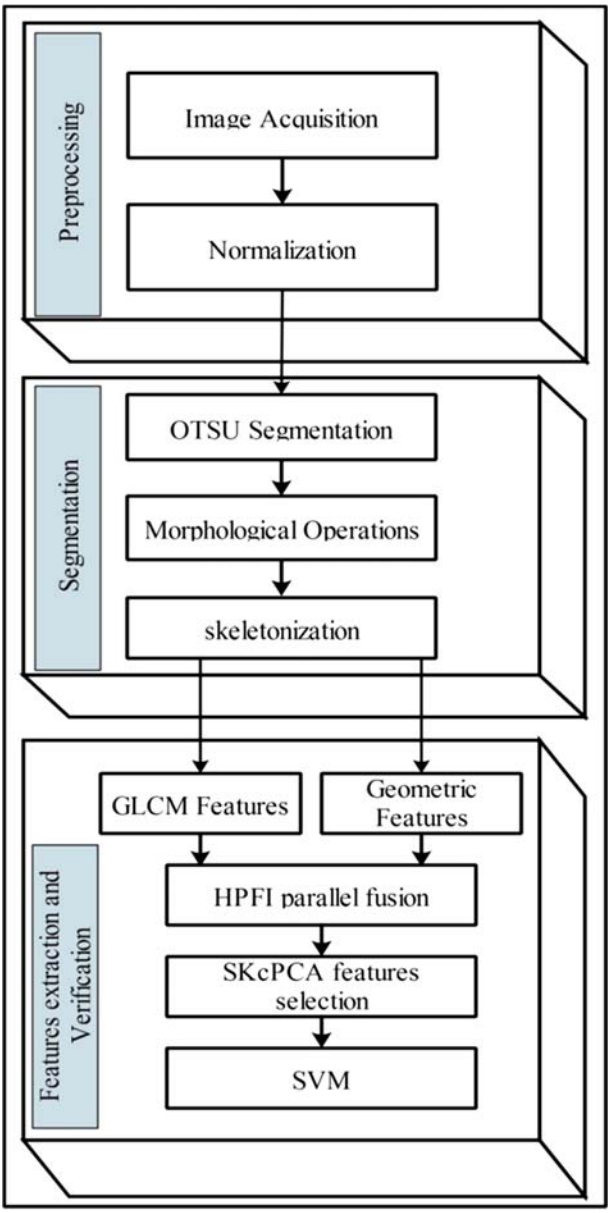

Fig. 1 The architecture of proposed offline signature verification system

Where φ^{η} is a floor function. To modify the function ϕ^{f_τ}, we consider:

$$\phi^{T_j} = \varphi^{\eta}\left(\left(\phi^{L}-1\right)\sum_{k=0}^{j}\phi^{p}(x)\right), \tag{3}$$

Then, the net yield is calculated as:

$$\widetilde{\phi^I} = \phi^{T_1}(i) = \left(\phi^L - 1\right)\int_0^i \phi^{P_i}(x)d \tag{4}$$

Where, ϕ^{P_i} is a PDF of ϕ^{f_x} and $T^{+} \sqcap 1$ is cumulative distribution function of i which is computed by multiplying with$(\phi^L - 1)$. The detailed description of histogram equalization is shown in Algorithm. 1.

After performing contrast stretching, $L^*a^*b^*$, color space [33] transformation is applied and split L^* channel. The aim of L^* channel selection is due to its maximum information in the frame as compared to other channels such as a* and b*. The selected L* channel is utilized for segmentation. The above normalization process is also presented in the form of Algorithm 1 for more deep detail for implementation.

Algorithm 1 Histogram equalization of gray channels

Step 1: For each color channel with L gray levels,

pixel intensity e k, $hist[k] = hist[k] + 1$, when i=0 to $\phi^L - 1$

Step 2: The cumulative frequency of histogram is given as $H_{cf}[k]h_{cf}[k-1] + hist[k]$

Step 3: The equalized histogram is generated as $H_{eq}[k] = \left\lfloor \frac{\phi^{L_*} h_{cf}[k] - N^2}{N^2} \right\rfloor$,

where L is number of gray levels, N are the number of pixels in X/Y direction.

$\lfloor\Theta\rfloor$ is truncation of Θ to nearest neighboor.

Step 4: For each k, replace previous values with the new mapping gray value $H_{eq}[k]$.

3.2 Image segmentation

In the segmentation step, significant regions of the image are segmented for the analysis purpose. The Otsu segmentation method is utilized with morphological operations. Later, skeletonization of image is achieved to get the one-pixel value of the input image. The brief description of Otsu segmentation is described below. The threshold function of Otsu segmentation is defined as:

$$\widetilde{\phi^T} = arg 0 \leq T_1 \leq \phi^L \ \ max\sigma_b^2(T_1), \tag{5}$$

Where, $\widetilde{\phi^T}$ denotes thresholded image. Bu this operation not prvides a better edges detection therefore, we performes morphologiccal operation such as opening, closing and filling to get

more accurate binarized image. The resultant images is further skeletonize for getting the one-pixel value of the signature and removes the variances in the width of a signature that happens due to the usage of different pens [5, 17]. It does not allow objects to break apart and remove pixels from the boundary of objects to preserve the topology of an image. To resolve this problem, perform Lantuéjoul's formula [23]. The visual affects of this step are shown in Fig. 2.

3.3 Features extraction

In OSV, the useful features are more important for better system accuracy because in many systems, the black and white image is put as an input. The each image has different shape characteristics which is a difficult process for accurate classification. In this article, GLCM [3] and Geometric features are applied for offline signature verification. The total twenty-two GLCM and eight geometric features are applied. Four metrics such as mean, range, variance and entropy are calculated of each feature. This result in the production of 1×88 features that later fused with geometric features. The process of features extraction is described below and flow is given in Fig. 3.

3.3.1 GLCM features

GLCM features are well known texture patterns of any image. It is a statistical technique for detecting texture that ponders the spatial connection of the pixels. It calculates second order statistical texture features and creates a grid in which the number of rows and columns are equivalent to the ϕ^L. GLCM describes the structure of an image by calculating frequent sets of the pixels in a predetermined spatial connection and particular qualities transpire in an image. In this work, 22 GLCM features are extracted from different angles. The co-occurrence matrix is calculated using the offset vector having angles 0°, 45°, 90°*and* 135°, by taking distance $\mathbf{d} = 1$. The typical textural measures extracted from the GLCM matrix are discussed in the Table 1. We have four directions (angles) 0°, 45°, 90°*and* 135° against each pixel to achieve the maximum count.

Four measures termed as, mean, range, entropy and variance are calculated against each feature to make them direction independent. The GLCM features are calculated as follows: Suppose $\phi^f$is any feature and $Mean \rightarrow (\phi^{\mu f})$, $Range \rightarrow \phi^{\text{Range } f}$, $Variance \rightarrow (\sigma^2)$ of ϕ^f, and entropy $\rightarrow \phi^{\text{ent}}$, then these parameters are computed as follows:

We have four angles 0°, 45°, 90°*and* 135°and their respective measures $\phi^{f(1)}$, $\phi^{f(2)}$, $\phi^{f(3)}$, $\phi^{f(4)}$ against each angle. Thus, the mean, range, and variance of ϕ^f is given as:

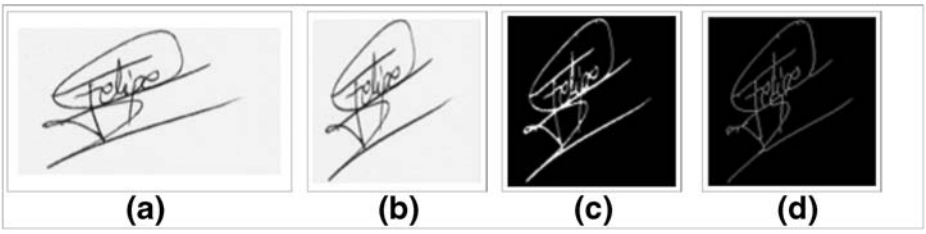

Fig. 2 Image Preprocessing and Segmentation: (**a**): Input signature image; (**b**): Resized and Normalized image; (**c**): Image Binarization through Otsu thresholding and morphological operations; (**d**): Skeletonized image

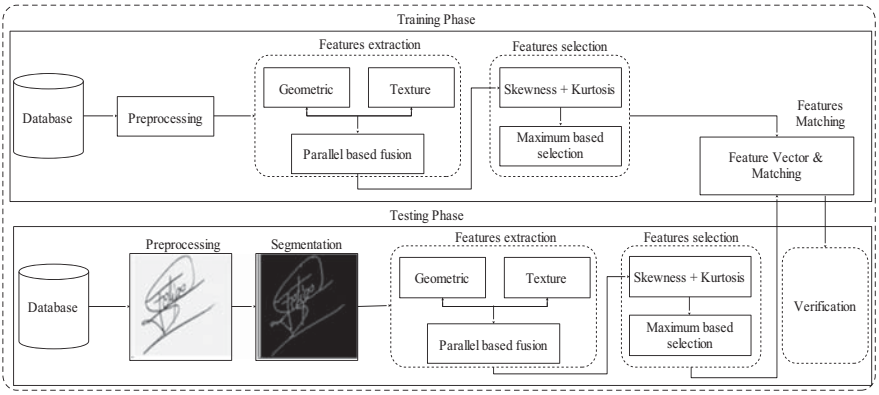

Fig. 3 Flow chart of features extraction and fusion

$$\phi^{\mu f} = \frac{1}{4}\left[\ \phi^{f(1)}\phi^{f(2)}, \phi^{f(3)}, \phi^{f(4)}\right. \tag{6}$$

$$\phi^{\text{Range } f} = \max\left(\phi^{f}\right) - \min\left(\phi^{f}\right) \tag{7}$$

$$\sigma^2 = \frac{1}{4}\left[\left[\phi^{f(1)} - \phi^{\mu f}\left(\phi^{f}\right)\right]^2 + \left[\phi^{f(2)} - \phi^{\mu f}\left(\phi^{f}\right)\right]^2 + \left[\phi^{f(3)} - \phi^{\mu f}\left(\phi^{f}\right)\right]^2 + \left[\phi^{f(4)} - \phi^{\mu f}\left(\phi^{f}\right)\right]^2 \right. \tag{8}$$

$$\phi^{\text{ent}} = \sum_k \sum_l P(k,l) \log P(k,l) \tag{9}$$

This process returns a feature vector of length 1 × 88. Moreover, the mathematical descriptions of extracted GLCM features are depicted in Table 9.

Table 1 Verification results of GPDS synthetic dataset

Dataset	Classifier				Measures		
	DT	W-KNN	EBT	SVM	FRR (%)	FAR (%)	AER (%)
Synthetic (CV = 5)	✓				**9.18**	10.5	**9.84**
		✓			11.6	10.1	10.8
			✓		10.7	10.1	10.4
				✓	10.1	**9.99**	10.0
Synthetic (CV = 10)	✓				10.1	12.8	11.4
		✓			10.1	12.4	11.2
			✓		10.2	14.7	12.4
				✓	**10.0**	**9.17**	**9.59**

Bold values shows the improved results

3.3.2 Geometric features

In this section, eight geometric features including area, orientation, solidity, perimeter, and few more are extracted. Mathematically, the listed features are computed as follows:

$$\text{Area} = \sum_{i=1}^{n}\sum_{j=1}^{m} A[i,j] \tag{10}$$

$$\textit{Major Axis Length} = x_1 + x_2, \tag{11}$$

$$\textit{Minor Axis Length} = \sqrt{(x_1 + x_2)^2 - d} \tag{12}$$

$$\textit{Perimeter} = 2l + 2w \tag{13}$$

where l describes length and w is width

$$\textit{Extent} = \frac{\textit{Area}}{\textit{Area of bounding box}} \tag{14}$$

$$\textit{Solidity} = \frac{\textit{Area}}{\textit{Convex Area}} \tag{15}$$

where x_1, x_2 are distance points of ellipse and $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ denotes the distance between $x_1 and\ x_2$. Hence, the resultant feature vector is having dimension 1×8 that are later fused by proposed parallel selection method.

3.4 Features fusion

Features fusion mean group the patterns of multilple features in one matrix which gives good accuracy as compare to individual descriptors [51–53]. In this article, a parallel approach is implemented name HPFI. In this technique, initially, compute the variance of extracted geometric features and perform padding to make the equal length of both vectors. After making the equal lengt of both vectors, the HPFI approach is applied in which high priority index features are selected for new fused vector. Methamatically, fusion is performed as follows:

Let $\phi(GL)$ and $\phi(GM)$ be two extracted set of features as shown in Fig. 4. The t_1, t_2 denotes the dimensionality of extracted features which are 1×88 and 1×8, respectively. The given signature dataset is $D \in \Omega$ with their corresponding feature vectors X and Y, where, $X \in \phi(\text{GL})$ represnted as GLCM features and $Y \in \phi(\text{GM})$ reprersnted as geometric features, respectively. Then, parallel fusion is performed in following steps.

1) In the first step, check the length of both vectors X and Y. If the length of both vectors are not equal o each other, then find out the maximum length vector and make equal length of minimum dimensional vector based on variance padding. In variance padding, initially, variance is calculated from minimum length vector and computed variance value is utilized as padding to make the equal length.

2) After performing variance padding, fusion is performed based on proposed HPFI approach. In HPFI pproach, maximum indexed feature is selected for new fused vector. The selection of highst value is feature is formulated as follows:

Let $X1$ and $Y1$ are feature vector of GLCM and Geometric of equal length $N \times 88$, where N denotes the number of images that utilized for features extraction. Then, matching the each *ith* feature index of $X1$ vector with *jth* feature index of $X2$ vector. The higher value index feature is selected and placed in new vector $X3$. The $X3(k)$ denotes the fused vector of dimension $N \times 88$. The higher value index is selected by following formulation.

$$C_1 = \{X1 + iY1 | X1 \in \phi(GL), Y1 \in \phi(GM)\} \tag{16}$$

Here C_1 denotes the length of fused feature vector which is $N \times 88$. Then the inner product is the calculation of complex vector as:

$(g, h) = g^T h$, where $(g, h) \in C_1$ and T is conjugate transpose of complex vector. Hence, the norm of complex vector and its distance is calculated as:

$$Norm = \|Z\| = \sqrt{\sum_{j=1}^{n} \left({g_j}^2 + {h_j}^2 \right)} \tag{17}$$

$$Distance = \sqrt{(z_1 - z_2)^T (z_1 - z_2)} \tag{18}$$

Where, Z= $(+ih_1, \ldots g_j + ih_j)$. Finally, a threshold function is defined below which select the high index features for fused vector.

$$F(C_1) = \begin{cases} Selected\ ith & if\ \ X1(i) \geq X2(j) \\ Selected\ jth & if\ \ X2(j) > X1(i) \end{cases} \tag{19}$$

The above funation explains that if *ith* feature of index $X1$ is greater or equal to *jth* index of $X2$ feature vector, then *ith* index is selected and put into fused vector $X3(k)$, else *jth* feature is selected. Hence, the dimension of resultant vector is $N \times 88$ which is later utilized in prominent features selection phase.

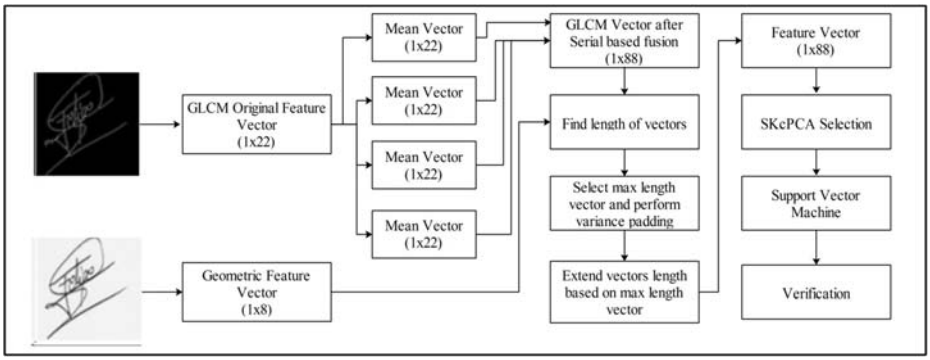

Fig. 4 Graphical representation of features fusion

3.5 Features selection

In this section, the most prominent features from the extracted features are selected. Feature selction is important because all the extracted features are not necessary for final classification [54, 55]. In this article, this problem is resolved by the implementation of a novel method which is based on SKcPCA. Initilally, PCA is peformed and compute the principle components for all features. Then, skewness and kurtosis features are calculated for each image seperatly and sort into ascending order. The major purpose of sorting into ascending order is to select the best features based on their mean value. The mean value is calcuated for both skweness and kurtosis vectors and fed into a threshold function for final selecion. Mathematically, formulation is defined as follows:

$$\phi(SK) = E[\![\frac{(x-\mu)}{\sigma}]\!]^3, x \in X3(k) \tag{20}$$

$$= \frac{E(x^3) - 3\ \mu\ E(x^2) - \mu E(x) - \mu^3}{\sigma^3} \tag{21}$$

$$\phi(SK) = \frac{E(x^3) - 3\ \mu\sigma^2 - \mu^3}{\sigma^3} \tag{22}$$

$$\phi(Kur) = E\left(\frac{x-\mu}{\sigma}\right)^4 \tag{23}$$

$$= n\ \frac{\sum_{i=1}^{n}(x_i-\mu)^4}{\left(\sum_{i=1}^{n}(x_i-\mu)^2\right)^2} \tag{24}$$

$$\phi(Kur) = \frac{\mu^4}{(\sigma^2)^2} \tag{25}$$

Where, $\phi(SK)$ denotes the skewness vector and $\phi(Kur)$ denotes the kurtosis vector of same length as fused vector $X3(k)$. By mean value μ, defined a threshold function which is used for best features selection. The threshold function is defined below:

$$F(sl_i) = \begin{cases} S & if\ \ \phi(SK)(i) \geq \mu 1 \\ N & Elsewhere \end{cases} \tag{26}$$

Table 2 GPDSsynthetic dataset error comparison with existing methods

Database	Reference	FAR(%)	FRR(%)
GPDSsynthetic	Solemani et al. [35]	29.49	5.80
GPDSsynthetic	Zhang et al. [44]	17.35	18.74
GPDSsynthetic	**Proposed**	**9.17**	10.0

Bold values shows the improved results

$$F(sl_j) = \begin{cases} S & if \quad \phi(Kur)(i) \geq \mu 2 \\ N & Elsewhere \end{cases} \tag{27}$$

Where, $\mu 1$ and $\mu 2$ denotes the mean value of skewness and kurtosis, respectively. $F(sl_i)$ and $F(sl_j)$ denotes the skewness and kurtosis vectors, respectively of length between $N \times 53$ *to* $N \times 72$. Finally, fused these features based on serial method and obtianed a new vector which is later fed to SVM for verification. The redial basis (RBF) kernel function is employed in this work where the the method is one vs one and box constraint is one. Mathematically, it is formulated as follows:

$$min \frac{1}{2} \; w^T w + \delta \sum_{i=1}^{\mathcal{K}} max\{1 - \tau^i w^T \rho^i, 0\} \tag{28}$$

$$min_\gamma \quad \zeta(\gamma) = \frac{1}{2} \; \gamma^T \; \upsilon\gamma - e^T \gamma \tag{29}$$

$$s.t \quad 0 \leq \gamma \leq \delta \; and \; i = 1, 2, 3, \ldots, \mathcal{K} \tag{30}$$

where $\delta > 0$ is called penalty parameter and $Q^{i,\,j} = \tau^i \tau^j \rho^i \rho^j$ and $e \in R^{\mathcal{K}}$ is a vector of ones, recpectively.

4 Experimental results

In this section, an OSV method is employed to verify the forged (F) and genuine (G) signatures. To reveal the algorithm's results with other related methods based on three publically available benchmark datasets MCYT [32], GPDSsynthetic [8], and CEDAR [18] respectively. K-Fold cross validation is performed on MCYT, GPDSsynthetic, and CEDAR to testify the accuracy of proposed algorithm in terms of False Acceptance Rate (FAR), FRR, and Average Error Rate (AER), respectively where K = 5 and 10.

4.1 GPDS synthetic signature database

GPDSsynthetic [8] is a computer based database in which automatically generate signatures of new persons. This datasets contains total of 4000 set of images where each set contains 24 G and 30 F signatures. For experiments, we randomly selected 200 sets. From 200 sets, we have been utilized 150 sets for trainig the classifier and 50 sets

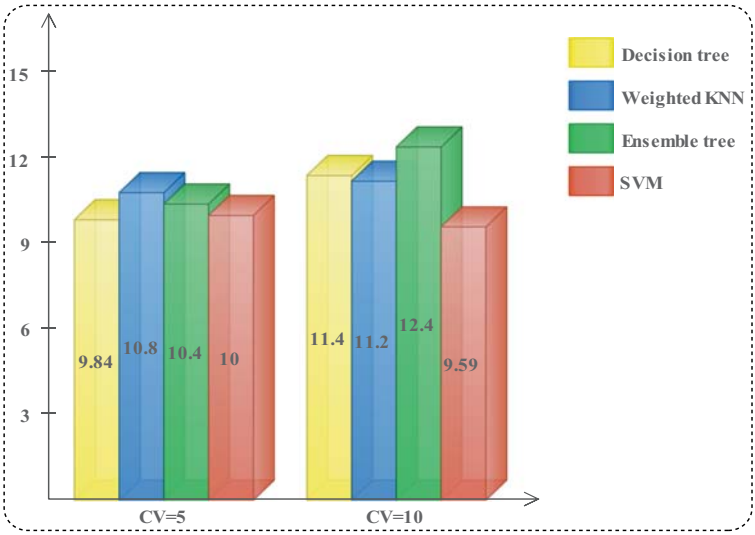

Fig. 5 Average results using CV = 5 and 10 on GPDS synthetic dataset

for testing. Results are compared with three classifier named Decision Trees (DT), Ensemble Boosted Tree (EBT) and Weighted K- Neareast Neighbour (W-KNN) [37]. The error rate of their outputs in term of quantitative values are expressed in Table 1. From Table 1, FRR is 10.0 and FAR is 9.17 by 10-fold cross validation (10FCV). The comparison of error rate is shown in Table 2, which shows that proposed method perform signaificantly good as compare to existing methods. Also, the average comparison of classifiers in terms of CV = 5 and CV = 10 are shown in Fig. 5.

4.2 MCYT signature database

This dataset contaisn total of 25 set of signatures [32]. The each set includes 15 G and 15 F signature samples of resolution 600 dpi. For experiments, the proportion of 75:25 is adopted for training and testing along 10FCV. The proposed fused features are tested by utilizing SVM

Table 3 Verification results of MCYT dataset

Dataset	Classifier				Measures		
	DT	W-KNN	EBT	SVM	FRR (%)	FAR (%)	AER (%)
MCYT (CV = 5)	✓				6.67	2.53	4.67
		✓			**2.67**	4.00	3.34
			✓		6.00	**2.00**	4.00
				✓	**2.67**	3.25	**2.96**
MCYT (CV = 10)	✓				4.67	4.00	4.34
		✓			4.01	2.67	3.34
			✓		3.99	**1.34**	2.67
				✓	**2.00**	2.66	**2.33**

Bold values shows the improved results

Table 4 Comparison of proposed results with existing method using MCYT dataset

Method	FAR(%)	FRR(%)
Zois et al. [45]	17.21	4.96
Soleimani et al. [38]	13.16	12.53
Vargas et al. [43]	6.77	8.59
Radika et al. [23]	10.0	7.0
Proposed Method	**2.66**	**2.00**

Bold values shows the improved results

and the error rate of their outputs in term of quantitative values are expressed in Table 3. Table 4 demonstrate the error rate, computed in terms of FAR and FRR to be 2.66 and 2.00, respectively. It is also evident from Table 4, that proposed strategy outperforms the previous state-of-the-arts methods, also results are plotted in Fig. 6.

4.3 CEDAR database

CEDAR [18] benchmark dataset includes total of 55 signature samples sets of 55 signers relating to educational experiences. This dataset comprises of total 1320 G and 1320 F signatures. Each set incorporates 20 G signatures, taken 20 min apart. The 3 F signatures are simulated, 8 times each, to produce 24 forged signatures. The signature images are in both gray and binary forms. For experiments, 35:20 is adopted as training and testing. The proposed fused features are tested by utilizing SVM and the error rate of their outputs in term of quantitative values are expressed in Table 5.

Table 6 demonstrate the error rate, computed in terms of FAR and FRR to be 3.34 and 3.75, respectively. It is also evident from Table 6, that proposed strategy outperforms the previous state-of-the-arts methods, also results are plotted in Fig. 7.

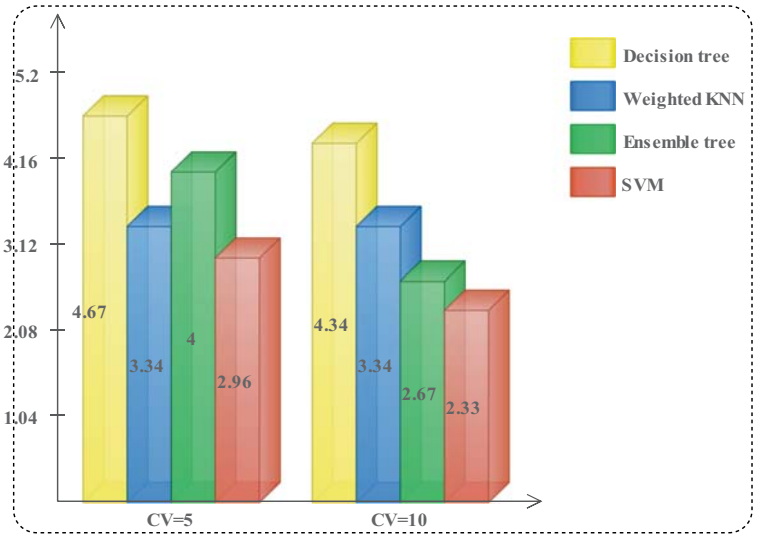

Fig. 6 Average results using CV = 5 and 10 on MCYT dataset

Table 5 Verification results of CEDAR dataset

Dataset	Classifier				Measures		
	DT	W-KNN	EBT	SVM	FRR (%)	FAR (%)	AER (%)
CEDAR (CV = 5)	✓				5.42	5.02	5.22
		✓			6.25	**3.34**	4.79
			✓		3.76	4.59	4.18
				✓	**5.0**	**3.34**	**4.17**
CEDAR (CV = 10)	✓				5.42	4.17	4.79
		✓			5.47	4.77	5.12
			✓		5.41	3.77	4.59
				✓	**3.75**	**3.34**	**3.55**

Bold values shows the improved results

5 Discussion

The depth analysis of proposed OSV system is presented in this section in terms of Far, FRR, and AER. The proposed algorithm is a unification of two major phases: First, features extraction and fusion, secondly, verification of forged and genuine signatures as shown in Fig. 1 and Fig. 3. We constructed a fused model of GLCM and geometric features that later improves by feature selection technique, results are shown in Tables 1, 3, and 5. Error comparison of fused features' sets for three benchmark datasets is provided in Tables 2, 4, and 6. It is apparent, that proposed approach achieves minimum FAR to be 9.17, 2.66, and 3.34 and FRR 10.0, 2.0, and 3.75, for GPDSsynthetic, MCYT and CEDAR databases, respectively. The average results on selected classifiers on CV = 5 and CV = 10 are shown in Figs. 5, 6, and 7. Also, the results are taken on hold-out ($p = 0.5$) as depicted in Table 7.

Table 2 shows FAR and FRR comparison of proposed method with Solemainia et al. [38] and Zhang et al. [44], where with current framework, achieved FAR is 9.17 which is comparable but FRR is calculated to be 10.0 which is lesser then [35]. The reason of high FRR is due to analysis with a higher number of training samples. So overall, our system aoutperforms as compare to existing ones. Further, we check the change in verfication accuracy after change the nature of input as results can be seen in Table 8. In this table, it is clearly seen that the proposed method performs well for binary images.

From the results on selected datasets, it is shown that the proposed method outperforms. The strengths of the proposed method are: (a) gives improved performance based on the HPFI fusion performance, and (b) selection of prominent features improves the accuracy and makes the system more efficient. The drawbacks of the proposed method are: (a) fusion of features

Table 6 CEDAR dataset error comparison with existing methods

System	FAR(%)	FRR(%)
Zois et al. [45]	11.52	5.83
Kumar et al. [21]	5.68	6.36
R.Kumar et al. [22]	8.33	8.33
Proposed	**3.34**	**3.75**

Bold values shows the improved results

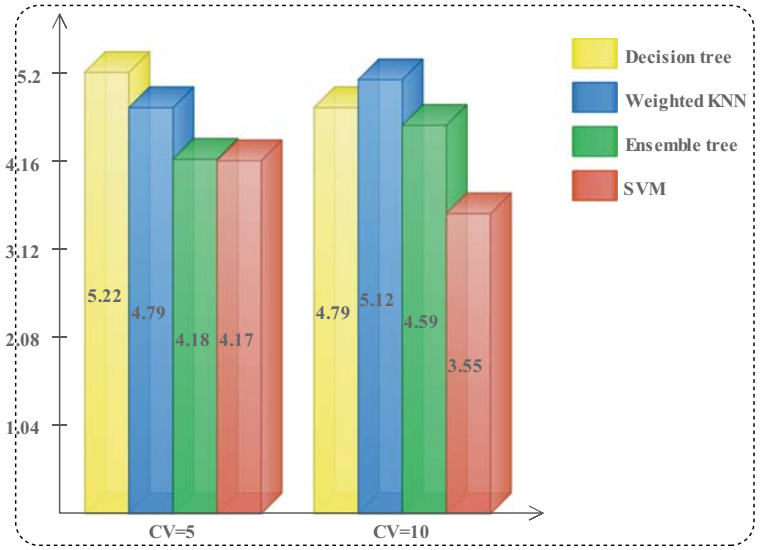

Fig. 7 Average results using CV = 5 and CV = 10 on CEDAR dataset

affects the system efficiency; (b) the process of preprocessing increases system complexity, and (c) the selection process ignore few features but it is possible that these features are essential for best accuracy on newly datasets.

6 Conclusion

A new approach is presented in this work for OSV based on HPFI fusion and SKcPCA based select best features. Two types of features such as GLCM and geometric are computed and fused by a proposed parallel HPFI approach. Later, SKcPCA based best features are selected for final verification. The SKcPCA approach is based on the PC for extracted features and their

Table 7 Proposed method results on four classifiers using hold out ($p = 0.5$) on selected dataset

Dataset	Classifier				Measures		
	DT	W-KNN	EBT	SVM	FRR (%)	FAR (%)	AER (%)
Synthetic (p = 0.5)	✓				9.69	10.3	10.0
		✓			9.67	10.0	9.8
			✓		**9.17**	8.98	9.08
				✓	9.63	**7.08**	**8.36**
MYT (p = 0.5)	✓				6.80	**0.06**	3.73
		✓			3.75	2.58	3.17
			✓		5.54	5.0	5.27
				✓	**1.25**	5.36	**3.31**
CEDAR (p = 0.5)	✓				**0.83**	7.92	4.38
		✓			3.33	3.33	3.33
			✓		2.50	1.67	2.09
				✓	2.1	**1.60**	**1.85**

Bold values shows the improved results

Table 8 Comparison of AER among original and binary nature of input for verification, where the experiment is conducted on Synthetic dataset of CV is 5

Method	Input Nature		AER (%)
	Original	Binary	
DT	✓		13.69
		✓	**9.84**
WKNN	✓		14.51
		✓	**10.80**
EBT	✓		14.97
		✓	**10.40**
SVM	✓		14.32
		✓	**10.00**

Bold values shows the improved results

mean value. Three publically available datasets including MCYT, GPDS synthetic, and CEDAR are utilized and achieved FAR 2.66%, 9.17%, and 3.34%, respectively. The results reveal that the proposed system outperforms on selected datasets. Also, the results are calculated on 5 fold and 10 fold and conclude that the proposed system gives a significant performance on 10 folds. In the future, increase the accuracy rate, a deep learning-based model will implement for OSV. The deep learning models required a huge amount of data to train a model [6]. Besides, an optimization algorithm name Grasshopper will be considered for best features selection as compared to heuristic techniques.

Appendix

Table 9 Extraction of twenty-two GLCM features

Feature Name	Equation
Auto Correlation	$\widetilde{\phi^{R}} = \sum_k \sum_l (k \times l) P(k,l)$
Contrast	$\widetilde{\phi^{C}} = \sum_{k=1}^{\phi^{\Gamma}} \sum_{l=1}^{\phi^{\Gamma}} \lvert k-l \rvert^2 P(k,l)$
Correlation 1	$\widetilde{\phi^{R_1}} = \sum_{k=0}^{\phi^{\Gamma}-1} \sum_{l=0}^{\phi^{\Gamma}-1} (k \times l) P(k,l) - \phi^{\mu_x} \phi^{\mu_y}$
Correlation 2	$\widetilde{\phi^{R_2}} = \frac{\sum_{k=0}^{\phi^{\Gamma}-1} \sum_{l=0}^{\phi^{\Gamma}-1} (k-\phi^{\mu_k})(l-\phi^{\mu_l})\ P(k,l)}{\sigma}$
Cluster Prominence	$\widetilde{\phi^{\mathbb{P}}} = \sum_{k=0}^{\phi^{\Gamma}-1} \sum_{l=0}^{\phi^{\Gamma}-1} \{k + l - \phi^{\mu_x} - \phi^{\mu_y}\}^4 P(k,l)$
Cluster Shade	$\widetilde{\phi^{S}} = \sum_{k=0}^{\phi^{\Gamma}-1} \sum_{l=0}^{\phi^{\Gamma}-1} \{k + l - \phi^{\mu_x} - \phi^{\mu_y}\}^3 P(k,l)$
Dissimilarity	$\widetilde{\phi^{D}} = \sum_k \sum_l P(k,l) \lvert k-l \rvert$
Energy	$\widetilde{\phi^{\mathbb{E}}} = \sum_k \sum_l P(k,l)^2$
Entropy	$\widetilde{\phi^{H}} = \sum_k \sum_l P(k,l) \log P(k,l)$
Homogeneity 1	$\widetilde{\phi^{\alpha_1}} = \frac{\sum_k^{\phi^{\Gamma}-1} \sum_l^{\phi^{\Gamma}-1} P(k,l)}{1+\lvert k-l \rvert}$

Table 9 (continued)

Feature Name	Equation
Homogeneity 2	$\widetilde{\phi^{\alpha_2}} = \frac{\sum_k^{\phi^{\Gamma}-1} \sum_l^{\phi^{\Gamma}-1} P(k,l)}{1+(k-l)^2}$
Maximum Probability	$\widetilde{\phi^{P}} = max_{k,l}\, P(k,l)$
Sum of Squares (Variance)	$\widetilde{\phi^{\Sigma\sigma^2}} = \sum_{k=0}^{\phi^{\Gamma}-1} \sum_{l=0}^{\phi^{\Gamma}-1} (k-\phi^{\mu})^2\, P(k,l)$
Sum Average	$\widetilde{\phi^{\Sigma\mathbb{A}}} = \sum_{k=2}^{2\phi^{\Gamma}-2} kP_{x+y}(k)$
Sum Entropy	$\widetilde{\phi^{\Sigma H}} = -\sum_{k=2}^{2\phi^{\Gamma}-2} P_{x+y}(k)\log(P_{k+l}(k))$
Sum Variance	$\widetilde{\phi^{\Sigma\sigma^2}} = \sum_{k=2}^{2\phi^{\Gamma}-2} \left(k-\widetilde{\phi^{H}}\right) P_{x+y}(k)$
Difference Variance	$\widetilde{\phi^{\overline{\sigma}^2}} = \sigma^2\left(P_{x-y}\right)$
Difference Entropy	$\widetilde{\phi^{H}} = -\sum_{k=0}^{\phi^{\Gamma}-1} P_{k-l}(k)\log(P_{k-l}(k))$
Information Measure of Correlation 1	$\widetilde{\phi^{MR_1}} = \frac{\widetilde{\phi^{H}}-H_{xy1}}{max\left(H_x,H_y\right)}$
Information Measure of Correlation 2	$\widetilde{\phi^{MR_2}} = \sqrt{\left(1-exp\left[-2.0\left(H_{xy2}-\widetilde{\phi^{H}}\right)\right]\right)}$
Inverse Difference Normalized	$\widetilde{\phi^{D^{-1}}} = \sum_k \sum_l \frac{P(k,l)}{1+\frac{\lvert k-l\rvert}{\phi^{\Gamma}}}$
Inverse Difference Moment Normalized	$\widetilde{\phi^{DM^{-1}}} = \sum_k \sum_l \frac{P(k,l)}{1+\frac{(k-l)^2}{\phi^{\Gamma}}}$

References

1. Abbas N, Chibani Y, Hadjadji B, Omar ZA, and Smarandache F (2016) A DSmT based system for writer-independent handwritten signature verification, in Information Fusion (FUSION), 2016 19th International Conference on, pp. 2213–2220.
2. Alajrami E, Ashqar BAM, Abu-Nasser BS, Khalil AJ, Musleh MM, Barhoom AM, Abu-Naser SS (2020) Handwritten signature verification using deep learning. International Journal of Academic Multidisciplinary Research (IJAMR) 3(12):39–44
3. Albregtsen F (2008) Statistical texture measures computed from gray level coocurrence matrices, *Image processing laboratory, department of informatics, university of oslo,* vol. 5
4. Bhunia AK, Alaei A, Roy PP (2019) Signature verification approach using fusion of hybrid texture features. Neural Comput & Applic 31:8737–8748
5. Biswas S, Kim T-h, Bhattacharyya D (2010) Features extraction and verification of signature image using clustering technique. Int J Smart Home 4:43–55
6. Calik N, Kurban OC, Yilmaz AR, Yildirim T, Ata LD (2019) Large-scale offline signature recognition via deep neural networks and feature embedding. Neurocomputing 359:1–14
7. Chandra S and Maheskar S (2016) Offline signature verification based on geometric feature extraction using artificial neural network, in Recent Advances in Information Technology (RAIT), 2016 3rd International Conference on, pp. 410–414.
8. Ferrer MA, Diaz-Cabrera M, and Morales A (2013) Synthetic off-line signature image generation, in Biometrics (ICB), 2013 International Conference on,pp. 1–7.
9. Ghanim TM and Nabil AM (2018) Offline signature verification and forgery detection approach, in 2018 13th International Conference on Computer Engineering and Systems (ICCES), pp. 293–298.

10. Gyimah K, Appati JK, Darkwah K, Ansah K (2019) An improved geo-textural based feature extraction vector for offline signature verification. Journal of Advances in Mathematics and Computer Science 32:1–14
11. Hadjadj I, Gattal A, Djeddi C, Ayad M, Siddiqi I, and Abass F 2019 Offline signature verification using textural descriptors, in Iberian Conference on Pattern Recognition and Image Analysis, pp. 177–188.
12. Hafemann LG, Sabourin R, Oliveira LS (2017) Learning features for offline handwritten signature verification using deep convolutional neural networks. Pattern Recogn 70:163–176
13. Hafemann LG, Oliveira LS, Sabourin R (2018) Fixed-sized representation learning from offline handwritten signatures of different sizes. International Journal on Document Analysis and Recognition (IJDAR) (21) 3: 219–232
14. Hafemann LG, Sabourin R, Oliveira LS (2019) Meta-learning for fast classifier adaptation to new users of signature verification systems. IEEE Trans Inf Forensics and Secur
15. Jadhav T (2019) Handwritten signature verification using local binary pattern features and KNN. Int Res J Eng Technol (IRJET) 6:579–586
16. Jain C, Singh P, Rana A (2017) Fuzzy logic based adaptive resonance Theory-1 approach for offline signature verification. Image Process Commun 22:23–30
17. Jan Z, Muhammad H, Rafiq M, Zada N (2015) An automated system for offline signature verification and identification using Delaunay triangulation. In: New contributions in information systems and technologies. Springer, pp 653–663
18. Kalera MK, Srihari S, Xu A (2004) Offline signature verification and identification using distance statistics. Int J Pattern Recognit Artif Intell 18:1339–1360
19. Kayaoglu M, Topcu B, and Uludag U (2015) Biometric matching and fusion system for fingerprints from non-distal phalanges, *arXiv preprint arXiv:1505.04028*.
20. Kennard DJ, Barrett WA, and Sederberg TW (2012) Offline signature verification and forgery detection using a 2-d geometric warping approach, in Pattern Recognition (ICPR), 2012 21st International Conference on, pp. 3733–3736.
21. Kumar MM, Puhan NB (2014) Off-line signature verification: upper and lower envelope shape analysis using chord moments. IET Biometrics 3:347–354
22. Kumar R, Sharma J, Chanda B (2012) Writer-independent off-line signature verification using surroundedness feature. Pattern Recogn Lett 33:301–308
23. Lantuéjoul C (1977) Sur le modèle de Johnson-Mehl généralisé, Intern Rep Cent Morph Math
24. Maergner P, Pondenkandath V, Alberti M, Liwicki M, Riesen K, Ingold R, et al.) (2018) Offline signature verification by combining graph edit distance and triplet networks, in Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR pp. 470–480.
25. Maergner P, Howe NR, Riesen K, Ingold R, and Fischer A (2019) Graph-based offline signature verification, *arXiv preprint arXiv:1906.10401*
26. Masoudnia S, Mersa O, Araabi BN, Vahabie A-H, Sadeghi MA, Ahmadabadi MN (2019) Multi-representational learning for offline signature verification using multi-loss snapshot ensemble of CNNs. Expert Syst Appl 133:317–330
27. Miskhat SF, Ridwan M, Chowdhury E, Rahman S, and Amin MA (2012) Profound impact of artificial neural networks and Gaussian SVM kernel on distinctive feature set for offline signature verification, in Informatics, Electronics & Vision (ICIEV), 2012 International Conference on, pp. 940–945.
28. Morales A, Ferrer MA, Cappelli R, Maltoni D, Fierrez J, Ortega-Garcia J (2015) Synthesis of large scale hand-shape databases for biometric applications. Pattern Recogn Lett 68:183–189
29. Neamah K, Mohamad D, Saba T, Rehman A (2014) Discriminative features mining for offline handwritten signature verification. 3D Res 5:2
30. Nguyen V and Blumenstein M (2011) An application of the 2d gaussian filter for enhancing feature extraction in off-line signature verification, in Document Analysis and Recognition (ICDAR), 2011 International Conference on, pp. 339–343.
31. Okawa M (2018) Synergy of foreground–background images for feature extraction: offline signature verification using fisher vector with fused KAZE features. Pattern Recogn 79:480–489
32. Ortega-Garcia J, Fierrez-Aguilar J, Simon D, Gonzalez J, Faundez-Zanuy M, Espinosa V et al (2003) MCYT baseline corpus: a bimodal biometric database. IEE Proceedings-Vision, Image and Signal Processing 150:395–401
33. Plataniotis K, Venetsanopoulos AN (2013) Color image processing and applications. Springer Science & Business Media, Berlin
34. Radhika K, Gopika S (2015) Online and offline signature verification: a combined approach. Procedia Comput Sci 46:1593–1600

35. Ruiz V, Linares I, Sanchez A, Velez JF (2020) Off-line handwritten signature verification using compositional synthetic generation of signatures and Siamese neural networks. Neurocomputing 374:30–41
36. Samuel D, Samuel I (2010) Novel feature extraction technique for off-line signature verification system. Int J Eng Sci Technol 2:3137–3143
37. 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
38. Soleimani A, Araabi BN, Fouladi K (2016) Deep multitask metric learning for offline signature verification. Pattern Recogn Lett 80:84–90
39. Sriwathsan W, Ramanan M, Weerasinghe A (2020) Offline handwritten signature recognition based on SIFT and SURF features using SVMs. Asian Res J Math:84–91
40. Stauffer M, Maergner P, Fischer A, and Riesen K (2019) Graph embedding for offline handwritten signature verification, in Proceedings of the 2019 3rd International Conference on Biometric Engineering and Applications, pp. 69–76.
41. Taşkiran M and Çam ZG, (2017) Offline signature identification via HOG features and artificial neural networks, in 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 000083–000086.
42. Thakare BS and Deshmukh HR (2018) A combined feature extraction model using SIFT and LBP for offline signature verification system, in 2018 3rd International Conference for Convergence in Technology (I2CT), pp. 1–7.
43. Vargas JF, Ferrer MA, Travieso C, Alonso JB (2011) Off-line signature verification based on grey level information using texture features. Pattern Recogn 44:375–385
44. Zhang Z, Liu X, and Cui Y (2016) Multi-phase offline signature verification system using deep convolutional generative adversarial networks, in Computational Intelligence and Design (ISCID), 2016 9th International Symposium on,pp. 103–107.
45. Zois EN, Alewijnse L, Economou G (2016) Offline signature verification and quality characterization using poset-oriented grid features. Pattern Recogn 54:162–177
46. Zois EN, Tsourounis D, Theodorakopoulos I, Kesidis AL, Economou G (2019) A comprehensive study of sparse representation techniques for offline signature verification. IEEE Transactions on Biometrics, Behavior, and Identity Science 1:68–81
47. Habiba A, Khan MA, Sharif MI, Yasmin M, Tavares JMRS, Zhang Y, and Satapathy SC. "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition." Expert Systems: e12541.
48. Khan MA, Javed K, Khan SA, Saba T, Habib U, Khan JA, and Abbasi AA (2020) "Human action recognition using fusion of multiview and deep features: an application to video surveillance." Multimedia Tools and Applications 1-27.
49. Muhammad S, Attique M, Tahir MZ, Yasmim M, Saba T, and Tanik UJ.(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.
50. Arshad H, Khan M A, 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. International Journal of Machine Learning and Cybernetics 10 (12):3601-3618
51. 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. Microscopy Research and Technique
52. Rehman A, Khan MA, Mehmood Z, Saba T, Sardaraz M, Rashid M, (2020) Microscopic melanoma detection and classification: A framework of pixel-based fusion and multilevel features reduction. Microscopy Research and Technique 83 (4):410-423
53. Khan MA, Rubab S, Kashif A, Sharif MI, Muhammad N, Shah JH, Zhang Y, Satapathy SC, (2020) Lungs cancer classification from CT images: An integrated design of contrast based classical features fusion and selection. Pattern Recognition Letters 129:77-85
54. Sharif MI, Li JP, Khan MA, Saleem MA, (2020) Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images. Pattern Recognition Letters 129:181-189
55. 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. Journal of Medical Systems 43 (9)