International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

A Signature Verification System with Ensemble Classifier

Alpana Deka

Abstract*: Handwritten signature is considered as one of the established authentication process to study the behavioral nature of a person. This paper focuses on verification of offline handwritten signatures (for English scripts) as either genuine or forgery. Here the considered samples are genuine, skilled and simple forgeries. The verification is carried out by ensembling the three base classifiers Naive Bayes (NB), K-Nearest Neighbor (KNN) and Kmeans classifiers. The accuracies for skilled and simple forgeries are obtained as 86 % and 92 % respectively.

***Keywords*: Ensemble Classifier, Features extraction, Offline, Preprocessing.**

I. INTRODUCTION

To establish the authentication of any written documents, signature is considered as a trusted tool to accept that document as a legal one. This document may be related to any official documents or banking cheques etc. But mostly signatures of the banking sector are required to be verified. Signature verification process measures the degree of variability of signatures which results the given signature as genuine or forgery. Although the signatures of a person may vary during his life time upto some considerable extent (i.e. length, height, length of stray marks etc.), but it cannot change completely. Although there are similarities between the signatures of same person but the variation which may occur themselves is known as intra personal variation. Again an imposter always tries to make an exact copy of the genuine signature such that visual inspection cannot find out the difference between original and forgery signatures. But still there remains some differentiation between them which is known as inter personal variation [1].

Above two variations can be tested with two modes: online and offline [2]. In online one, special hardware device is used where both static and dynamic features from the signatures can be captured. But in offline mode, since no such device is used therefore only static features can be extracted from the input image which is taken from scanned document. Although there remains lack of dynamic features in offline one, but still it is more user friendly verification mode than the online one and hence in proposed system, the offline verification process is applied. Whether the mode is offline or online, but main aim of both of these is to establish the genuine signature as genuine and forgery as forgery. Here, genuine means the signature which is signed by the actual owner of that signature. But in forgery case, depending on the variations between genuine and forgery, there may be three types: random, simple and skilled [3]. The rate of degree of variability is lower to higher for random, simple and skilled forgeries respectively. Since without any automatic verification process, the random forgery can be detected easily therefore in proposed system we have considered simple and skilled forgeries along with the genuine samples.

Manuscript published on November 30, 2019.
* Correspondence Author
Alpana Deka*, department of Computer Science, NERIM Group of Institutions, Guwahati, India.

II. METHODOLOGY

The following figure Fig. 1 illustrates about clear indication of the proposed system.

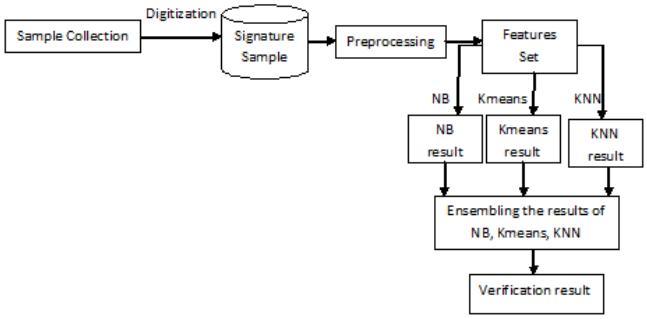


Fig. 1. Overall idea about the system

A. Sample Collection

In proposed system, 2262 signature samples are collected from a total of 58 persons. These persons are from research scholars and officials of different levels. For signing process, white piece of papers were provided to each person without giving any restrictions such that they can freely sign. This sample collection step was completed with 3 steps keeping 3 months time gap between each of the steps. After completion of data collection, training and testing sets are arranged with the help of trial and error method such that we can get maximum efficiency of the system. Since SVM, KNN are supervised classifiers and Kmeans is an unsupervised classifier therefore for training set of SVM and KNN both genuine and forgery signatures are taken into consideration but in case of Kmeans only genuine signatures are taken.

B. Digitization

Since the signature samples were collected on the white papers, therefore a scanner is used such that the images from a document form can be transformed to digitalized form.

C. Preprocessing

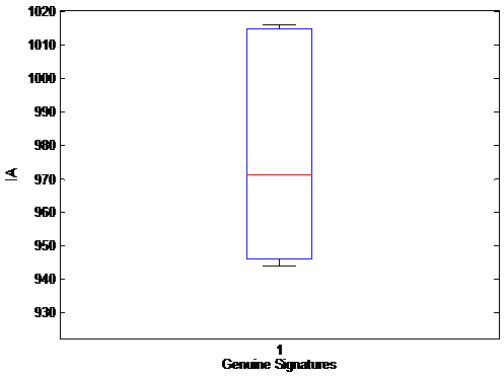
Now the digitalized signature images are taken as input image in order to clear the unwanted part from it. The unwanted part may be in terms of adjustment of size of the signature, removal of noises, extra spaces, unwanted stray marks etc.

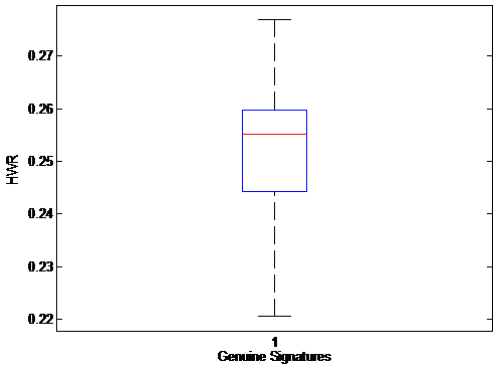
D. Features Set

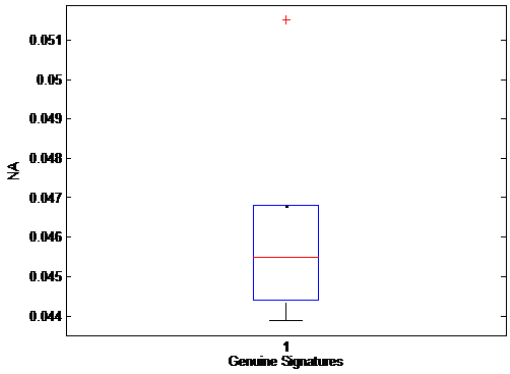
Since handwritten signature verification relates with the field of pattern recognition system and hence to study the pattern, features of the signature samples are required. Features represent the characteristics that differentiate signatures of one person to another. The set of all required features collected from a signature sample signifies a features set. In proposed system, features set consist of six features such as Image Area (IA), Height to Width Ratio (HWR), Normalized Area (NA), Maximum Horizontal Projection (MHP), Maximum Vertical Projection (MVP), Sum of Local Normalized Area (SLNA) [4] [5]. Each of them is explained below.

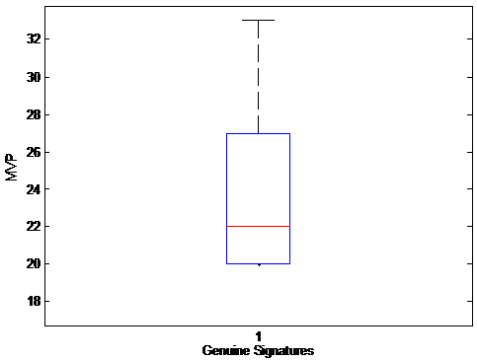
1. **IA:** Here, the active pixels (or on pixels) of each signature sample is taken under consideration to sum up such that we can get area of that sample. In our case, the white pixels of the signature (as given in Fig. 4(b)) indicate the active pixels.
2. **HWR**: Here, first height is calculated by scanning the signature image horizontally i.e. from left to right of the image to count number of rows. Then the image will be scanned vertically from top to bottom to get the number of columns which indicates width of that signature. Finally, the ratio between height and width is calculated to obtain HWR.
3. **NA:** Here, the height is multiplied by width. Then NA is obtained by taking the ratio between IA and the multiplication result.
4. **MHP and MVP**: The MHP and MVP are reverse to each other. Both examine the maximum number of white pixels but in opposite directions. MHP studies in horizontal direction with rows where MVP searches for columns in vertical direction.
5. **SLNA**: SLNA is obtained by summing up the values of NA taken from four partitions with respect to columns of the signature.

To give the overview representation of above mentioned features, following boxplots are drawn for two persons, user 1 and user 2 as given below:









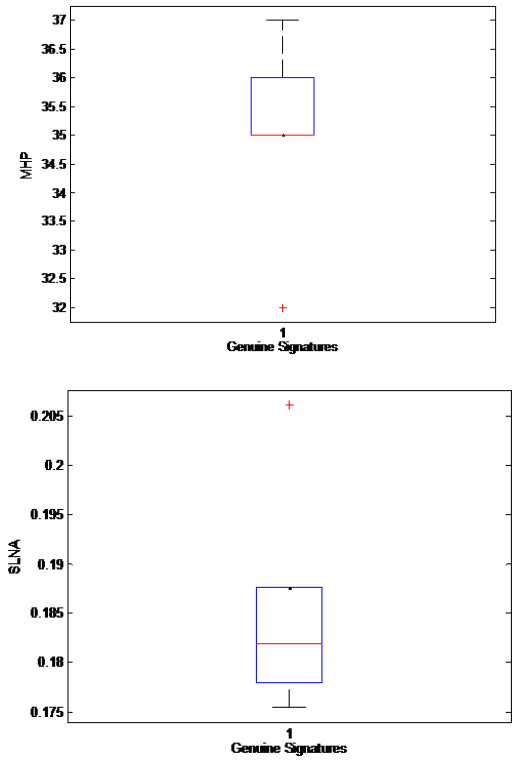
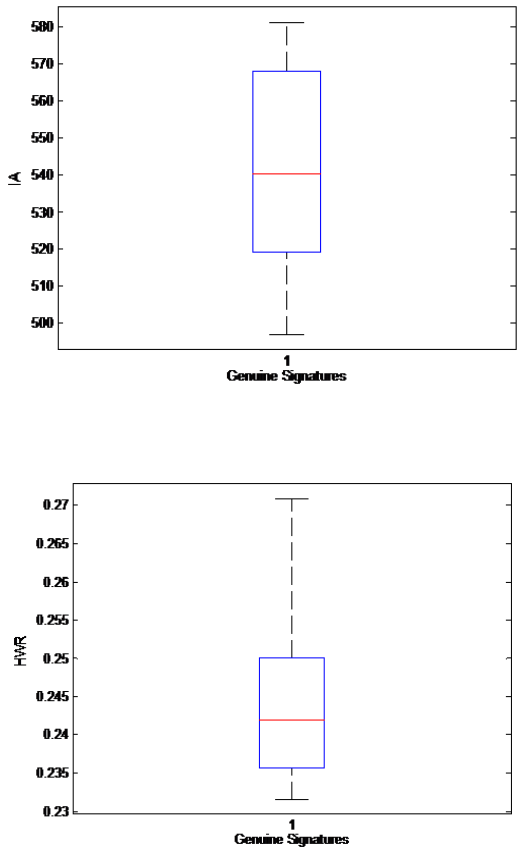
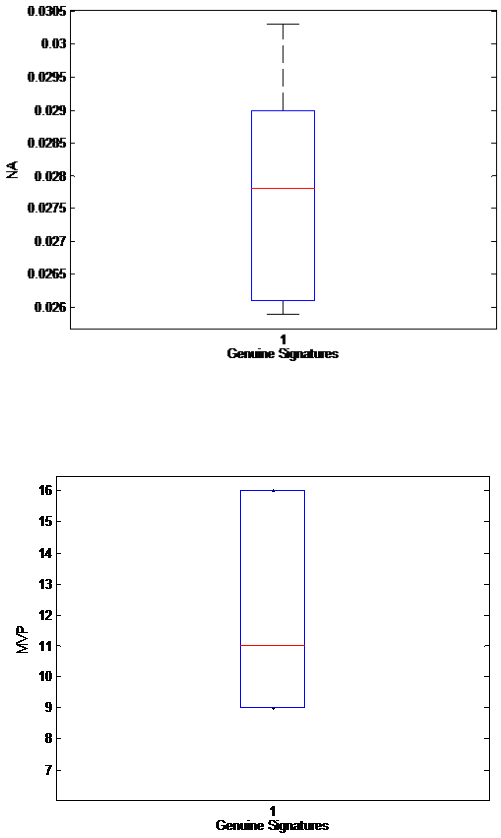


Fig. 2. Boxplot of different features for user 1





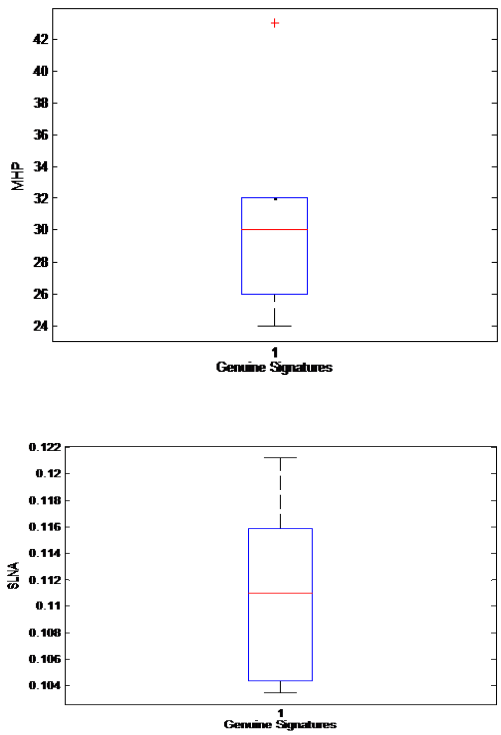


Fig. 3. Boxplot of different features for user 2

Here, boxplots of both users show different measures. For example, the medians are (970, 540) for IA, (0.26, 0.24) for HWR, (0.045, 0.0275) for NA, (22, 11) for MVP, (35, 30) for MHP and (0.18, 0.11) for SLNA of both users respectively.

E. Classification Stage

Classification is the final stage of pattern recognition system where pattern of an object is recognized. Here, to classify the pattern of signatures, three classifiers NB, KNN and Kmeans are applied.

The NB method is probabilistic in nature which is based on Bayes theorem [6]. Here, we have considered equal probabilities i.e. 0.5 to each of the classes of genuine and forgery such that prior probabilities can be predefined and also the distribution is taken as gaussian probability distribution which are the prerequisites of naive bayes classifier. Finally, class label of the tested signature is declared on the basis of maximum posterior probability between the two classes [7].

The second classifier, KNN classifies an object on the basis of closest training data of the feature space. That means decision is taken on the basis of the training signatures from which the distance is minimum of the tested signature. Here, by trial and error method, the distance function and number of nearest neighbor are decided to take as cosine and 4 respectively [8] [9].

The K-means clustering technique is also known as hard clustering technique. By taking the parameter k, it partitions data items into clusters such that the degree of similarity is low within the cluster but more between the clusters. In proposed system, value of k is taken as 2 [10].

In our system, we propose to apply ensemble classifier keeping in mind that the recognition result of the system will be better than the application of single classifier. Ensemble classifier classifies an object by combining results from multiple classifiers [11]. These multiple classifiers are called base classifiers. In proposed system, the above mentioned NB, KNN and Kmeans are taken as base classifiers. Since amongst the available combiners, majority voting method is easy to apply, therefore it is applied in the proposed system to combine the results of these three base classifiers. By applying the majority voting method, the decisions of patterns from the three base classifiers are combined and final decision is taken on the basis of maximum contribution to that pattern. It is well known that signature verification is a binary verification process that means it declares a signature as either genuine or forgery. Thus in our proposed system, if at least two base classifiers recognize the tested signature as genuine then the resultant ensemble result will be genuine and if they recognize as forgery then the tested signature will be forgery.

F. Experimental Result

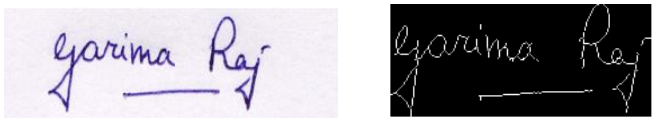
Step 1. Preprocessing

(a) **(b)**

Fig. 4. A genuine signature before and after preprocessing

(a) **(b)**

Fig. 5. A simple forgery signature before and after preprocessing



(a) **(b)**

Fig. 6. A skilled forgery signature before and after preprocessing

Step 2. Preparation of training and testing set for signature sample as in Fig. 4(b), Fig. 5(b) and Fig. 6(b).

i) For genuine signature verification:

Table- I: Training set for the figure as in Fig. 4

IA	HWR	NA	MVP	MHP	SLNA
805	0.4536	0.0530	18	27	0.2126
782	0.4278	0.0486	19	27	0.1943
716	0.3398	0.0497	16	26	0.1986
625	0.3855	0.0588	14	28	0.2351
678	0.4217	0.0583	12	29	0.2332
551	0.4812	0.0647	15	31	0.2587
600	0.3744	0.0421	17	34	0.1689
698	0.4018	0.0346	17	39	0.1385
708	0.4258	0.0381	14	31	0.1520
688	0.4478	0.0380	19	29	0.1519
634	0.8532	0.0625	33	22	0.2502
769	0.4140	0.0537	17	26	0.2141

Table- II: Testing set for the figure as in Fig. 4

IA	HWR	NA	MVP	MHP	SLNA
738	0.4111	0.0544	16	27	0.2216

ii) For simple forgery verification:

Table- III: Testing set for the figure as in Fig. 5

IA	HWR	NA	MVP	MHP	SLNA
724	0.3210	0.0307	39	35	0.1227

iii) For skilled forgery verification:

Table- IV: Testing set for the figure as in Fig. 6

IA	HWR	NA	MVP	MHP	SLNA
675	0.4388	0.0400	19	39	0.1602

Since same user is taken for all the verifications that means genuine, simple and skilled forgery verification therefore the training set will be same for all the cases. But difference will come in testing set only.

Step 3. Testing Phase

i) For genuine signature verification:

By taking the parameters as explained in above, finally we get as below:

Table- V: Verification result

NB	KNN	Kmeans	Ensemble classifier
Genuine	Genuine	Genuine	Genuine

ii) For skilled forgery verification:

Table- VI: Verification result

NB	KNN	Kmeans	Ensemble classifier
Forgery	Forgery	Forgery	Forgery

iii) For simple forgery verification:

Similarly proceeding as above

Table- VII: Verification result

NB	KNN	Kmeans	Ensemble classifier
Forgery	Forgery	Forgery	Forgery

III. RECOGNITION RATE

In proposed system, recognition rate of each of the base classifiers and ensemble classifier is obtained as in Table VIII.

Table- VIII: Recognition rate (accuracy) of different classifiers

Signature type \ Classifier	NB	KNN	Kmeans	Ensemble
Skilled forgery	83%	79%	76%	86%
Simple forgery	89%	84%	82%	92%

Now, Table VIII is represented in pictorial form as given in Fig. 7 and Fig. 8.

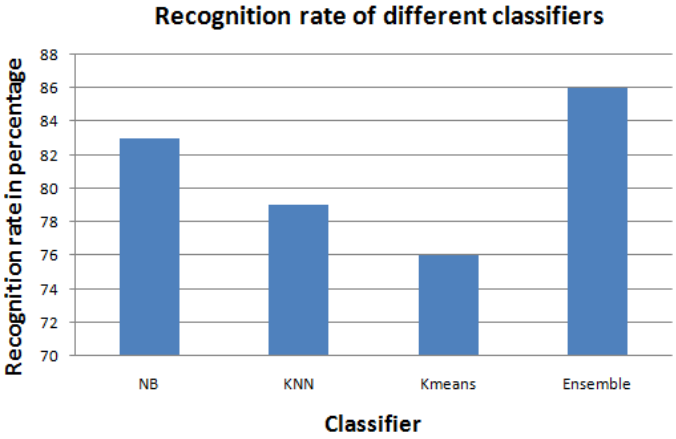


Fig. 7. Recognition rate (accuracy) for skilled forgery

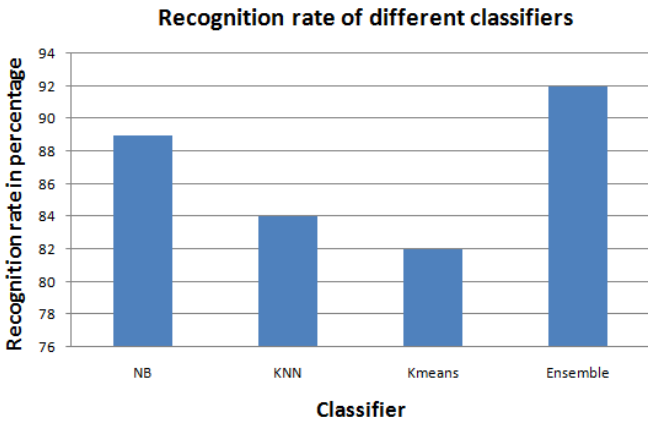


Fig. 8. Recognition rate (accuracy) for simple forgery

The above table Table VIII indicates that although the recognition rate of each individual classifier is lesser than that of the ensemble classifier but ensemble of all three base classifiers increases the system efficiency.

IV. CONCLUSION

The proposed system consists of English handwritten signature samples with genuine, simple and skilled forgery. From the tabulated values as given in Table IX, result of the proposed system is found to be encouraging and superior.

Table- IX: Recognition of different classifiers

Author	Database	AER	Accuracy
Sigari, M. H. et al. [12]	Persian/Skilled	19.25 %	-
Swanepoel J. P. et al.[13]	English/Simple	10.23 %	-
Proposed system	English/Skilled	19 %	86%
	English/Simple	8%	92%

REFERENCES

1. Garhawal, S., Shukla, N., "A Study on Handwritten Signature Verification System", International Journal of Advanced Research in Computer Engineering & Technology, Vol. 2, Issue 8, 2497-2503. (2013).
2. Ashok, K. D., Dhandapani, S.: Offline Signature Verification System for Bank Cheques Using Zerinke Moments, Circularity Property and Fuzzy Logic. International Journal of Engineering and Computer Science, Vol. 6, Issue 9, 22442-22449. (2017).
3. Panchal, S. T., Yerigeri , V. V.: Offline Signature Verification based on Geometric Feature Extraction using Artificial Neural Network. IOSR Journal of Electronics and Communication Engineering, Vol. 13, Issue 3, 55-59. (2018).
4. Ahmed, H., Shukla, S.: Comparative Analysis of Global Feature Extraction Methods for Off-line Signature Recognition. International Journal of Computer Applications, Vol. 48, Issue 23, 15-19. (2012).
5. Kisku, D. R., Gupta, P., Sing, J. K.: Offline Signature Identification by fusion of Multiple Classifiers using Statistical Learning Theory. International Journal of Security and its Applications, Vol. 4, Issue 3, 35-45. (2010).
6. Park, D.C.: International Journal of Computer Science and Electronics Engineering, Volume 4, Issue 3, 135-139. (2016).
7. Murty , M. N., Devi , V. S.," Pattern Recognition: An Algorithmic Approach" Springer, 86-96. (2011).
8. Jaafar, H., Ramli, N. H., Nasir, A. S. A.: An Improvement To The K-Nearest Neighbor Classifier For ECG Database. IOP Conference Series: Materials Science and Engineering, 1-10. (2018).
9. Hu, L. Y., Huang, M. W., Ke, S. W., Tsai, C. F.: The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus, pp: 1-9. (2016)
10. Ali, H. H., Kadhum, L. E: K-Means Clustering Algorithm Applications in Data Mining and Pattern Recognition, International Journal of Science and Research, Vol. 6, Issue 8, 1577-1584. (2017)
11. Dietterich, T. G.: Ensemble Methods in Machine Learning. In proceedings of the first international workshop on multiple classifier systems, Italy, 1-15. (2000)
12. Sigari, M. H., Pourshahabi, M. R., Pourreza, H. R.: Offline Handwritten Signature Identification and Verification Using Multi-Resolution Gabor Wavelet, International Journal of Biometrics and Bioinformatics, Vol. 5, Issue 4, pp: 234-248. (2011)
13. Swanepoel, J. P.: Off-line signature verification using classifier ensembles and flexible grid features, Thesis presented in partial fulfillment of the requirements for the degree of Master of Master of Science in Applied Mathematics at Stellenbosch University, South Africa, 19-57. (2009)

AUTHORS PROFILE

Alpana Deka is an assistant professor in the Department of Computer Science at NERIM Group of Institutions, Guwahati, Assam. She received her Ph. D degree from Gauhati university. Her research interest includes pattern recognition, image processing and different statistical approaches etc.