An Ensemble Based Offline Handwritten Signature Verification System

Alpana Deka [1], Lipi B. Mahanta [2,*]

[1]*Department of Computer Science, NERIM Group of Institutions, Guwahati, Assam, India*
[2]*Central Computational and Numerical Sciences, Vigyan path, Paschim Boragaon, Garchuk, Guwahati, Assam, India*

Abstract In the field of security and forgery prevention, handwritten signatures are the most widely recognized biometric since long and also most practical. Although handwritten signature verification systems are studied using both On-line and Off-line approaches, Off-line signature verification systems are more difficult to compare to On-line verification systems. This is due to the lack of dynamic information, viz. a database which constantly stores the latest signature of the person. In the paper, an approach using ensemble methods are adopted to classify a signature as forgery or not. In the proposed system, three classifiers, viz, one unsupervised, viz. Fuzzy C-Means (FCM) and two supervised classifiers, viz. Naive Bayes (NB) and Support Vector Machine (SVM) are used as base classifiers. An attempt is made to compare the different approaches. We attempt both the categories of classification not only because both of them are applicable in this particular case but also intending to find out which performs better. In most cases, it is observed that Naive Bayes and Ensemble are comparable as they exhibit better performance than the other two. But among them, in most of the cases Ensemble classifier performs better than the Naive Bayes and consequently, we have taken the Ensemble as a final classifier.

Keywords Handwritten signature, Offline, Verification, Ensemble classification

1. Introduction

A signature is nothing but a combination of letters, words and special symbols. Signatures are utilized as a means of identity verifier for authentication, attestation and authorization in different legal transactions especially in banking environment [1]. Depending on the nature of the samples acquisition method, a signature verification system may of online and offline [2, 3]. The off-line handwritten recognition systems are more difficult and a more practical application area as compared to the on-line one. In a pattern recognition system, an ensemble classifier is the fusion of multiple base classifiers to predict a class label of an unseen instance [4]. Here, the base classifiers can be aggregated with the help of combiners like simple majority voting, weighted majority voting, Bayesian combination, probabilistic approximation etc. Among them, the simple majority voting is found to be immensely used combiner due to its simplicity [5]. Thus the output which appears most frequently as a vote for the base classifiers is considered as a required pattern of an instance [6].

Ensemble classifier may be of two types, viz. parallel and serial [7]. In the proposed system, we have applied parallel ensemble classifier where Fuzzy C-Means (FCM), Naive Bayes (NB) classifier and Support Vector Machine (SVM) is applied as base classifiers such that they can perform the classification task independently and finally their outcomes are aggregated to predict the pattern of a given signature. The choice of these three classifiers are based on the fact that we do not have a readily available large training dataset and NB is the best

---

choice when the training set is small with high bias/low variance. SVM are popular in text classification problems where very high-dimensional spaces are the norm. It also gives high accuracy, nice theoretical guarantees regarding overfitting, and with an appropriate kernel, they can work well even if the data isn't linearly separable in the base feature space. FCM was adopted as a writer-independent verification model, where a single classifier deals with the whole population of writers. The main advantages of writer independent model are - reduction in the number of signature samples required, for training and validation, and ability to absorb new writers without generating new personal models.

The prediction results of these three classifiers are fused with the simple majority voting combiner. Thus based on the following possible voting conditions (as given in Table 1), the ensemble classifier decides whether a tested signature is genuine or forgery. We present in Figure 1 the overview of our proposed work.
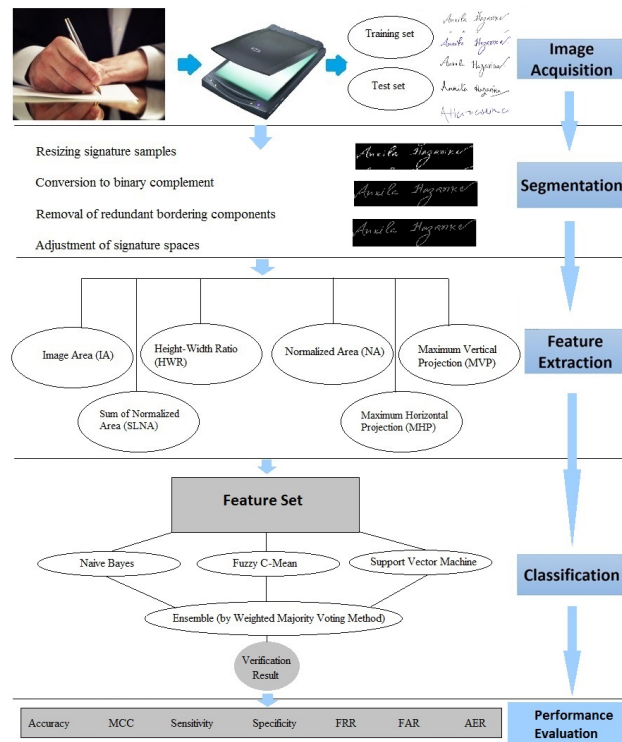

Figure 1. Parallel ensemble classifier for proposed signature verification system.

The objective of the study was to detect forgery for all the three classes of counterfeit, viz. random, simple and skilled. It has enormous practical utility because the system can be applied in and Banking, Financial and other Document registration organizations, where every day signatures have to be documented and verified in bulk. The contribution of the study may be summarized as follows:

- Development of a well-documented and real-time database consisting of genuine and forged samples of all the three categories: random, simple and skilled.
- Development of an algorithm for several classification methods and creating the ensemble for the most accurate output.
- Development of the algorithm for all three categories of forgery. Hence it is a robust algorithm.

Table 1. Possible Majority voting conditions of proposed Ensemble Model.

FCM	NB	SVM	Ensemble Decision
Genuine	Genuine	Genuine	Genuine
Forgery	Genuine	Genuine	Genuine
Genuine	Forgery	Genuine	Genuine
Genuine	Genuine	Forgery	Genuine
Genuine	Forgery	Forgery	Forgery
Forgery	Forgery	Forgery	Forgery
Forgery	Forgery	Genuine	Forgery
Forgery	Genuine	Forgery	Forgery

2. Methodology

2.1. Data Acquisition

An indigenous data set is created, where genuine signature samples are collected from each of 85 persons in two stages keeping a sufficient time gap. Out of this, for the training set of FCM technique, we have considered a total of 510 (i.e. 85x6) numbers of genuine signatures. Again for supervised classifiers, predefined knowledge about the classes should be known. Therefore for training sets of NB and SVM classifiers, we have taken another 510 (i.e. 85x6) numbers of forgery (2 from each of random, simple and skilled forgery) signatures apart from the above mentioned genuine signatures. In both cases, the testing dataset consists of 765 genuine signatures and 2295 (765x3, each of the 765 signatures is taken for random, simple and skilled) forgery signatures. (Here, both genuine and forgery signature samples are collected on a white sheet of paper). The details are tabulated below in Table 2.

Table 2. Data set.

	Training Set	Test set
$Classifier$	Number of training samples	Number of testing samples
FCM	6x85=510	765+2295=3060
NB	12x85=1020	765+2295=3060
SVM	12x85=1020	765+2295=3060
$Ensemble$	—-	765+2295=3060

2.2. Preprocessing

Here, a binary signature image (which is the conversion of RGB to grayscale and then to binary form) is taken as an input image where the operations like complement binarization, removal of redundant bordering components, adjustment of signatures spaces etc. are performed [8, 9].

1. Conversion to binary complement signature: The complement of a binary signature is the reverse of the binary signature. A complement binary signature is displayed in Figure 2. Here, to invert the binary signature, the logical operator NOT is applied on that binary signature such that

$$\overline{B(x,y)} = \begin{cases} 0, & \text{if B(x,y) = 1} \\ 1, & \text{if B(x,y) = 0} \end{cases} \tag{1}$$

2. Removal of redundant bordering components: During adjustment of signatures in Photoshop software, the bordering components as shown in Figure 3(a) are found to be present in some of the samples. Here, we have discarded such elements in order to interpret only the signature region. Now, consider a marker and original

Figure 2. A complement binary signature.

signature images as F_{mark} and I_{bor} respectively.

$$F_{mark}(x,y) = \begin{cases} \begin{cases} I_{bor}(x,y), & if\, pixel(x,y) \\ present\, on\, the\, border\, of\, I_{bor}, \\ 0 & otherwise \end{cases} \end{cases} \quad (2)$$

Next, to extract only bordered elements from the given signature, we have performed reconstruction taking I_{bor} as a mask image by following way:

$$J = R(F_{\mathrm{mark}}) \quad (3)$$

Here, the resultant J is the image of only bordered elements. Now, to obtain only the signature region which is free from bordered elements, we have performed the following operation as below:

$$K = 1 - J \quad (4)$$

The final resultant signature is displayed in Figure 3(b).

Figure 3. Before and after elimination of bordering components from a signature.

3. Adjustment of signatures spaces: In order to perform spaces adjustment, the output of Fig 3 (b) has been followed horizontally and vertically to detect the first and last rows with white pixels. This gives the indices about the first and last rows as well as columns with white pixels along with their respective directions. Suppose, $[h_{first}, h_{last}]$ be the indices of desired rows along the horizontal direction and $[v_{first}, v_{last}]$ be the indices of desired columns along the vertical direction. This means that $[h_{first}, h_{last}]$ and $[v_{first}, v_{last}]$ are horizontal and vertical ranges within which the signature is to be confined. Therefore, I $(h_{first} : h_{last}, v_{first} : v_{last})$ will be the resultant signature with rows $(h_{first} : h_{last})$ and columns $(v_{first} : v_{last})$. An overview of positions of desired rows and columns and signature after adjustment are displayed in Figure 4-5.
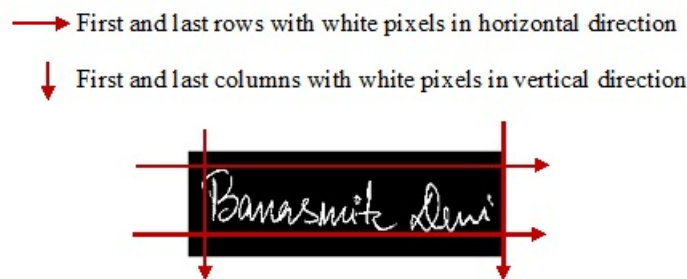

Figure 4. An overview of actual initial and final positions of rows, columns with white pixels.

Figure 5. Adjustment of a signature after removal of extra spaces.

2.3. Features Extraction

In the proposed system, we have extracted the global and local features as mentioned below [10, 11, 12, 13, 14]:

1. Image area (IA): Number of on pixels represents image area of a signature.
2. Height-Width Ratio (HWR): It is a ratio of height to width of a signature which is calculated along vertical and horizontal direction respectively.
3. Normalized area (NA): It is calculated by taking the ratio of the area of a signature image to the area of signature.
4. Maximum vertical projection (MVP): It represents a column with maximum value in the horizontal direction.
5. Maximum horizontal projection (MHP): MHP is the same as a maximum vertical projection but done in a vertical direction.
6. Sum of four local Normalized Areas (SLNA): Here we divide the final preprocessed image into four parts, then sum up normalized areas of each of the part.

2.4. Classification

The three base classifiers are explained below:

2.4.1. FCM clustering technique: The FCM technique partitions data elements or objects into clusters based on their similarity of behavior [15, 16, 17]. We have developed our algorithm for verification of signatures with FCM technique on the assumption that if the cluster size (i.e. the number of data items along with the testing data present in a cluster) is more than 2, the process will continue. Thus when the cluster size becomes 2 then a given signature will be assumed as genuine and when it becomes 1 (i.e. no training points remain with the testing data) then the signature will be recognized as a forgery.

2.4.2. NB Classifier: Formally, NB classifier is based on the following three conditions [18]:

- Pre-defined knowledge about the classes in terms of prior probabilities should be known (Here, we have taken prior = 0.5, since before testing a signature *p(genuine) = p(forgery)*).
- To estimate parameters, a suitable distribution is selected. (In the proposed system, Gaussian distribution is implemented, being the general system of distributions).
- To assign a class label to a test pattern, posterior probabilities of available classes are evaluated. Finally, proper class label of the pattern is declared based on maximum posterior probability. That is, the tested signature is genuine if *post(gen)>post(frg)* and forgery if *post(gen)<post(frg)*.

2.4.3. SVM Classifier: In our proposed system, we have applied the LS-SVM classifier with the help of trial and error method. In LS-SVM classifier, every data point acts as a support vector where some of them have more contribution than others [19]. We have selected the following parameters.

- Kernel function: Radial Basis Function where the value of δ is taken as 1.
- Boxconstraint: Boxconstraint is nothing but the tradeoff parameter between error and margin which is taken as 1e-2 in the system.

SVM can classify both linearly and non-linearly separable data. Data in our proposed system are non-linearly separable data.

3. Results and Discussion

Experiments were conducted with the methodology described above with various test signatures belonging to the categories genuine and skilled, simple and random type of forgery cases. One signature of each type is illustrated below.

3.1. Verification with Genuine Signature

Figure 6. A tested genuine signature with before and after application of preprocessing steps.

We illustrate the result with the help of a signature (which is originally genuine), as indicated in Figure 6. The extracted features after requisite pre-processing are indicated in Table 3.

Table 3. Values of extracted features of test signature.

Features	Values of features
IA	642
HWR	0.3455
NA	0.0384
MVP	10
MHP	24
SLNR	0.1536

Using FCM: From Table 3, co-ordinate of the tested signature concerning features Area and Height-Width Ratio are (642, 0.3455). In Figure 7, this tested data item is indicated by an arrow mark. From the alignment of the training data items, it is clear that the indicated data item is in the left-hand sides cluster. Since our verification result with FCM is based on the assumption that the number of data items in the final cluster < 3, therefore, we have to again divide that particular cluster where the tested data item lies, based on the features Normalized Area (NA) as well as Sum of Local Normalized Areas (SLNA). Now, the co-ordinate of the tested signature or data item for features like Normalised Area and Sum of Local Normalized Areas is (0.0384, 0.1536) as indicated by an arrow mark in Figure 8. From the figure, it is clear that the indicated data item lies in the right cluster. Since this data item is making a cluster with one another data item of the training set, hence we can conclude that tested signature is genuine.

Using NB: Here the probability values obtained for the given signature were post(gen) = 0.4923 and post(frg) = 6.2974e-004. Hence based on maximum posterior probability, the tested signature is recognized as genuine.

Using SVM: Applying SVM classifier (support vectors :<12x6>, kernel: linear kernel, alpha:<12x1>, bias: -0.0022) for the above mentioned tested signature, we have obtained the signature as a genuine one.

Thus based on simple majority voting algorithm, ensemble classifier correctly signifies the given signature as genuine.

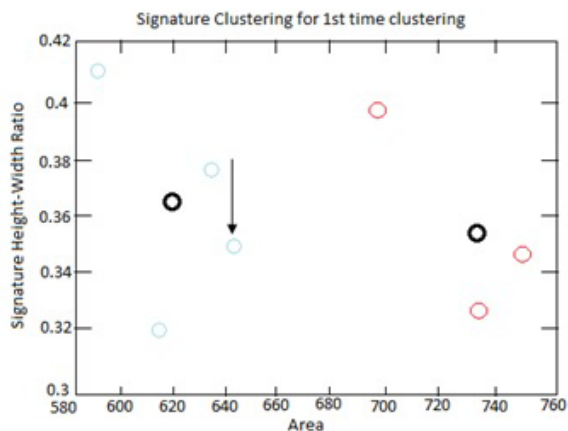| FCM	NB	SVM	Ensemble Decision
Genuine	Genuine	Genuine	Genuine

Figure 7. Data clustering of genuine signature with FCM for Area and Height-Width Ratio.
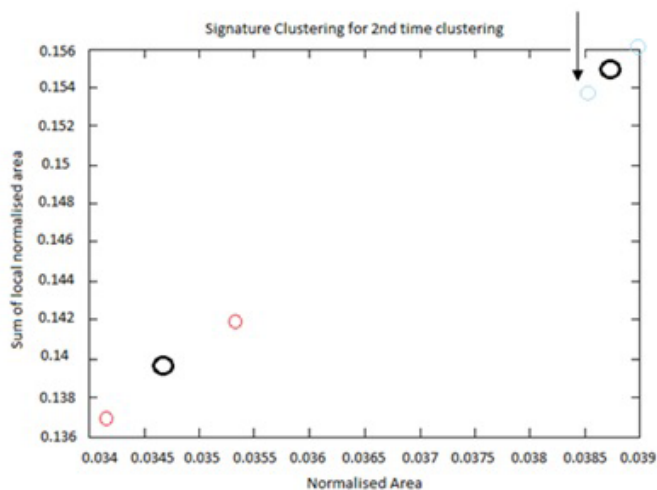

Figure 8. Data clustering of genuine signature with FCM for Normalized Area and Sum of Local Normalized Area.

The method is then applied to one each Skilled, Simple and Random forgery signatures (Figures 9, 11 and 13) of the same signature. It may be mentioned that here the values of the test data will be different but the training sets will be the same as before.

3.2. *Verification with Skilled Forgery Signature*

Figure 9. A tested shilled forgery signature with before and after application of preprocessing steps.

Using FCM: The FCM technique has to be conducted three times: first with IA and HWR, then with NA and MVP and finally with MHP and SLNA, from which it is clear that the tested signature is a forgery (Figure 10).
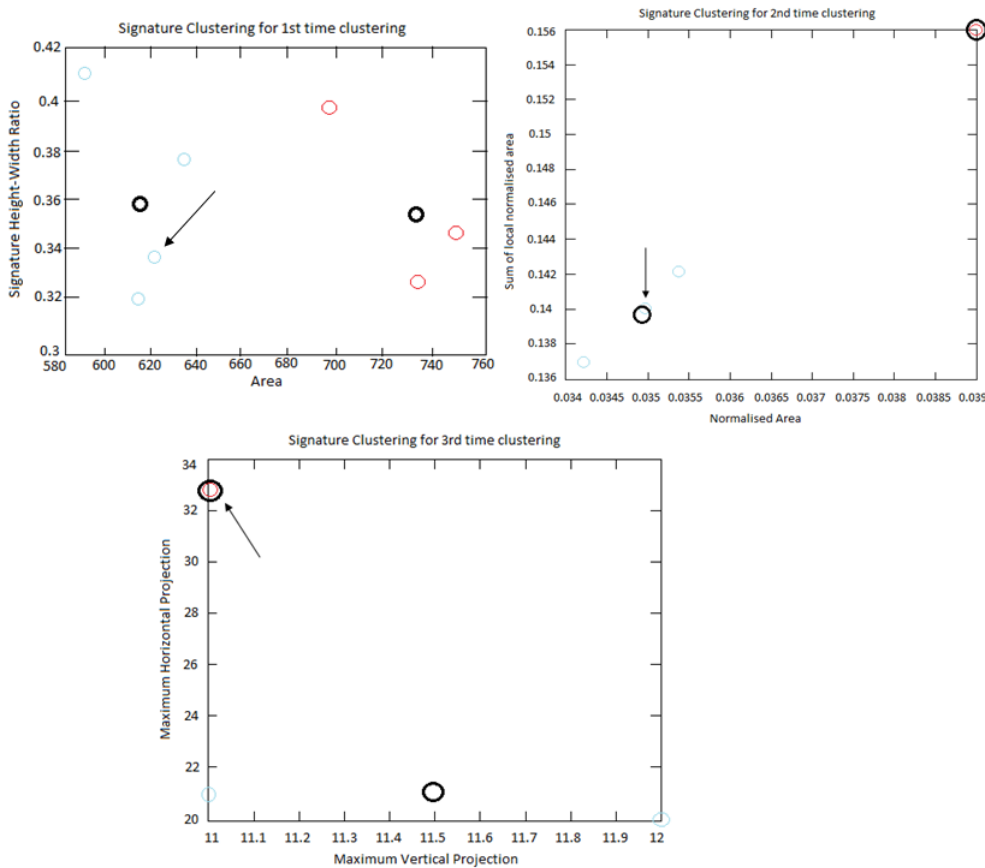

Figure 10. Data clustering of skilled forgery signature with FCM.

Using NB: Here post (gen) = 0.1273 and post (frg) = 4.1712e-004 Hence, by the method mentioned above, the tested signature is wrongly recognized as genuine.

Using SVM: Since the training set of tested signature is same as in the genuine case, therefore its support vectors, values of alpha and bias will also be same (as in above genuine case) which exhibit the tested signature as a forgery one. Thus,the results of the individual classifiers and the Ensemble classifier is as follow:

FCM	NB	SVM	Ensemble Decision
Forgery	Genuine	Forgery	Forgery

3.3. Verification with Simple Forgery Signature

Figure 11. A tested simple forgery signature with before and after application of preprocessing steps.
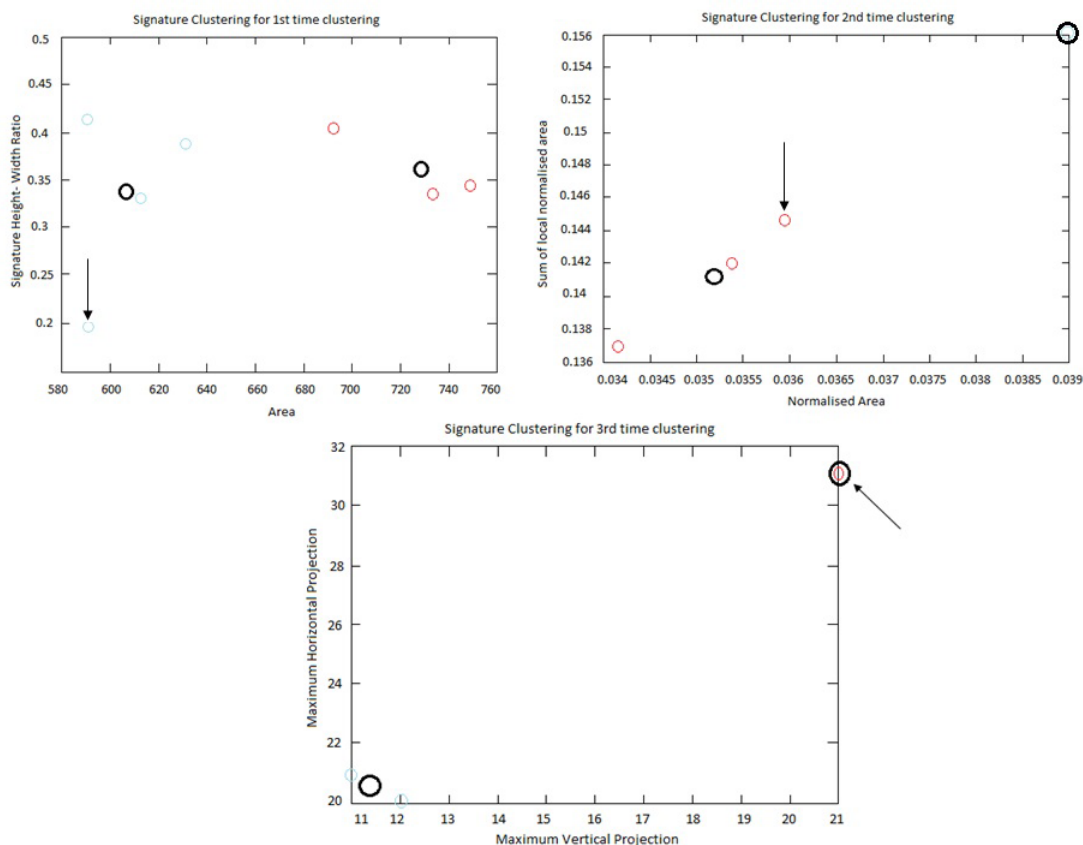
Figure 12. Data clustering of simple forgery signature with FCM.

Using FCM: Here also, the FCM has to be conducted three times. As a result, the tested signature is classified as forgery (Figure 12).

Using NB: Here, post (gen) = 7.3725e-010 and post (frg) = 4.8588e-009, hence result is forgery.

Using SVM: Here also, the training set of tested signature is same.

Thus,

FCM	NB	SVM	Ensemble Decision
Forgery	Forgery	Forgery	Forgery

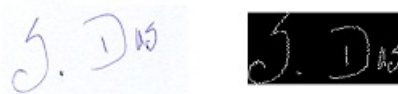### 3.4. Verification with Random forgery signature.

Figure 13. A tested random forgery signature with before and after application of preprocessing steps.
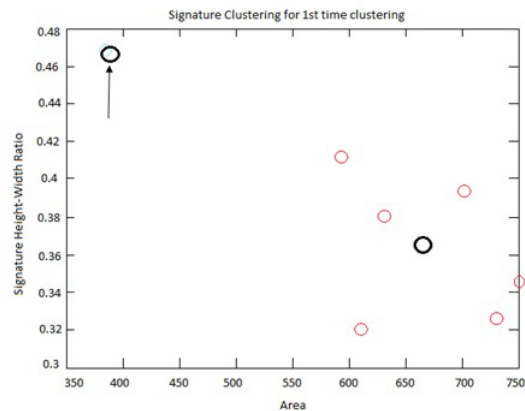
Figure 14. Data clustering of random forgery signature with FCM.

Using FCM: Here the test signature is classified as Forgery after the first step of clustering, with features IA and HWR (Figure 14).

Using NB: Here, post (gen) = 7.3725e-010 and post (frg) = 4.8588e-009, hence forgery.

Using SVM: Here also the training set of tested signature is same. Thus,

FCM	NB	SVM	Ensemble Decision
Forgery	Forgery	Forgery	Forgery

3.5. *Performance measurement*

The proposed method is evaluated with the following evaluation criteria [20](Table 4):

Table 4. Performance indicators.

Measurement (in %)	Mathematical Formula
Accuracy	$\frac{(A+B)}{(A+\overline{A}+B+\overline{B})}$ x 100
MCC	$\frac{A \times B - \overline{A} \times \overline{B}}{\sqrt{(A+\overline{A})(A+\overline{B})(B+\overline{A})(B+\overline{B})}}$
Sensitivity	$\frac{A}{(A+\overline{B})} \times 100$
Specificity	$\frac{B}{(B+\overline{A})} \times 100$
FRR	$\frac{\mathbb{C}}{D} \times 100$
FAR	$\frac{E}{F} \times 100$
AER	$\frac{(FRR+FAR)}{2} \times 100$

The smaller value of AER indicates higher significance of the system.

Here, the terms are:

A: A genuine signature as genuine

$\overline{A}$: A genuine signature as forgery

B: A forgery signature as forgery

$\overline{B}$: A forgery signature as genuine

C: Number of original signatures rejected

D: Number of original signatures tested

E: Number of forgery signatures rejected

F: Number of forgery signatures tested

Table 5. Performance with skilled forgery.

Measurement (in %)	FCM	NB	SVM	Ensemble Decision
Accuracy	75.36	81.18	80.46	83.46
MCC	0.51	0.63	0.61	0.67
Sensitivity	72	77.32	81.23	80.99
Specificity	79.94	86.30	79.72	86.36
FRR	17	11.76	20.78	12.55
FAR	32.29	25.88	18.30	20.52
AER	24.65	18.82	19.54	16.54

Table 6. Performance with simple forgery.

Measurement (in %)	FCM	NB	SVM	Ensemble Decision
Accuracy	84.12	89.74	81.55	90.59
MCC	0.68	0.8	0.63	0.81
Sensitivity	84.89	90.97	81.23	93.31
Specificity	83.38	88.58	81.83	88.19
FRR	17	11.76	20.78	12.55
FAR	14.77	8.76	6.41	6.27
AER	15.89	10.26	13.59	9.41

Table 7. Performance with random forgery.

Measurement (in %)	FCM	NB	SVM	Ensemble Decision
Accuracy	90.92	93.79	89.35	93.59
MCC	0.83	0.88	0.80	0.88
Sensitivity	98.60	99.26	99.34	99.70
Specificity	85.32	89.41	82.72	88.82
FRR	17	11.76	20.78	12.55
FAR	1.18	0.65	0.52	0.26
AER	9.09	6.21	10.65	6.41

From Tables 5-7, it is observed that Naive Bayes and Ensemble are comparable as they exhibit better performance than the other two. But among them, in most of the cases Ensemble classifier performs better than the Naive Bayes and consequently, we have taken the Ensemble as a final classifier.

Jana et al. [21] developed an euclidean based offline signature verification system (database size = 175, language = English) with global and local features. The computed FRR and FAR are 2.86% and 17.14% respectively. With the help of global and local features, Bhausaheb et al. [22] presented a euclidean based offline signature verification system (language = English) with accuracy 90% to 100% for 60 to 360 signature samples.Hatkar et al. [23] explained a neural network based offline signature verification system (database size = 1000, language = English) with geometrical features. The accuracy of the proposed system was obtained as 86.25%. Chugh et al. [24] illustrated a Kohonen networks based offline signature verification system where database size is 250. They obtained FAR and FRR as 2.8% and 5% respectively.

M.H. Sigari et al. [25] elaborated an ensemble offline signature verification system based on three simple classifiers with Nearest Neighbor using different combination methods where Majority Voting obtained the better result with 19.25% AER. Here, the database consists of 30 signature samples of 20 classes. Swanepoel, JP.[26] developed an offline signature verification system with database size 4800 where AER is obtained as 10.23%. Here, the combined classifier is constructed by ensembling DTW and HMM base classifiers.

Finally, we have compared the performance of the proposed system with other systems (which deal with ensemble classifier with majority voting method) as tabulated below in Table 8:

Table 8. Comparison of proposed system with other systems.

Author	Input Signatures	Forgery Type	Performance Measurement (AER)
M. H. Sigari et al.	Persian	Skilled	19.25%
J. P. Swanepoel	English	Simple	10.23%
J.-B. Fasquel et al. [27]	English	Random	19.3%
Proposed system	English	Skilled Simple Random	19.25% 9.41% 6.41%

4. Conclusion

Above table, Table 8, indicates that our proposed system gives more significant result than the other three systems in terms of their AER. Hence this technique may be adopted to resolve the confusing and routine cases of handwritten signature forgery cases faced in real-life situations.

REFERENCES

1. Shanker AP and Rajagopalan AN, *Off-line signature verification using DTW*, Pattern Recognition Letters 28:1407-1414, 2007.
2. Bertolini D, Oliveria LS, Justino E, Sabourin R, *Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers*, Pattern Recognition 43:387-396, 2010.
3. Scheidat T, Vielhauer C, Dittman J, *Handwriting verification-Comparison of a multi-algorithmic and a multi-semantic approach*, Image and Vision Computing 27(3):269-278, 2009.
4. Rokach L, *Pattern classification using ensemble methods*, World Scientific 75:19, 2010.
5. Yang P, Yang YH, Zhou BB, Zomaya AY, *A review of ensemble methods in bioinformatics*, Current Bioinformatics 5(4): 296-308, 2010.
6. Bramer M, *Principles of Data Mining*, Springer Science & Business Media 209, 2013.
7. YPonti MP, *Combining classifiers: from the creation of ensembles to the decision fusion*, IEEE Proceedings of 24th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials, Alagoas 1-10, 2011.
8. Gonzalez RC, Woods RE, Eddins SL , *Digital Image Processing Using Matlab*, Tata McGraw Hill Education Private Limited 1-498, 2011.
9. *Image Processing Toolbox*, www.mathworks.in (accessed on 24.07.2013)
10. Kumar S, Raja KB, Chhotaray RK, Pattnaik S, *Off-line signature verification based on fusion of grid and global features using neural network*, International Journal of Engineering Science and Technology 2(12):7035-7044, 2010.

11.   Ahmed H, Shukla S, *Comparative analysis of global feature extraction methods for off-line signature recognition*, International Journal of Computer Applications 48:15-19, 2012.
12.   Rathi A, Rathi D, Astya P, *Offline handwritten signature verification by using pixel based method*, International Journal of Engineering Research & Technology (IJERT) 1:1-4, 2012.
13.   Kisku DR, Gupta P, Sing JK, *Offline signature identification by fusion of multiple classifiers using statistical learning theory*, International Journal of Security and its Applications 4:35-45, 2010.
14.   Sharif M, Khan MA, Faisal M, Yasmin M, Farnandez SL, *A framework for offline signature verification system: best features selection approach*, Pattern Recognition Letters XXX, 2018.
15.   Bezdek JC, Ehrlich R, Full W, *FCM : the fuzzy c-means clustering algorithm*, Computers & Geosciences 10:2-3 and 191-203, 1984.
16.   Alguliyev R, Aliguliyev R, Imamverdiyev Y, Sukhostat L, *Weighted Clustering for Anomaly Detection in Big Data, Statistics*, Statistics, Optimization and Information Computing (6), 178-188, 2018.
17.   Macedo E, *Two-Step Semidefinite Programming Approach to Clustering and Dimensionality Reduction*, Statistics, Optimization and Information Computing (3), 294-311, 2015.
18.   Murty MN, Devi VS, *Pattern Recognition: An Algorithmic Approach*, Springer Science & Business Media 86-87, 2011.
19.   https://en.wikipedia.org/wiki/Least_squares_support_vector_machine (accessed on 02.04.2015).
20.   https://en.wikipedia.org/wiki/Sensitivity_and_specificity (accessed on 15.05.2014).
21.   Jana R, Mandal S, Chhaya K, *Offline Signature Verification for Authentication*, International Journal of Computer Applications, Image Processing and Pattern recognition 126(6):20-23, 2015.
22.   Bhausaheb RR, Kumar S, Sachin K, *Centroidal Distance Based Offline Signature Recognition Using Global and Local Features*, IPASJ International Journal of Computer Science 3(3):36-42, 2015.
23.   Hatkar PV, Salokhe BT, Malgave AA, *Offline Handwritten Signature Verification Using Neural Network*, International Journal of Innovations in Engineering Research and Technology 2(1):1-5, 2015.
24.   Chugh A,Jain C, *Kohonen Networks for Offline Signature Verification*, International Journal of Recent Research Aspects 4(2):18-23, 2017.
25.   Sigari MH, Pourshahabi MR, Pourreza HR, *An ensemble classifier approach for static signature verification based on multi-resolution extracted features*, International Journal of Signal Processing, Image Processing and Pattern recognition 5:21-35, 2012.
26.   Swanepoel JP, *Off-line signature verification using classifier ensembles and flexible grid features*, SThesis presented in partial fulfillment of the requirements for the degree of Master of Science in Applied Mathematics at Stellenbosch University, South Africa 19-57, 2009.
27.   Fasquel JB and Bruynooghe M, *A hybrid opto-electronic method for fast off-line handwritten signature verification*, International Journal on Document Analysis and Recognition 7(1):56-68, 2004.