# Auto-Upgrading of Signature Case-Base used for Efficient Detection of Identity and Authenticity

**Chitrita Chaudhuri**[1],
Department of Computer Science and Engineering, Jadavpur University
188 Raja S.C. Mallik Road, Jadavpur, Kolkata-700032, India

**Atal Chaudhuri**[2]
Department of Computer Science and Engineering, Jadavpur University
188 Raja S.C. Mallik Road, Jadavpur, Kolkata-700032, India

**Abstract :** Handwritten signatures, used for identifying a person, suffer from two major drawbacks: they are inherently inconsistent, and they can easily be counterfeited. In this work a classifier is built applying Case-Based Reasoning techniques to resolve these issues by preserving authentic sets of off-line signature images of people within cases. For identification purpose, a bit-sequence containing the most frequent discretized mode value for each global feature is retained per person, and the corresponding pattern forms an index to the case pertaining to each person. Identity is established by finding the nearest matching case, while searching the base with discretized global feature pattern obtained from a test signature. Authentication of the test signature is achieved by comparison with weighted central metrics obtained from feature sets and dynamic time warping values. Essentially median vectors and inter quartile ranges are calculated and preserved for the cases apriori to serve as authenticity indices. As part of an incremental upgradation scheme, a newly authenticated test signature automatically gets preserved in the base, either as an additional specimen, or by replacing the worst of the existing lot if it qualifies as a better one, thus improving the discerning power of the system. The proficiency of the classifier is assessed through accuracy measurements on two sets of data - one downloaded from a standard database [19] and the other collected by the researchers.

**Keywords:** Case-Based Reasoning (CBR), Global Features, Mode, Median, Inter Quartile Range (IQR), Dynamic Time Warping (DTW).

## 1 Introduction

The work being presented here is an attempt to explore the viability of applying Case-based reasoning techniques [8][16] to resolve the following problems : first, to identify humans through their off-line hand-written digitized signatures; second, to authenticate a presented signature and detect an impostor, if so be the case; and third, to improve the signature base temporally, by replacing earlier samples with better and more recent ones.

Soft identity verification schemes are becoming a part and parcel of the modern civilized world. Signatures have been the most popular and least bothersome technique of establishing

identity proof for a long time. Other biometric measures may be more reliable, but they pose a lot of inconvenience in the mode of collection and checking. Moreover, signatures can ultimately be scrutinised very easily by a human expert to resolve controversies.

In some domains, such as in the banking sector or a university examination centre, identifying a customer or a participant may simply be done by matching account numbers or roll numbers. But, as in most spheres, norms demand that the signature be presented for identification, so a connection needs to be established between the author's rightful case-entry and the presented signature. This process requires appropriate clustering techniques through measurement of similarity metrics between the test signature and those in the base. The success of the procedure depends on the derivation of a representative feature-vector for each case in the base. The test-signature is supposed to belong to the nearest-neighboring case or person.

Once a matching case has been identified, further comparison techniques need to be applied to declare the signature as either genuine or fraud. This constitutes the authentication module, where the process of signature verification can be posed as a problem to determine whether a particular signature is indeed written by the person claiming to be its author and, if not, whether attempts to forgeries can be established. The idea of applying Case-based Reasoning (CBR) and Dynamic Time Warping (DTW) techniques to solve this problem has been initially stimulated by studying pioneering works in these fields [16] [14]. Given enough granular evidence for a fake signature, the nature of forgery may also be predicted to be of random, unskilled or skilled variety [15].

Several offline signature verification methods have been found to utilize DTW methodologies in the literatures surveyed. Yoshimura and Yoshimura (1997)[18] proposed signature verification using DTW to segment the signature into a fixed number of components and then compute a component wise dissimilarity measure. In [17] Shankar and Rajagopalan propose a modified DTW algorithm that takes into account stability of various components of the signature for enhanced performance in verification.

A unique phase of this work involves adapting the case-base to changed circumstances, by updating with more recent and truer specimen. This is felt to be a necessary and vital appendage to most databases of offline signatures. An argument favoring the issue may here be forwarded with a suitable example. A valued customer of a bank may not like to have a genuine signature being rejected due to changes caused by age or infirmity. Any such implicit progression are automatically retained by the updation module included in this work.

The choice of classifier was guided by some advantages of the CBR - it does not need any separate training with plausible forged sets. Additionally, since each case preserves the metrics derived from a fixed number of pre-recorded signatures of a person, there is no scalability issue associated with the growth of the total dataset.

The remaining part of this paper has been divided into several more sections. The following section 2, entitled *Case Base Design Methodology*, describes techniques for acquisition and preprocessing of signature images, extraction of features[9][4][2] and preservation of central tendencies of such features in the case-base. Identity of a signatory can be established by utilizing the minimal set of associated feature value sequences frequently reappearing within signature sets[13]. Augmentation by dynamic time warping distances[5][18][14][17] between compared signature images enabled to develop characteristic functions for proving the authenticity of the signatory. Finally, the procedure

adopted to update the case-base, with the best recent-most set of genuine signatures presented to the system, is described and methods discussed to utilize this updated information for improving the process of identification and authentication.

To benchmark the system the signature data have been tested on other classifiers : the Multi-Layer Perceptron (MLP) Network, the Support Vector Machine (SVM), the Naïve Bayesian (NB) model, the Decision Tree (DT), and the K-Nearest-Neighbour (KNN) model. Section 3, describing the details of the experiments performed to assess the system, is named *Performance Analysis*. Here the results of the experiments are tabulated and the outcomes visualized in the form of charts and graphs. Textual comments ensuing at the end of this section leads to the formal conclusion appearing in the last section 4, appropriately named *Concluding Remarks and Future Works*, where a glimpse of further research envisaged in this domain is also included. The *Bibliographic Reference* at the end contains a list of the pioneering works which strengthened the foundation of the presented research.

## 2 Case Base Design Methodology

Historically the Case-Base Reasoning paradigm has often been utilized to mimic problem-solving techniques adopted by men. In the human society, whenever a problem occurs in any domain, expert's advice is sought either directly or from a preserved source of collected prior experience. The procedure embraces incremental learning, where an unprecedented event calls for new solutions created from existing knowledge, which in turn further enriches the learner's knowledge bank.

In the context of machine intelligence, a similar environment may be set up using a Case-Base in the backbone, where each case is a combination of a problem along with a corresponding solution. Many instances of Case-Base-Reasoners have been successfully utilized in various domains of predictive nature such as Medical Diagnosis[7][11], Market viability[6] etc. A brief outline of the underlying strategy of most such systems follows.

The actual nature of information to be stored in a case-base would necessarily be domain-dependant. The in-built reasoning part helps to extract ready solutions for pre-stored problems. If a fresh problem crops up, the system should record it as a new case and provisions should be there to add an adequate solution to it. Storing the new problem-solution pair as a separate case must be done in a manner which ensures ease of retrieval, whenever necessary. Last, but not the least, automatic quality-enhancement of the case-base should be incorporated by providing proper update techniques for the system.

In the following sub-section is discussed the procedure to build up a case-base of genuine off-line signatures. The case-base allows storing of adequate information to identify and authenticate a newly presented test signature and shortlist it for subsequent updates.

### 2.1 Image-acquisition and Preprocessing of Signature Images

There are two sets of off-line handwritten signature images that were acquired and pre-processed, designated hencefort as Dataset1 and Dataset2 respectively.

The first set, i.e. Dataset1, consisted of 2250 signatures belonging to the standard *MCYT Bimodal Biometric Database* [19] scanned at a resolution of 300 dpi with 15 genuine and 15 skilled forgeries for each of 75 persons. The signatures were obtained as gray scale images.

The second set, i.e. Dataset2, was collected, as a part of a research project, from 75 volunteers, mainly University students and their acquaintances. Each person provided twenty signatures, mostly phased out over a period of about two years. Some volunteers further offered to provide twenty forged specimens for each of the original 75 signatories. Forged signatures belonged to 5 skilled , 5 unskilled and 10 random variety for every person. The categorization [15] depended mainly on the degree of familiarity with the genuine signature, the skilled variety demanding an almost exact copy of the original, while the random one was supposed to be generated without any prior knowledge of the style of the signatory, the forger never having the chance to see the original. The unskilled variety, as the name indicates, were generally produced by an inexpert forger after a brief glance at the original.

Participants in this program were asked to sign using black ball-point pen having 0.5 micron tip points. The space demarcated for each signature was a rectangular box of 9 cm x 3 cm area, 10 such boxes being accommodated on an A4 size white bond paper. All the signatures in this second set were scanned at a resolution of 200 dpi to obtain gray scale images. Out of the twenty genuine signatures in Dataset2, ten were set aside for building up the initial case base. The remaining ten authentic signatures and the set of forged signatures were used to assess the classifier accuracy.

Standard techniques required to prepare the raw signature images prior to feature extraction are mentioned below, the details of the techniques being readily available in existing works [3] [12]:

1. *Image Binarization*
2. *Noise Reduction*.
3. *Minimal Area Cropping*.
4. *Width Normalization* .
5. *Skeletonization* .

## 2.2 Feature Extraction

The main objective of this sub-section is to describe the sets of features utilized to represent the signature images in this work. Mainly, three types of features have been considered : global features [9][4][2], 96 grid features [4] and 48 texture features [4].

***Global Features.***

Features related to the structure of the signature image as a whole are categorized as global. These features are usually extracted from the pixels that lie within the region circumscribing the signature image. Some of the advantages associated with global features are that they are easily extractable, less sensitive to noise as small distortions in isolated regions of signature does not cause a major impact on the total image, and they captivate overall idiosyncrasies of authors. They can easily reflect style variations thus providing better scope of forgery detection. The set of 20 global features extracted and used in the system are mentioned below :

***Pure Height, Pure Width, Aspect Ratio, Image Area, Signature Height, Vertical Center, Horizontal Center, Baseline Shift, Top Heaviness, Horizontal Dispersion, Maximum Vertical Projection, Maximum Horizontal Projection, Vertical Projection Peaks, Horizontal Projection Peaks, Number***

***of Edge Points, Number of Cross Points, Number of Closed Loops, Mean Ascender Height, Mean Descender depth, Interior to Exterior pixel ratio***

The outer contour of the signature needs to be extracted for most of the above features. [Figure 1].

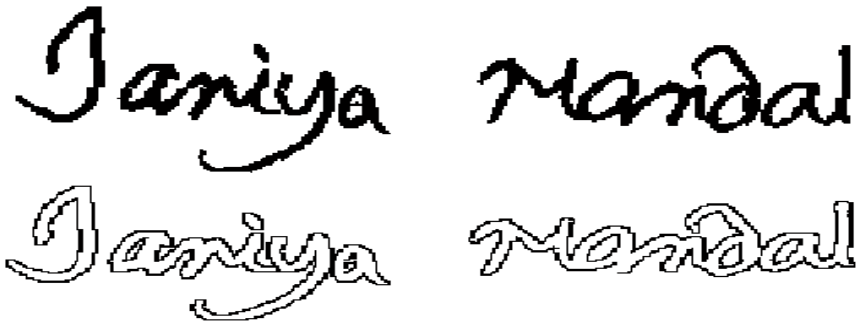

**Figure 1** Original signature and its Contour line in black

***Grid Features.***

For extraction of these features the skeletonized signature image is subdivided into twelve vertical segments, each of which are further subdivided into eight horizontal segments. The number of black pixels in each of the resulting 96 rectangular areas are counted. The lowest and highest of these counts are assumed to be 0 (zero) and 1 (one) respectively. All other segments are then normalized to a value between 0 and 1. The signature , in terms of grid features thus contains rectangular patches of white, black or different gradations of grey [Figure 2].

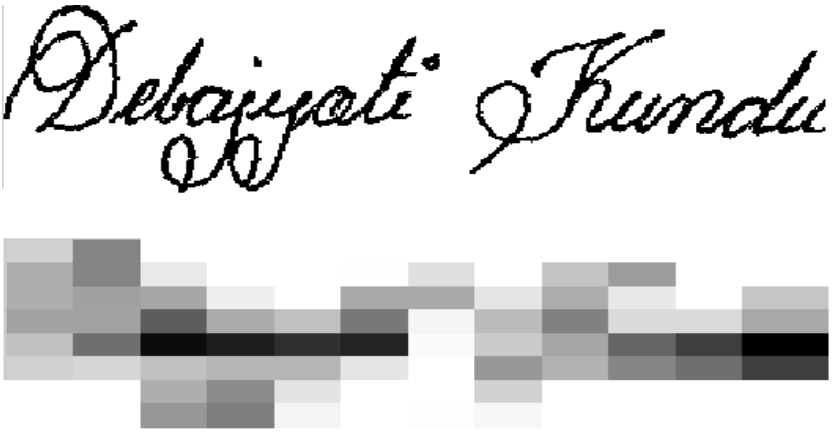

**Figure 2** Original signature and its Grid Feature segments

***Texture Features.***

To capture these features, the binarized, skeletonized signature image is subdivided into 6 rectangular regions, 3 partitions in the horizontal direction and 2 in the vertical direction. The texture information of each of these regions is calculated by considering the co-occurrence matrices of the signature image. For the binary image, this is a 2x2 matrix defined by the 1st row elements $n_{00}$ and $n_{01}$, and the 2nd row elements $n_{10}$ (= $n_{01}$) and $n_{11}$, where only the two values $n_{01}$ and $n_{11}$ are considered. The first of these, i.e. $n_{01}$ (or alternatively $n_{10}$ ), is the number of times a white pixel occurs at a distance of $d$ from a black pixel, while the other i.e. $n_{11}$, is the number of times a black pixel occurs at a distance of $d$ from a black pixel within a region. The four distances considered are $d$ = (0,1) i.e. same column next row, $d$ = (1,0) i.e. next column same row, $d$ = (1,1) i.e. next column next row, and lastly $d$ = (-1,1) i.e. previous column next row. Considering the two counts for each of these four distances within the six regions, 2 x 4 x 6 = 48 texture features are obtained for each of the signatures.

All 164 features are further discretized by mapping the total range of each individual feature values to lie within the scope of an integral number between 0 and 15. It has been experimentally verified that higher number of partitions such as 32 or 64 does not improve the result significantly. Thus, after descretization, a feature vector containing 164 hexadecimal-digits represent the signature points within the feature space. The discretization helps to eliminate noise as well as smoothen out weightage factors from all different types of features. Moreover, since signatures are highly variant in nature, even two signatures, obtained consecutively from the same person under perfect conditions, will still provide a large difference in their absolute feature values. So, this disctretization process does help to improve discerning qualities of the feature space considered.

Here each signature image $S_k$ can be represented as a feature vector, $F_k = (f_{k1}, f_{k2}, \ldots, f_{k164})$, and the Euclidean distance between two signature images $S_i$ and $S_j$ is given by

$$Dist(S_i,S_j) = \left( \sum_{m=1}^{164} (f_{im} - f_{jm})^2 \right)^{1/2} \quad (1)$$

The distance measure calculated on the basis of Equation 1 between two signatures in this discretized feature space is then utilized to compare their proximity to each other as described in sub-section 2.4.

## 2.3 DTW for signature verification

At this juncture, context demands a detour for explaining the basics of DTW techniques employed to further strengthen the similarity check between signatures in a case. DTW is a dynamic programming technique widely used in speech processing, bio-informatics and handwriting communities to match one-dimensional sequences. Two such waveforms mainly differ from one another in the degree of compression and deflation at contiguous portions, maintaining the order of variation all the while. This quality is found to be most suitable in comparing signatures, since no two signatures of a person are ever identical, yet genuine signatures retain the author's style to a large degree. Here DTW has been used to find an optimal distance between two signatures' contour lines [Figure 3].

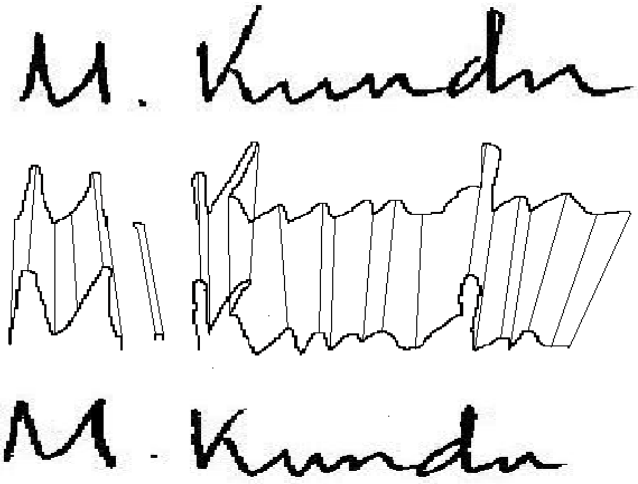

**Figure 3** Upper contours of two signatures matched by DTW technique

Plotting the cumulative distances between two time sequences $\boldsymbol{X = x_1, x_2, \ldots x_M}$ and $\boldsymbol{Y = y_1, y_2, \ldots y_N}$, in a MxN rectangular grid, the total minimum DTW distance D(M,N) as estimated in the uppermost rightmost position, by considering the following recurrence relation:

$$\begin{aligned} D(i,j) &= d(x_1,y_1), \quad \text{at } i=1 \text{ and } j=1, \\ &= D(1,j-1) + d(x_1,y_j), \text{ at } i=1 \text{ and } j>1, \\ &= D(i-1, j) + d(x_i,y_1), \text{ at } i>1 \text{ and } j=1, \\ &= \min \{ D(i,j-1), D(i-1,j), D(i-1,j-1) \} + d(x_i,y_j), \\ &\qquad \text{at } i>2 \text{ and } j>2, \end{aligned} \tag{2}$$

where $d(x_i,y_j)$ = Manhattan distance between $x_i$ and $y_j$

and $D(i,j)$ = DTW distance as recorded at position i,j

***D(M,N)*** thus gives us the DTW distance between the two sequences.

The DTW path is evaluated under the following constraints, as outlined by Sakoe and Chiba [5] :

- Monotonic condition : $x(k-1) \le x(k)$ & $y(k-1) \le y(k)$
- Continuity condition : $x(k)-x(k-1) \le 1$ & $y(k)-y(k-1) \le 1$
- Boundary condition : $x(1)=1, y(1)=1, x(K)=M, y(K)=N$
- Adjustment window condition : $|x(k) - y(k)| \le r$, where r = window width
- Slope constraint condition : neither too steep , nor too gentle a gradient for the path.

Since there are no time sequences associated with an offline signature, the contours of two signatures being compared can be considered as the sequences in this context, generated from the vertical offsets of the maxima and minima of the upper and lower contour of a signature against their horizontal positions. [1].

In the next sub-section is discussed how such DTW distance comparison values between two signatures lend itself to fine-tune the authentication process.

## 2.4 Categorization and Classification of Test Signatures

The primary objective of creating the case base is to provide the following three functionalities:-

(A) identifying the author of a presented signature,

(B) authenticating a signature once identity had been established, and

(C) upgrading the case with a better-quality authenticated test signature.

These tasks demand some threshold knowledge built into the case base. This has been achieved by saving the feature vectors of 10 authentic signatures per person at the outset. The identification process introduced here demands some pre-processing on the archived signatures, which may automatically reduce the archive cardinality below 10 by dropping weaker specimens, while generating index values for individual cases. The authentication process is further assisted by calculating and preserving some intra-person central tendencies amongst the archived specimens. Lastly the upgradation process entails automatic experience enhancement of the system by allowing addition / replacement with a newer / better authenticated specimen. Following is a description of the actual techniques utilized to facilitate all the above.

### 2.4.1 Signature Identification:

The global features are felt to be the most discerning ones while discriminating between signatures of two different persons. Here each signature of a person is represented as a transaction containing the 20 feature values in lieu of items. A coding system is adopted to represent the features by an index ranging between the numbers 0 to 9 for the first ten features, the hexadecimal digits A to F for the next 6 features between 11th and 16th, and lastly the four alphabets G,H, I and J to indicate the 17th, 18th, 19th and 20th global feature in this context. The value of each feature being discretized using 16 values within each respective range, the items in a transaction is designated as feature-codes followed by discretized values ranging between 0 and F. So a prospective item in a transaction can have a value ranging between 00 and JF at the most.

A few sample transactions *T1 to T5* are presented here as an example of five preserved signatures of a person:

***T1*** **= <0A,13,25,3F,41,55,61,72,82,9B,A1,B3,C3,DD,E7,F1,G2,H9,I0,J9>**

***T2*** **= <0A,13,25,3F,41,56,63,74,80,9B,A1,B3,C2,DD,E7,F1,G2,H7,I1,J9>**

***T3*** **= <09,13,25,3F,42,55,63,73,82,9B,A1,B2,C1,DE,E7,F1,G3,H8,I0,J8>**

***T4*** **= <0B,13,24,3F,41,55,63,75,81,9B,A1,B3,C0,DD,E7,F1,G2,H9,I2,J9>**

***T5*** **= <0A,13,25,3F,41,55,63,74,82,9B,A1,B3,C2,DC,E7,F1,G4,H9,I0,J9>**

In the next step, the algorithm introduced in [13] is used to extract the longest frequent sequence from the above transactions, assuming the minimum support-count to be 60% , i.e. 3 out of 5 here in this example. Designating the sequence as ***F***, the value of the same is given as follows:

$$F = \langle 0A,13,25,3F,41,55,63,\_\_,82,9B,A1,B3,\_\_,DD,E7,F1,G2,H9,I0,J9 \rangle$$

The feature sequence may or may not have any missing values. For example, the above sequence has two missing features, the eighth and the thirteenth one, the values for these being less frequent than the support count threshold (60%). The technique demands filling up these values by the average of the existing values for each of these features rounded off to the nearest bucket value, ie. **74** and **C2** respectively in the present context. So the transformed sequence ***F'*** is represented as

$$F' = \langle 0A,13,25,3F,41,55,63,74,82,9B,A1,B3,C2,DD,E7,F1,G2,H9,I0,J9 \rangle$$

For a feature sequence, ***F*** or ***F'***, the corresponding bit-representation of the signature for identifying the i-th person is preserved as a sequence $B_i$, containing the nibble values for each of the 20 features, in a memory-based table, serving as index to the case-base under consideration.

$$B_i = \langle A35F15342B132D712909 \rangle$$

The scheme for reducing signature specimens is next activated, and any of the specimens failing to match the index code by more than 40%, ie. 8 out of 20 feature values, is removed, for example *T3* in the instance cited above.

The feature values of the test signature form a sequence, say $B_t$. with which a binary search is made on the table of $B_i$-s. For each move in the search process, $B_t$ is compared with the corresponding $B_i$ and the count of matched bits retained until the search ends. The smallest count (ideally zero) helps to identify the nearest case. The experiments show an accuracy rate of 82.133% of successful identification for the ATVS dataset and a still higher rate of 93.8% for the indigenously collected set of data, using this scheme. The same is recorded in Table 1 of section 3.

### 2.4.2 Signature Authentication :

At the outset, each case of the preserved genuine signatures for the identified person is considered for predicting the authenticity of a test signature as described more minutely in a previous research paper [3]. While explaining the procedure, some deviations from the earlier work are pointed out highlighted in bold :

- Every authentic signature is assumed to be a point in a **164** dimensional feature-space.
- Let $\xi$ be the set of genuine specimens in a case, with say $n=|\xi|$. **The value of *n* depends on the identification process, where for each person only the signatures, tallying with the index feature values by at least the minimum support count percent, are retained for further consideration. This is an improvement from the earlier work, which had no identification module in it and hence no indexing scheme.** Distance between each pair of signatures $S_i, S_j \in \xi$ in the feature space is computed according to Equation 1 to obtain the median distance value $M_{ftr}$ and the inter quartile range $IQR_{ftr}$ .
- Similarly, the median $M_{dtw}$ and the inter quartile distance range $IQR_{dtw}$ are calculated and these four parameters along with the signatures are preserved in the case for further processing.
- When a new test signature ***T*** arrives, the median distances $MT_{ftr}$ and $MT_{dtw}$ for the test signature is obtained. The test signature ***T*** is classified as a genuine signature if it satisfies the following inequality:-

$$\alpha * DTW\_Comp + \beta * FTR\_Comp \leq 1 \quad (3)$$

where,

$DTW\_Comp = (abs(MT_{dtw} - M_{dtw}) / \gamma * IQR_{dtw})$,

$FTR\_Comp = (abs(MT_{ftr} - M_{ftr}) / \delta * IQR_{ftr})$,

$\alpha$ is the DTW similarity weight with value ranging between $0$ and $1$,

$\beta$ is the feature similarity weight, such that $\beta = 1 - \alpha$,

$\gamma$ is the allowed percentage of $IQR_{dtw}$ with value ranging between $0.1$ and $1$,

$\delta$ is the allowed percentage of $IQR_{ftr}$ with value ranging between $0.1$ and $1.5$.

### 2.4.3 Case-base Upgradation :

The upgradation process for a case-base usually entails accommodating new cases to improve the knowledge content of the base. The signature case base should not have been an exception but for two factors which affect the performance of the system adversely if the normal mode is followed without any other consideration.

The first factor is one which concerns the increased space and time complexity for proliferate growth of the case base caused by in-flow of too many new candidate samples such as for pro-active bank customers. The system needs to draw the line at some stage or the other.

The second factor concerns another important aspect - the natural process of degeneration and ageing that affects handwriting and signatures of a person. There may be other causes for changing ones style of signing as well. If the process is a continuous one, it may get unnoticed. Yet by changing imperceptibly and slowly it will cause the system to reject a perfectly legitimate signature as a fraud all of a sudden.

So the automated upgradation process in a signature case base has to allow replacement of older and/or poorer specimens. To achieve this, the first step is to create a time stamp for each signature vector $S_i$ retained for a case $k$. The time-stamp is the date portion of the real-time clock value extracted from the system when the signature vector is first created. A time index $t_k$ points to the signature with the oldest time-stamp for each case $k$.

The next step involves finding out the poorest specimen index $p_k$ in each case on the basis of the Euclidean distances of the signatures from the median vector of the case. Thus considering a case $k$ with median vector $MV_k$, the distances $D_{Si}$'s from $MV_k$ are calculated as follows :

$$D_{Si} = Dist(S_i, MV_k) \text{ for } i = 1, 2 .. 10 \quad (4)$$

$p_k$ contains the $i$ value corresponding to the highest $D_{Si}$.

Everytime a fresh authentic signature $T_C$ is detected by the system, it automatically becomes an update candidate for the case $k$ with which it is identified. The time stamp $T_{Ck}$, generated for this candidate signature, is checked against the time stamp $T_{Stk}$ for the signature with the index value $t_k$. The indexed signature is replaced by the fresh candidate signature, under the following condition:

$$(T_{Ck} - T_{Stk}) > T_{thrshld}, \quad (5)$$

where $T_{thrshld}$ is a preset time limit (1 year for Dataset2).

If the candidate signature $T_C$ does not satisfy relation (5), the next criterion must also be checked. Here the distance of $T_C$ from $MV_k$ is calculated and preserved in $D_{Tc}$ and the value checked against the distance of the poorest specimen from $MV_k$ . So the condition for replacement is given by :

$$D_{Tc} < D_{Si} \text{ for } i = p_k \quad (6)$$

In either case of replacement, the time-stamp $T_{Ck}$ and the distance measure $D_{Tc}$ of the new authentic signature, preserved in the case-base, are utilized to recalculate the time index $t_k$ and the poorest specimen index $p_k$ for the identified case. The index and central tendencies of the newly updated case need also to be recalculated and restored for future use.

## 3 Results and Performance Analysis

The experiments mentioned in this section are mainly carried out using Matlab R2013a Version 8.1. In the identification phase, starting with an experience threshold of ten authentic signatures per person in both datasets, the accuracy obtained are summarized in Table 1 below :

**Table 1** Identification Accuracy

| Dataset | Accuracy out of 75 cases |
|---|---|
| 1 | 82.133% |
| 2 | 93.8% |

During authentication, errors are calculated corresponding to the False Acceptance Rate (FAR), the False Rejection Rate (FRR) and the Total Error Rate (FAR+FRR) based on inequality (3), for a range of values of $\alpha$, $\beta$, $\gamma$ and $\delta$. The Lowest Total Error positions and the Equal Error Rate (EER) positions for both the Datasets 1 and 2 are obtained by plotting partial error curves in the relevant zone [ Figures 4, 5].

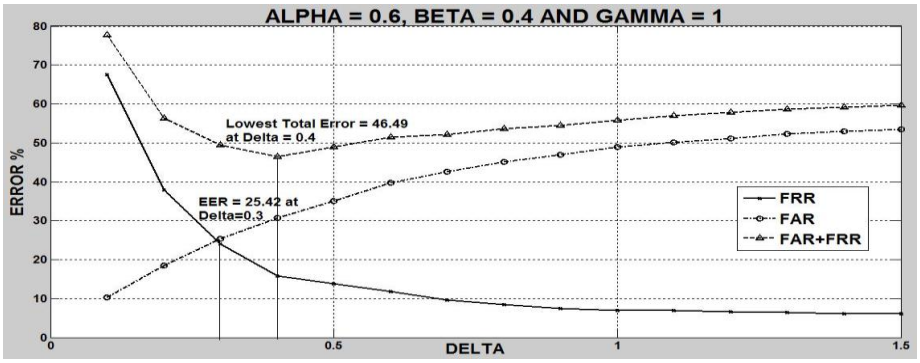


**Figure 4** FAR, FRR, Total Error and EER values for Dataset1

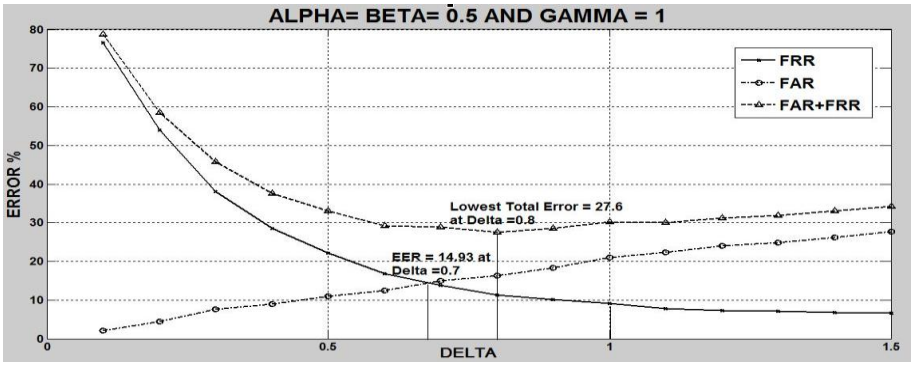


**Figure 5** FAR, FRR, Total Error and EER values for Dataset2

The ultimate values of $\alpha$, $\beta$ and $\gamma$ for both Datasets, and the $\delta$ values obtained at EER and Lowest Total Error position for both are summarized in Table 2 below.

**Table 2** EER and Lowest Error Values

| Data Set | *B=1- α,* and *γ=1.0 for both Datasets* | | | | |
|---|---|---|---|---|---|
| | ***A*** | ***EER%*** | $\Delta_{EER}$ | ***Lowest Error*** | $\delta_{LowestError}$ |
| 1 | 0.6 | 25.42 | 0.3 | 46.49 | 0.4 |
| 2 | 0.5 | 14.93 | 0.7 | 27.6 | 0.8 |

To establish the usefulness of the CBR system, its performance have been compared with that of several standard classifiers on both sets of data. Amongst these other classifiers the MLP model is built with a single hidden layer and based on a sigmoidal activation function and using the WEKA Data Mining Software [10]. The remaining classifiers include the SVM, the NB, the DT and the KNN model and are all developed on MATLAB R2013a(ver 8.1). The SVM used linear kernel and sequential minimal optimization technique for separating the hyperplane. The NB was designed having Normal(Gaussian) Distribution with uniform prior. The split criterion for DT was based on Gini's diversity index, and the KNN model was built for k = 3. The accuracy percentage of recognizing the different categories of signatures for all the classifiers are depicted as a series of barcharts in Figures 6 and 7 below for Datasets 1 and 2 respectively .

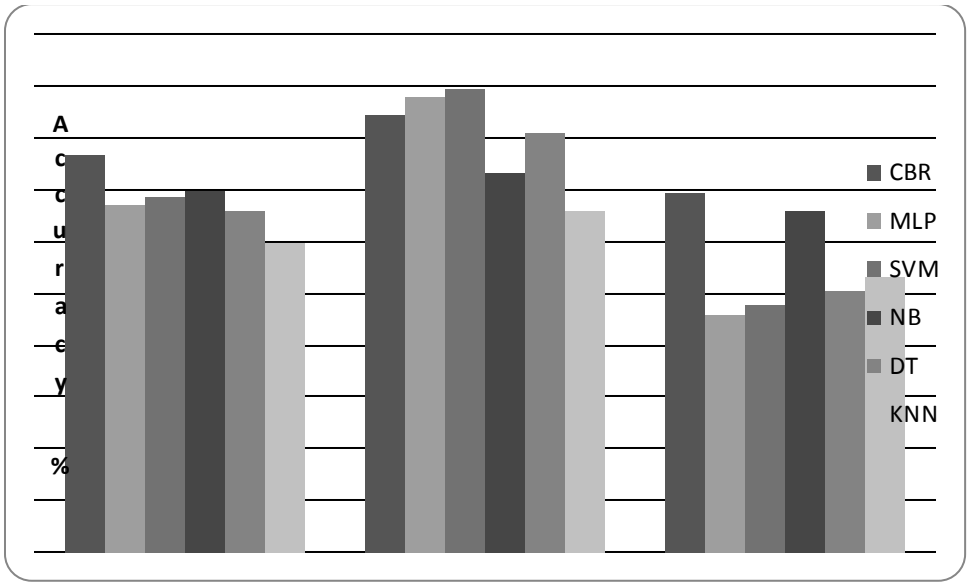


**Figure 6** Accuracy BarGraph for Dataset1

The results demonstrate a clear trend : the overall accuracy of the CBR system is better than all the other classifiers for both datasets, as verifiable from Figures 6 and 7. This system is also quite efficient in detecting forgery, and that too without prior training by forged samples, unlike all other eager learners. Both MLP and SVM which seem to perform better in recognizing authentic signatures, and the Naïve Bayesian, which appears to be the winner for the second dataset in case of forgery, are heavily dependant on such training. Only the CBR system enjoys a natural advantage in this regard.

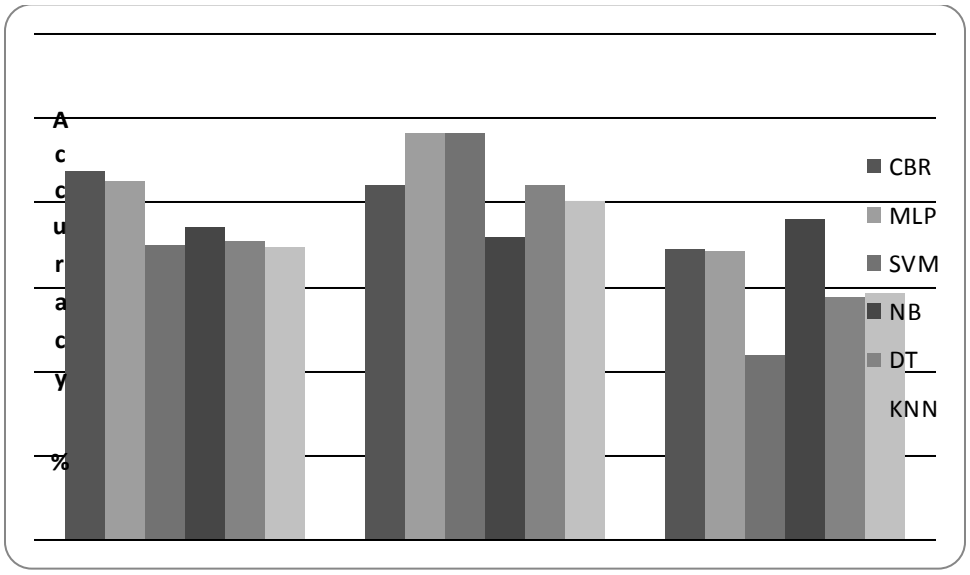


**Figure 7** Accuracy BarGraph for Dataset2

It leads one to believe that the CBR system would be a better solution in real life. The partial shortcoming of the CBR system in detecting the authentic signatures may as well be caused by the fact that the skilled and unskilled category of forgery used in Dataset2 were

mostly made by a single highly competent person. Since the MLP and SVM got a chance to train on that person's style, they could discern the authentic person's signature more accurately.

In the last phase, to assess the effect of upgradation, the efficiency of the updated system is checked with the original CBR system metrics. The results are depicted below in Figure 8.

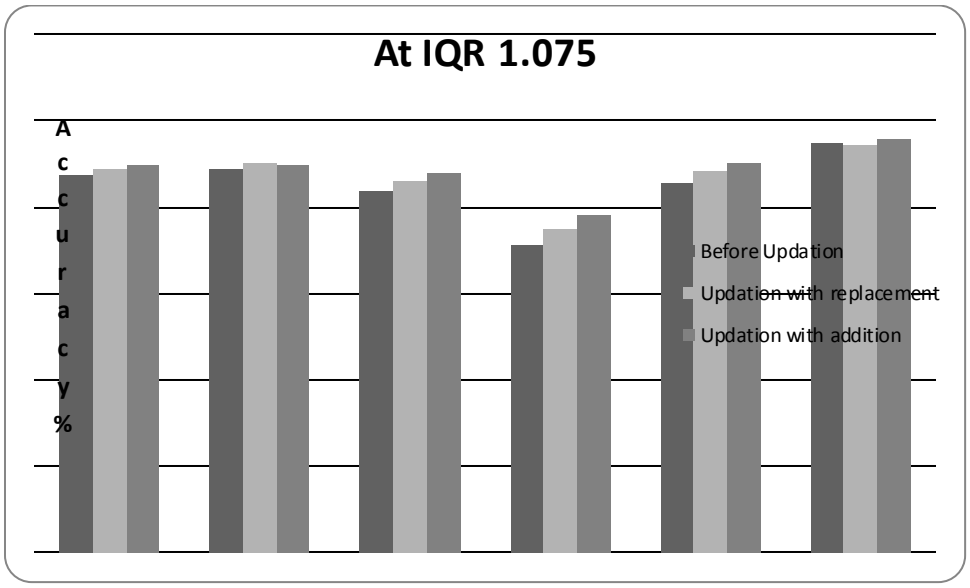


**Figure 8** Accuracy before and after upgradation for Dataset2

Here, only Dataset2 is tested as no upgradation could be affected on Dataset1 due to lack of time stamping and scarcity of test specimen. Since Dataset2 contained three variety of forged signatures, all of these (except the random signatures) benefit from being updated by more recent and/or better test signatures. Actual experiments are carried out with two types of upgradation – one where the existing cases are replaced by better quality or later updates – and the other where the case base is augmented by additional update information. The results, as plotted in Figure 8, show some improvement in both cases of upgradation – understandably better when addition is allowed.

## 4 Concluding Remarks and Future scope

The CBR system is found to perform reasonably well in all three of its proposed functionalities on the collected signature database. More specifically, the performance of both the identification and the authentication module can be improved by upgrading with recent and better signatures. The indexing technique utilized as part of identification module allows the case base to be accessed efficiently. The experimental results during the authentication phase indicate that the CBR system gives the best performance overall. Although both SVM and MLP are found to authenticate original signatures better, the SVM is far below any reasonable standard in detecting fraud. The Naïve Bayesian seems to be a better detector of forged signatures as per Dataset2, but this may have been caused by external factors.

One great advantage of the CBR system is its ability to detect fraud after training by original signatures alone. All the other eager classifiers need to be trained by forged samples as well. This poses a natural fallacy to the reasoning behind the working principle of the

practical real-life domain – how can one ask a forger to give sample signatures?

Another advantage of the CBR system is the augmentation by DTW. Since this required a direct end-comparison between two inputs, this technique can not be merged with other classifiers successfully. This particular feature is felt to possess tremendous potential in discerning the trend of writing and can be harnessed to explore better techniques based on this aspect in future endeavors.

Lastly an appreciable value-addition is achieved by the ability of the CBR system to upgrade its experience threshold automatically through preservation of the most recent and better representative specimens, thus improving its efficiency progressively. This may well be considered as an extra strength of the system and a corroboration of the virtue of incremental learning as practiced by a lazy learner such as the CBR.

A scope of improvement exists in case of identification as well as authentication, by considering only a few of the global features, instead of involving all 20. The specific global features to be utilized for this purpose may be chosen using dimensionality reduction methods such as Principal Component Analysis (PCA). A simple heuristic such as retaining only those features, which gets corroborated with the preserved index values most frequently, may also help. These attributes can be isolated by applying minimal frequent set extraction procedures hinted in [13] on the set of all persons data values. This would help to reduce both time and space complexity of the ensuing techniques. Some future work may as well be envisaged towards the formation of a better authentication module by utilizing an ensemble model which boosts up the performance of the CBR system with MLP and SVM classifiers.

## References

[1] A. P. Shanker, A.N. Rajagopalan, 2007, "Off-line signature verification using DTW", Pattern Recognition Letters 28, pp 1407-1414

[2] Alan McCabe, et al, August 2008, Neural network-based handwritten signature verification, Journal of Computers, Vol. 3, No. 8.

[3] C. Chaudhuri, et al, Dec. 20-21 2014, "Authentication of Offline Signatures based on Central Tendency of Features and Dynamic Time Warping values preserved for Genuine Cases", EAIT 2014, ISI Kolkata, pp. 256-261

[4] H. Baltzakis and N. Papamarkos, 2001, A new signature verification technique based on a two-stage neural network classifier, Engineering Applications of Artificial Intelligence 12, 95-103.

[5] H. Sakoe, S. Chiba, 1978, Dynamic Programming Optimization for Spoken Word Recognition, IEEE Transactions on Acoustics, Speech and Signal Processing,Vol.ASSP-26,No.1, pp 43-49.

[6] Hongkyu Jo, et al, 1997, Bankruptcy Prediction Using Case-Based Reasoning, Neural Networks, and Discriminant Analysis, Expert Systems With Applications, Vol. 13 No. 2, pp. 97-108.

[7] Isabelle Bichindaritz, et al, 1998, Case-Based Reasoning in CARE-PARTNER: Gathering Evidence for Evidence-Based Medical Practice, Advances in Case-Based Reasoning, Volume 1488.

[8] Janet Kolodner, 1993, Case-based Reasoning, Morgan Kaufmann Inc.

[9] Kai Huang and Hong Yan, 1997, Off-line signature verification based on geometric feature extraction and neural network classification, Pattern Recognition, Vol. 30, No. 1, pp. 9-17.

[10] Mark Hall, et al, 2009, The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[11] Mu-Jung Huang, et al, 2007, Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis, Expert Systems with Applications 32 856–867.

[12] Nobuyuki Otsu, 1979, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics. 9 (1): 62–66.

[13] Pinaki Mitra and Chitrita Chaudhuri, 2006, "Efficient algorithm for the extraction of association rules in data mining", M.Gavrilova et al. (Eds.) : ICCSA 2006, ©Springer-Verlag Berlin Heidelberg, LNCS 3981, pp 1 – 10.

[14] S. Chen and S. N. Srihari, August 2005, Use of Exterior Contours and Word Shape in Off-line Signature Verification, *Proc. International Conference on Document Analysis and Recognition*, Seoul, Korea, pp. 1280-1284.

[15] Sanjay N. Gunjal, Manoj Lipton, November 2011, Robust Offline Signature Verification Based on Polygon Matching Technique, International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 1, Issue 1.

[16] Sankar K. Pal, Simon C.K. Shiu, 2004, Foundations of Soft Case-Based Reasoning, John Wiley & Sons, Inc.

[17] Shankar A.P and A. N. Rajagopalan, Off-line signature verification using DTW, Pattern Recognition Letters, Vol.28,2007, pp. 1407-1414.

[18] Yoshimura, M.,Yoshimura, I., 1997, An application of the sequential dynamic programming matching method to off-line signature verification,Lecture Notes in Computer Science. In: Proc. of First Brazilian Symposium on Advances in Document Image Analysis 1339. pp. 299–310.

[19] http://atvs.ii.uam.es/mcyt75so.html ( ATVS - Biometric Recognition Group >> Databases >> MCYT - SignatureOff - 75).