

# Conventional and Computational Features in Document Examination

Saini M<sup>\*1</sup> and Kapoor AK<sup>2</sup>

<sup>1</sup>UGC-JRF, Department of Anthropology, University of Delhi, Delhi, India

<sup>2</sup>Professor, Department of Anthropology, University of Delhi, Delhi, India

\*Corresponding author: Saini M, UGC-JRF, Department of Anthropology, University of Delhi, Delhi, India  
110007, E-mail: mini.1901@yahoo.com

Citation: Saini M, Kapoor AK (2015) Conventional and Computational Features in Document Examination. J Forensic Sci Criminol 3(3): 301. doi: 10.15744/2348-9804.3.301

Received Date: April 14, 2015 Accepted Date: June 26, 2015 Published Date: June 29, 2015

## Abstract

Document examination has been around for more than a century. The field of Document examination has become more diverse and requires authenticity or validation in many areas of examination which involves determination of authorship, fraud detection and personal identification. Forensic document examination branch is continuously facing challenges due to the availability of revolutionary computer methods. The conventional or traditional features used in document examination are prone to fatigue and human error owing to non-validation. Computational approaches of the Document examination formalize the conventional methods by means of precisely stated algorithms. This paper efforts to give a comparative account of conventional and computational features used in Document examination. Observable differences were seen in selected handwriting characteristics between conventional and computational methods.

**Keywords:** Document examination; Fraud detection; Personal identification; Conventional Features; Computational Features

## Introduction

The scientific examination of Document has been intrinsic part of forensic science for hundred years. This examination is mostly consists of analysis and comparison of handwriting and signatures. A number of techniques have been developed in that period to: (a) Identify the individual through their handwriting pattern (For the forensic or non-forensic purpose) (b) Detect whether the handwriting is forged (process of making fake documents by alteration) or not (c) Determine the origin and history of documents.

A number of conventional tools are used by forensic document examiners for the examination of handwriting. These tool and techniques can be categorized into following categories:

**Basic Measuring tool:** It includes metric rulers, calipers and glass alignment plates for the measurement of angles, height, width and spacing of handwriting.

**Magnifiers and Microscope:** Handheld magnifiers and comparison microscope.

**Lighting sources of different wavelength:** Transmitted light, long and short wave ultraviolet light and infra-red light.

**Special Instruments:** To reveal text from intended impressions document examiners use electrostatic detection device, ESDA (electrostatic detection apparatus) which applies charges and toner to visualize areas of indented writing, making them visible to the eye.

To detect writing that has been added with a different ink, or has been altered or removed by exploiting variations, an imaging instrument video spectral comparator (VSC) is used. This instrument applies radiation filtered at different wavelength on handwriting that enhances and make the handwriting content more visible. In recent years a new area of potential document examination has developed around computer. The image processing software offers advantages to analyze handwriting in more attractive way. These systems also narrow down the search by comparing the questioned documents with the dataset of documents which a system has received from known writers and provide visualization to assist Document examiners.

## Conventional Versus Computational Features

Features are quantitative measurements that can be obtained from a handwriting sample in order to obtain a meaningful characterization of the writing style. These measurements can be obtained from the entire document or from each paragraph, word, or even a single character [1]. Two types of features can be used in Document examination:

**Conventional Features:** These features are primarily used by forensic document examiners. These features can be measured both quantitatively and qualitatively and can be grouped into twenty-one discriminating elements of handwriting. A discriminating element is defined as “a relatively discrete element of writing or lettering that varies observably or measurably with its author and may, thereby, contribute reliably to distinguishing between the inscriptions of different persons, or to evidencing the sameness in those of common authors” [2]. These twenty –one elements are grouped into four categories:

**Element of Style:** Arrangement, Class of allograph, Connections, Designs of allographs and their construction, Dimensions, Slant or Slope, Spacings.

**Element of Execution:** Abbreviations, Alignment, Commencements and Terminations, Diacritics and Punctuation, Embellishments, Legibility or Writing Quality (Including Letter Shapes), Line Continuity, Line Quality, Pen Control, Writing movement.

**Attributes of all Writing Habits:** Consistency or Natural Variation, Persistency.

**Combinations of Writing Habits:** Lateral Expansions, and Word proportions.

**Computational Features:** The establishment of scientific basis for handwriting examination has been approached through the development of a computational theory. A computational theory is an application of computer vision which consists of representations, algorithms and implementations.

Computational Features are computed algorithmically for e.g. Software operation on scanned/digital image of handwriting. These features are quantitative and remove subjectivity from the process of feature extraction. All the conventional features can be converted into computational features eventually if correct algorithms are defined. A number of researches have been done in which computational features are used for the handwriting recognition [3-5]. Handwriting recognition is the process of recognizing the content of handwriting by transforming the written input presented in its spatial form of graphical marks into its symbolic representation through computer software.

Handwriting recognition is different from handwriting identification in that they are two opposite processes. The objective of handwriting recognition is to filter out individual variability from handwriting and recognize the message. The objective of handwriting identification is to capture the essence of the individuality, while essentially ignoring the content of the message. The two share many aspects of automated processing, such as determining lines, strokes, etc. [1].

In the present study both conventional and computational features have been used in the document examination which is a novel method. So far, no work on document examination based on the combination of both features has been introduced in the public literature. In addition to it, some advantages of computational features over conventional features have also been discussed.

## Methodology

The present study was conducted in Delhi on two population groups, namely Brahmin and Punjabi (Khatri and Arora). These population groups have shown a higher rate of literacy and economy as compared to the other population groups of Delhi. A total of 250 handwriting sample were collected for the present study. Out of 250 samples, 112 samples belonged to Brahmin population group and 138 were of Punjabi population group. The age of the subjects among Brahmin population group ranged from 16-42 years whereas among Punjabi group ranged from 16-46 years.

All the subjects were asked to copy the source document in his/her most natural handwriting. The source document consists of all capital and small letter alphabets, punctuation marks, 0-9 numerals and some handwriting characteristics of interest. All the subjects were provided with uniform writing material and writing instrument (Plain unruled paper pad and black Reynolds ball pen). Both conventional and computational features have been used for the analysis. In the present study five handwriting features were selected for experimental purpose. The following table presents the studied handwriting features and methods used for examination (Table 1).

Handwriting Features	Conventional (Manual)	Computational (Automated)
Slant	Protractor and Visual examination	Inverse tangent angle
Pressure	Visual examination	Grey-level threshold Value
Writing Connectivity	Visual examination	Number of Interior Contours
Height	Ruler	Euclidean Distance
The lower loop of alphabet 'y'	Visual examination	Template Matching

**Table 1:** Methods used for Handwriting Examination

### Conventional Features

**Pen Pressure:** The type of grip pressure used to hold the writing instrument is an important characteristic of writer identification. If the writer puts heavy pressure on the writing instrument, the writing will appear cramped and rigid. If the pressure is light, the pen will slide around and leave air strokes and stray marks on the paper. Unskilled writers tend to hold the pen too tightly, limiting their control of the writing instrument. In the present study pen pressure was classified into five categories on the basis of visual observation of imprints left on the underlying pages: Heavy, Slightly heavy, Light, Slightly light and Normal. Pen pressure was examined by observing stroke width and indentation marks.

**Slant:** Slant is the angle of a letter in relation to the baseline. It is also known as slope and measured with the help of a protractor (Figure 1). The angle is measured from the baseline to the top of the letter on the upstrokes above the baseline. In the present study, the measured slant was classified into following corresponding categories:

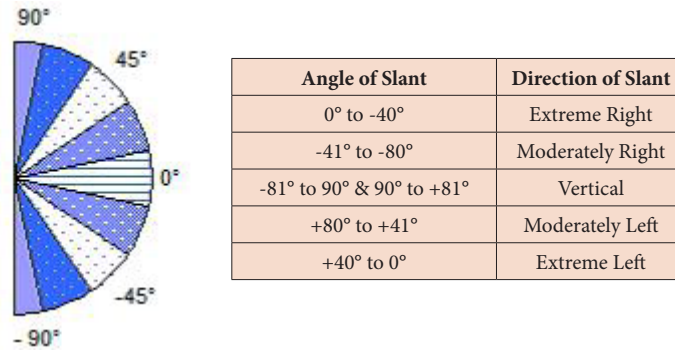


Figure 1: Classification of Slant into corresponding categories

**Handwriting Connectivity:** There are three styles of handwriting connectivity: Cursive, Printed, Cursive-printed hybrid. Handwriting is considered to be in cursive form when letters are connected with each other. On the other hand in printed writing the letters of one word are not joined. In third category, writers write in hybrid form where some elements of script resemble cursive- writing and other resemble printed writing. Handwriting connectivity is examined by visual inspection of the handwritten script.

**Height of Handwriting:** Height of handwriting is an important characteristic of handwriting examination. Handwriting is composed of three zones: upper, middle and lower zones. The average height of handwriting is 9mm. Capital letters e.g. 'F' occupy all three zones of the handwriting. In the present study height of the handwriting is measured from capital letters (Figure 2).

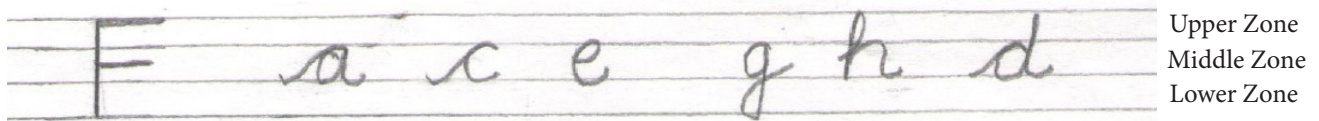


Figure 2: Three zones of handwriting

**The lower loop of alphabet 'y':** The lower loop of a 'y' letter descends below the baseline and returns to the baseline. It shows a wide variety of variations among different writers. These loops were analyzed visually with the help of a magnifying glass (Figure 3). In the present study 'y' loops has been classified into eight categories:

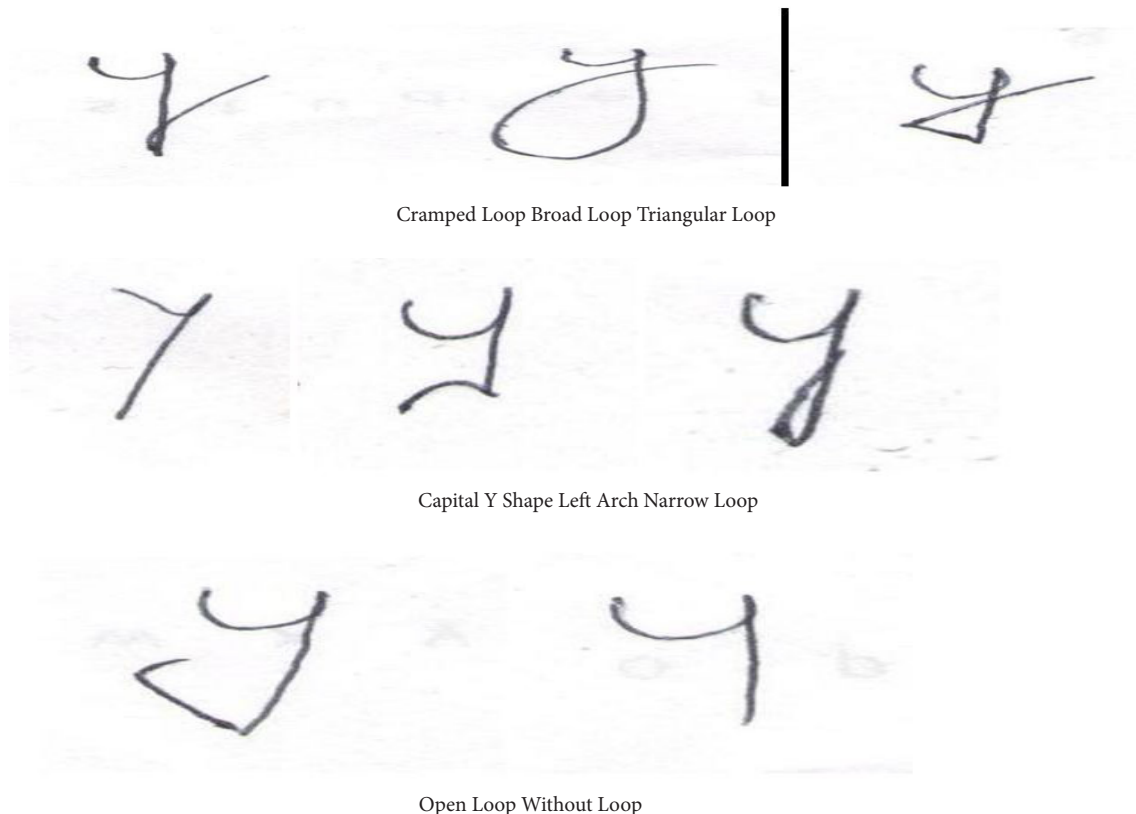


Figure 3: Different types of Lower loop 'y' alphabet

## Computational Features

For the computational analysis each of the collected handwritten samples were digitally scanned through high resolution scanner at 600 dpi (dots per inch). After all the handwritten documents were digitally scanned, noise removal was done to remove noise from the scanned handwritten image. In the present study, median filters (nonlinear method) were used for noise removal. Computational features were extracted with the help of MATLAB 8.3 (high-level language and interactive environment for numerical computation, visualization, and programming) software. These features are as follows:

**Writing Pressure:** The handwriting pressure was determined by the grey-level threshold value. The scanned image (RGB) was converted into binary image by threshold algorithm. This algorithm measures the grey-level pixel values in the image that are below a particular threshold to pure black (foreground) and those above the threshold to pure white (background). The threshold value is a measure of writing pressure, where higher values are indicative of light pressure and lesser values indicate heavy pressure.

**Slant:** Slant of the handwriting is computed by measuring the angle of the letters with the baseline, the ginput command stores the value of respective x and y coordinates as a matrix in a defined variable A and B, the elements of the matrix for different x and y coordinates were computed to get the angle of slant. It measures the inverse tangent angle using the following equation:

$$\tan \theta = \frac{y_2 - y_1}{x_2 - x_1}$$

The overall slant of writing was taken as the average of all *angles* of all the line elements i.e. letters and numerals with vertical shaft e.g. B, D, E, H, I, 1, 4 etc.

**Handwriting Connectivity:** The thresholded binary images were further processed to measure the connected components in the handwritten image. The boundary or contour of each connected component were stored and manipulated (Figure 4). A binary image of a line of text from the handwritten image and the corresponding contour image are shown in Figure 1.

The average number of connected components can be used as a measure of writing connectivity. The obtained number of was connected components classified into three categories: (a) Cursive (b) Printed-Cursive hybrid (c) Printed (Figure 5).

Printed handwriting has a greater number of connected components. Examples of connected components for two handwritten sample are shown in Figure 2.



Figure 4: (a) Threshold Binary Image (b) Corresponding Contour Image





(b)

Figure 5: Handwriting Connectivity: (a) Number of connected components = 23 (b) Number of connected components = 4

**Height:** In the computational approach, height of the handwriting was computed using Euclidean distance. The Euclidean distance between point's p and q is the length of the line segment connecting them (pq). Euclidean distance formula, the distance between two points in the plane with coordinates  $(x_1, y_1)$  and  $(x_2, y_2)$  is given by:

$$\text{distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

**The lower loop of alphabet 'y':** The correlation based template matching technique was used for automatic recognition of 'y' loop. In this technique nine predefined templates were created. A template represents a shape pattern, and the relationship between two shape patterns is captured by the relationship between templates, which reflects the probability of their co-occurrence in the same image [6]. The process of template matching using correlation consists of the following steps: (1) acquiring a two dimensional array of pixels, (2) locating an unknown character in the two dimensional array, (3) computing the correlations between the unknown character and every member of a trained set of characters (otherwise known as a font), (4) recognizing the unknown character as the trained character with the highest associated correlation coefficient above a threshold [7]. The minimum value of the correlation represented the best match.

### Results

The data obtained through conventional and computational methods were analysed statistically using SPSS Version 16.0 for Windows to find out the significant differences and correlation between these two methods. To make comparisons between quantitative and qualitative values, quantitative data has been classified into different categories in case of pressure, connectivity and 'y' alphabet lower loop examination.

Table 2 shows the mean, standard deviation and t-value for handwriting slant angle. The difference in the mean values of slant was found to be non-significant. The direction of the slant was found to be similar in both methods (Table 2).

Handwriting Global Slant Angle (N = 250)	Conventional Method	Computational Method	t-value
	Mean ± sd.	Mean ± sd.	
	81.80 ± 5.75	81.66 ± 5.58	

Table 2: Handwriting Slant measured using Conventional and Computational method

Table 3 depicts the correlation of conventional handwriting height with computational height. The relationship between these two methods for handwriting height was found to be positive and significant (0.922\*\*) at the same level (Table 3) (Figure 6).

	Conventional Height	Computational Height
Conventional Height	1	0.922**
Computational Height	0.922**	1

\*\* Correlation value is significant at 0.01

Table 3: Correlation between Conventional and Computational Handwriting Height

The classification of pen pressure, connectivity and 'y' loop was done in the same way regarding both conventional and computational methods (Figure 9). The cut- off values of these parameters for computational method is given as:

**Pen Pressure:** The obtained grey-threshold values were classified as follows: (a) Heavy: Below 0.7000 (b) Slightly Heavy: 0.7000-0.7200 (c) Normal: 0.7201-0.7500 (d) Slightly Light: 0.7501-0.7600 (e) Light: Above 0.7600 (Figure 7).

**Handwriting Connectivity:** The obtained number of connected components was classified as follows: (a) Cursive: 1-10 (b) Cursive-Printed Hybrid: 10-15 (c) Printed: 15-25 (Figure 8).

### Discussion

The present study presents that conventional methods of handwriting examination employ many reasonable but scientifically unproven techniques. These methods are not very effective in measuring the minor peculiarities of handwriting in terms of pressure and character recognition. The study depicts that visual examination (quantitative features) of handwriting features are subjected to human error because of non-availability of tools to measure these features efficiently. On the other hand, quantitative conventional features are also based on uncertainty for evaluating handwriting comparison. Computational approaches of the handwriting examination overcome these problems with a scientific basis and formalize human expert-based approaches. The computational theory also has the advantage of repeatability, i.e., the same results are obtained when applied to the same documents as opposed to expert human document examiners who are using conventional methods [1].

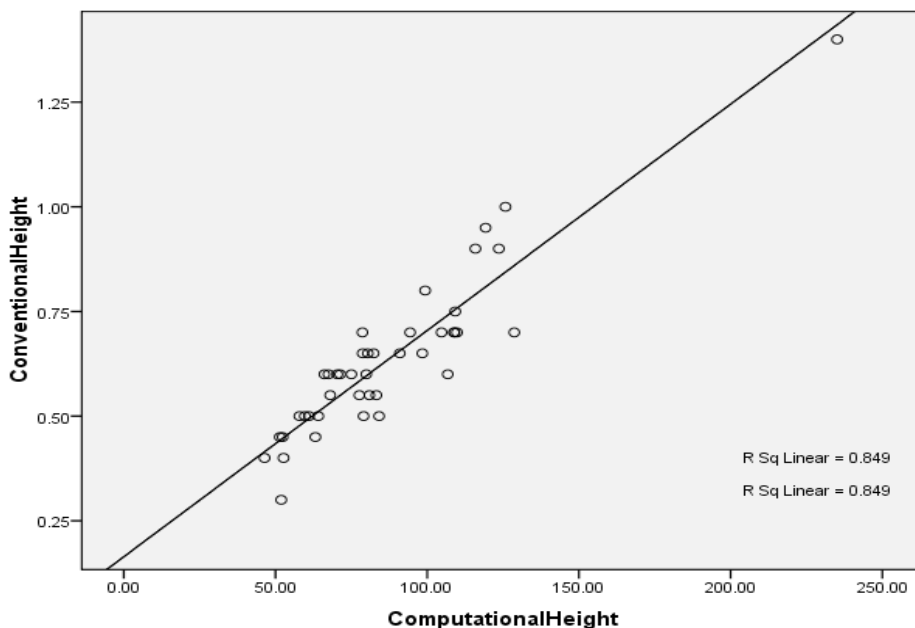


Figure 6: Relationship between Conventional and Computational Handwriting Height

Figure 5-7: Conventional and Computational Methods for Pen Pressure, connectivity and 'y' lower loop examination

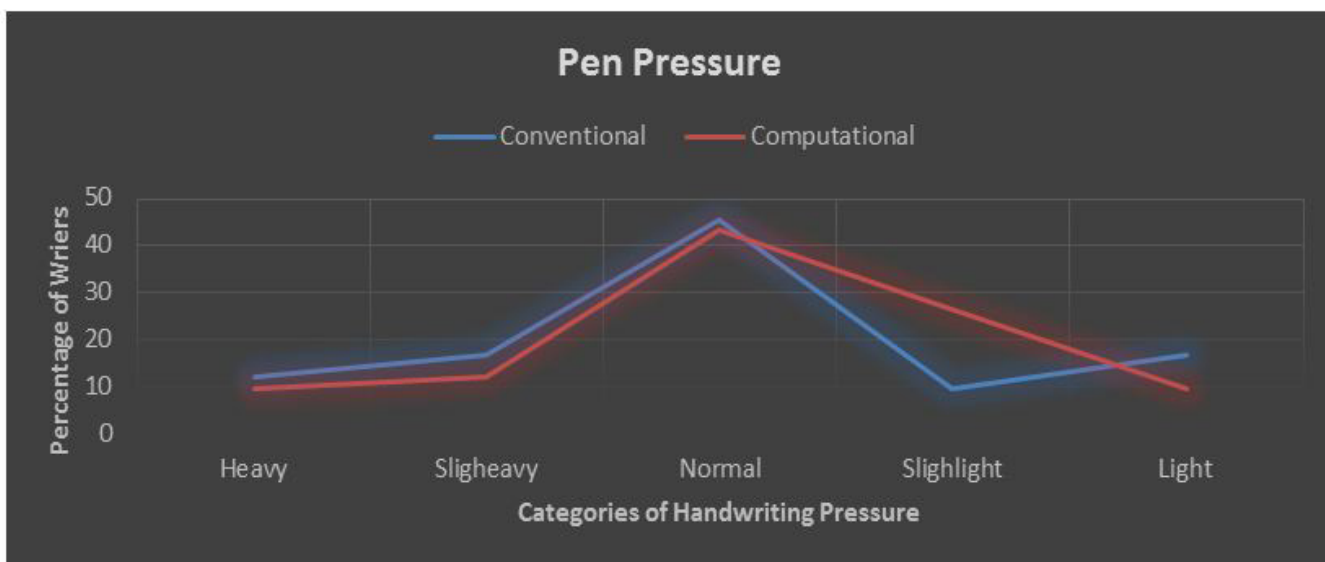


Figure 7

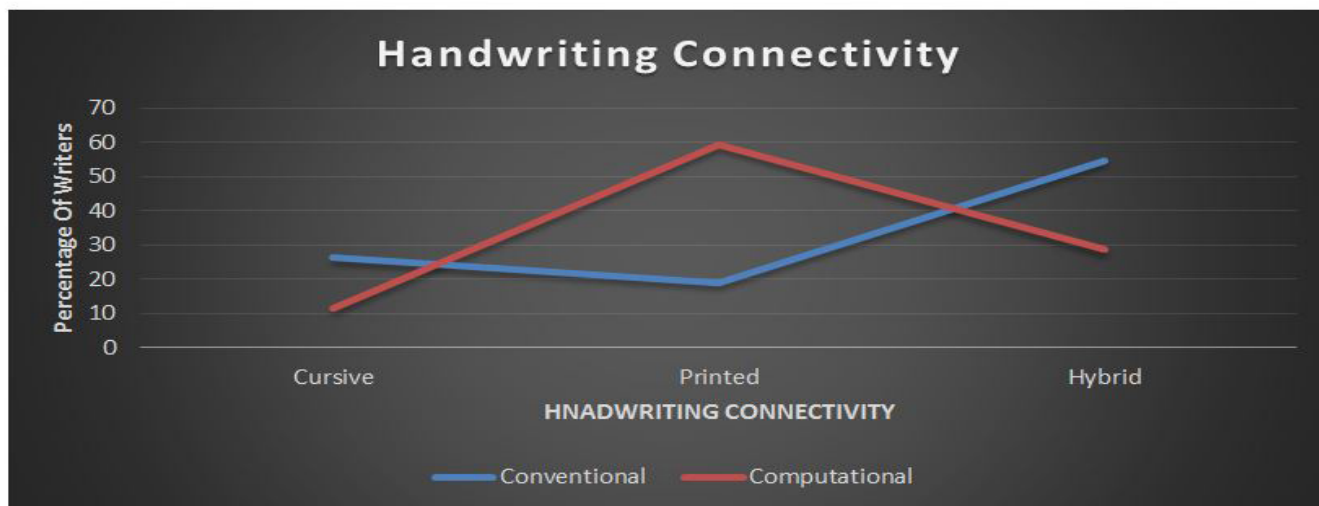


Figure 8

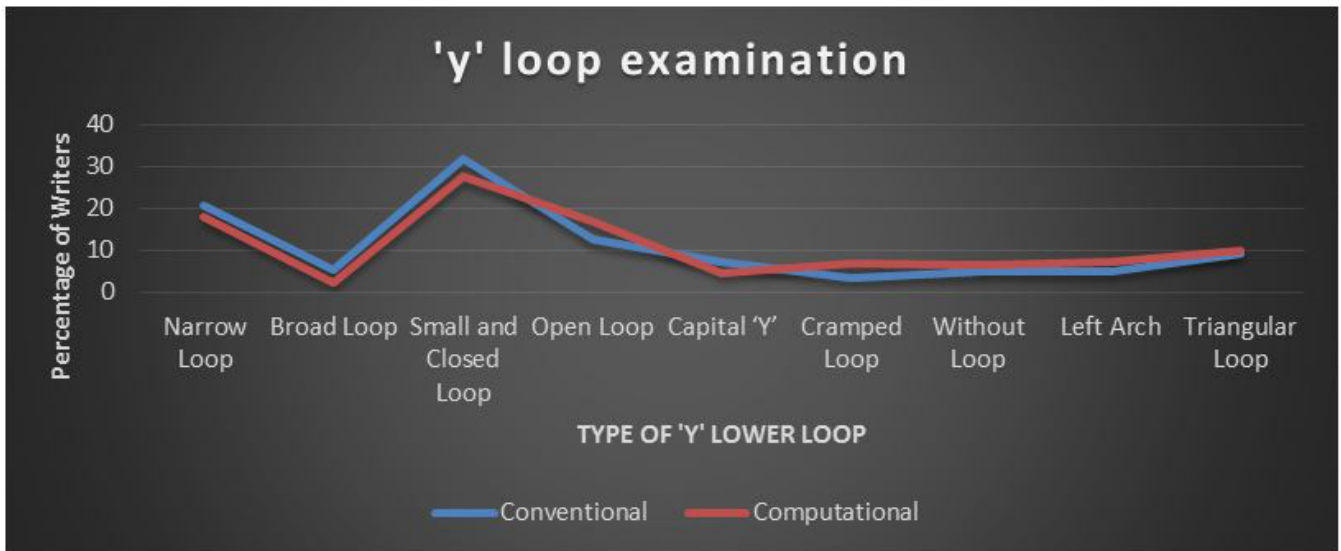


Figure 9

**Figure 7-9:** Presents the conventional and computational methods for Pen Pressure, connectivity and 'y' lower loop examination. Observable differences were seen in case of light pressure and printed script measurement through conventional and computational methods

## Conclusion

A method has been proposed to examine handwriting from conventional and computational features. Computational methods of Document examination offer the promise of validating conventional methods but these computational methods are still at an initial stage. Many of the conventional features e.g. line quality, rhythm are too subjective to be validated through algorithms. Therefore, Computer assisted Document examination needs to be more strengthened for complete validation of handwriting examination and general acceptance by forensic document examination community.

## Acknowledgement

The author heartily acknowledges her research supervisor for his persistent guidance and support in this research work. The author is also thankful to University Grant Commission for providing financial assistance in the form of Junior Research Fellowship and equally grateful to all the subjects for their cooperation in data collection.

## References

1. Srihari SN, Cha SH, Arora H, Lee S (2002) Individuality of handwriting. J Forensic Sci 47: 856-72.
2. Huber RA, Headrick AM (1999) Handwriting identification: facts and fundamentals. CRC Press, USA.
3. Hu J, Brown MK, Turin W (1996) HMM based online handwriting recognition. IEEE Trans Pattern Anal Mach Intell 18: 1039-45.
4. Marti UV, Bunke H (1999) A full English sentence database for off-line handwriting recognition. IEEE Docum Anal Recog ICDAR'99 Proceedings of the Fifth Inter Confer 705-8.
5. Srihari SN (1989) Feature extraction for locating address blocks on mail pieces. From Pixels to Features 261- 75.
6. Yang S, Bo L, Wang J, Shapiro LG (2012) Unsupervised template learning for fine-grained object recognition. Adv Neural Infor Process Sys 3122-30.
7. Martin GJ, Chen RY, Cook JF, Mason JA (1992) Character recognition method using correlation search US Patent No. 5,150,425. Washington, DC, USA.

Submit your next manuscript to Annex Publishers and benefit from:

- ▶ Easy online submission process
- ▶ Rapid peer review process
- ▶ Online article availability soon after acceptance for Publication
- ▶ Open access: articles available free online
- ▶ More accessibility of the articles to the readers/researchers within the field
- ▶ Better discount on subsequent article submission

Submit your manuscript at

<http://www.annexpublishers.com/paper-submission.php>