International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 11, November - 2013

Automated Tool Based on Statistical Study for Predicting Probability of Female Authorship

Kriti Nigam[1], Vaibhav Saran[1], Syeed Ahmed [2], A. K. Gupta[1]
[1]*Department of Forensic Science, SHIATS, Allahabad, U.P., India*
[2]*Central Forensic Institute*
Bhopal, M.P., India

Abstract

Establishing gender of writer in suspected document is a critical task and is often encountered by the forensic document examiners. The Present paper emphasizes on estimating the probability of female authorship based on statistical validation of features observed in 200 handwriting samples of female writers. The frequency of features in samples was computed and relative probability was calculated. A simple tool is proposed here which computes the input values given by user and suggests the probability of female authorship. The tool was validated by 50 unknown samples and found to be 83.2% accurate.

I. INTRODUCTION

Huber (1999) defined Handwriting as a complex motor skill which is a combination of sensory, neurological and physiological impulses. Factors such as visual perception and acuity, comprehension of form, central nervous system pathways, and the anatomy and physiology of the bones and muscles of the hand and arm all combine to produce the desired output [1] [2].

Handwriting is affected by various internal and external features like handedness, gender, age, educational level of the writers. However no significant classification has been proposed which classify the handwriting on such traits.

Classification of offline handwriting based on various attributes like gender, handedness, age and educational level has a vital role in forensic document examination, it may help in minimizing the suspect list and hence the accuracy of positive identification in less time could be achieved. Document expert rely on visual observation of handwriting features and sometimes their manual measurements, a computational approach which may measure the desired features of handwriting may assist the document expert in examination of document with higher accuracy level within less time consumption. Previous studies have reported many computational methods which are based on high level programming and personal models which are majorly based on optical character recognition, and no such classifier is proposed which may serve the document expert to categorize the handwriting into certain groups. Different character (micro-features) have been found to be powerful for handwriting discrimination, however, their capability in identifying the gender/handedness/age has not been evaluated yet [3].

The present work is based on development of a simple tool for checking the probability of writer being female. In the present paper an attempt has been made to identify various discriminating features of handwriting in respect to gender and out of these features, to sort out the most reliable features. The results are based on statistical analysis of frequency of features observed in handwriting of a dataset of 200 writers.

II. METHODOLOGY

Handwriting Samples of male and female were randomly collected from peoples of Allahabad of age range between 17 to 45 years old without letting them know the purpose of the study. They were seated in a comfortable condition and asked to complete the handwriting specimens. The participants were requested to copy the control passages in the unlined space of the same make of paper using the pen provided. They were required to copy the passage for in handwriting specimen forms in their normal handwriting. The writing was preserved for further analysis. The digital image processing of the scanned handwritten samples were performed in several steps as followed:

A. Image Acquisition- The handwriting samples were acquired in a digital format with the use of a High resolution scanner at 600 DPI using hp Scanjet 2400 Series scanner and a dataset of 200 offline handwritings is prepared by compiling the scanned handwritten images and storing in memory of the computer.

B. Noise Removal- The noise appeared during image acquisition resulting from a number of factors, which affect the intensity of actual image, is removed using Gaussian filters [4].

C. Image segmentation- Noise free image is subjected to Otsu method for the segmentation of the preprocessed image which chooses the threshold to minimize the intra class variance of black and white pixels.

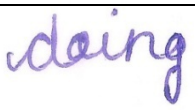
D. Feature Selection- The image is further segmented to extract the features mentioned in Table 1 and the frequency of these features were statistically computed for F value at 5 % level of significance and compared with the tabulated value to prove the alternate hypothesis that there is significant variation in male and female handwriting. The features which showed maximum significance were selected to calculate the probability level of the writer, being female. The selected features as shown in Table 2 were then framed to use in the Program designed to calculate the probability of feminine writing [4].

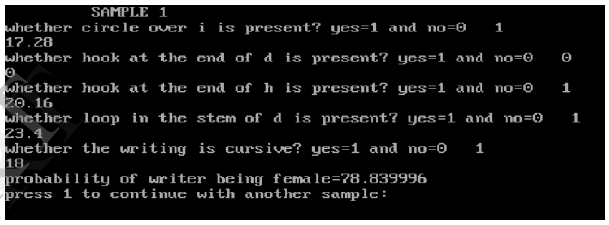
Table 1: Total Features Studied for Gender based Classifier

Features	P_1	P_2	Fcal
Circle over i	48	15	5.030
Cramped counter of a	25	43	5.29
Cramped counter of d	17	31	2.32
Cramped counter of o	5	20	3.21
Fullness of counter of a	48	23	3.69
Fullness of counter of o	36	15	3.40
Open counter of a	16	30	2.35
Hook at beginning of c	63	49	1.99
Hook at end of u	78	57	3.17
Hook at end of e	78	56	3.313
Hook at end of d	56	23	4.77
Hook at end of n	42	23	2.8
Hook at end of h	56	26	4.316
Hook at end of c	77	56	3.14
Loop in stem of h	49	34	2.17
Loop in stem of l	84	66	2.94
Loop in stem of b	10	2	2.38
Loop in stem of d	65	35	4.24
Flourish start of c	16	5	2.38
Flourish start of d	6	0	3.64
Flourish start of a	10	2	2.48
Flourish start of o	6	0	2.48
Flourish start of u	6	0	2.48
Consistency in angle of crossbar in t	63	49	1.99
Consistency in angle of crossbar in f	66	50	2.29
Upward flourish at end of s	6	1	2.0
Dispersive writing	23	42	2.87
Cursive writing	50	24	3.81
Arched form of r	59	37	3.12
Parochial form of r	27	47	2.93
Knot formation in w	43	29	2.08
Straight down stroke in y	9	21	2.4

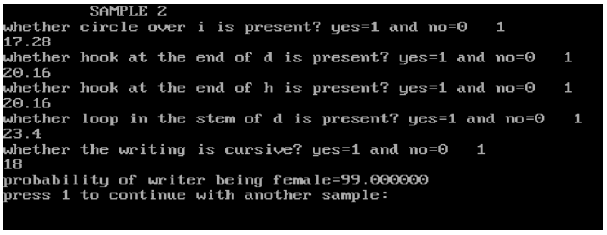
At 5% degree of freedom, Ftab = 1.96

Table 2 Features Selected for Automated Tool

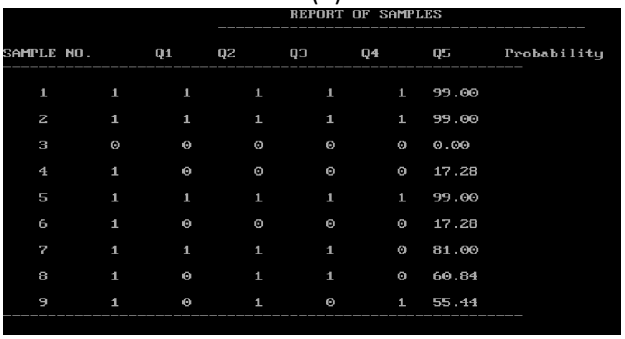
Sr. No	Feature		Probability of Being Female
1	Circle over i		17.28
2	Hook at end of d		20.16
3	Hook at end of h		20.16
4	Loop in stem of d		23.4
5	Cursive writing		18.0



(a)



(b)



(c)

Figure 1 (a), (b), (c) Screen shots of tool developed to check probability of female authorship

III. RESULT

The proposed tool is based on the observation of features in handwriting samples, the examiner carefully examines the handwriting sample and enters the value 0 or 1 for the presence or absence of the features in that sample respectively. The tool then computes the values and displays the overall probability of the writer being female. The tool was formulated on statistical observation of 500 samples and them validated by 50 unknown samples. The False Acceptance rate (FAR) was highest in cases of cursive writing 14% and circle over i 11% whereas the False Rejection Rate (FRR) was maximum in case of cursive writing 11%. The tool has showed an accuracy of 83.2 %

Features	No of Samples	False Acceptance Rate	False Rejection Rate	Accuracy
Circle over i	200	11%	9%	80%
Hook at end of d	200	5%	7%	88%
Hook at end of h	200	6%	7%	87%
Loop in stem of d	200	6%	8%	86%
Cursive writing	200	14%	11%	75%

IV.CONCLUSION

The above mentioned features proved to be highly significant in differentiating the female writings. Along with this each feature is associated with its fixed probability value on the basis of their relative occurrence in the samples analyzed. Thus, a combination of different features will give a different probability of the handwriting being feminine. The tool proposed in this paper will help the forensic document examiners to estimate the gender of the writer of any anonymous document in a comparatively lesser period of time with greater accuracy.

References-

[1].Huber, R. A. and Headrick A. M. *Handwriting Identification: Facts and Fundamentals*. CRC Press-Taylor & Francis, Boca Raton, Florida, 1999, pp. 10–14.
[2].Hilton, O. *Scientific Examination* of Questioned Documents. Elsevier, New York, 1982, pp. 10, 17, 153–157, 174.
[3]. Srihari, S. N., Cha, S.-H., Arora, H., and Lee, S. Individuality of handwriting, *Journal of Forensic Sciences* (2002) 47:856–872.
[4] Saran, V, Kumar, S, Ahmed, S, & Gupta, A., K., Computational method for forensic verification of offline signatures, (2013), *IOSR Journal of Computer Engineering*, ISSN: 2278-0661, Volume 14, Issue 2 (Sep. - Oct. 2013), PP 81-83.