

Automated Tool Based on Statistical Study for Predicting Probable Medium of Education of Writers

¹Kriti Nigam, ¹Ankit Srivastava, ¹Vijay Kumar Yadav and ²Sayed Ahmed

¹Assistant Professor, ²SSO-I

¹Dr. APJ Abdul Kalam Institute of Forensic Science & Criminology
Bundelkhand University, Jhansi-284128, Uttar Pradesh, India

²Directorate of Forensic Science Services,
MHA, Govt. of India, New Delhi

Abstract: Establishing educational medium of writer of any suspected document could be a critical task which may prove to be assistive for the investigating authorities during their investigation, especially while dealing with anonymous letters. In the present study emphasis has been given to the estimation of probable educational medium of writers, on the basis of statistical validation of features observed in 400 handwriting samples by using chi square test. Analyzed samples were comprised of two most common mediums of education in India i.e. English Medium and Hindi Medium. The frequency of differentiating features observed in samples was computed and their relative probability was calculated. Present work proposes a computational tool by using C++, designed to compute the input in terms of values given by user and suggests the probability of writer belonging to English medium educational background. 50 unknown samples were used to validate the tool and it is found to be 87.85% accurate.

Keywords: *Handwriting, Automated Tool, Educational Medium*

1. INTRODUCTION

Handwriting is a complex acquired motor skill which involves together neurological, physiological and sensory impulses. Comprehension of form, visual perception, acuity, pathways of central nervous system, and physiology and anatomy of the muscles and bones of arm are the various factors which combine to generate the desired output^[1,2]. Handwriting is affected by various intrinsic and extrinsic factors like gender, handedness, age, educational level of the writers, etc. Classification of offline handwriting based on various attributes like gender, handedness, age and medium of education has a vital role in forensic document examination, it may help in minimizing the list of suspect and corollary to this, the accuracy of positive identification in lesser time could be achieved.

Document experts rely on visual observation of handwriting characteristics and sometimes their manual measurements along with some sort of computational approach which may measure the desired characteristics of handwriting. Number of studies have been conducted earlier, reporting many computational methods based on high level programming and personal models majorly based on optical character recognition^[3]. Different characteristics (micro- features) have been found to be powerful for handwriting discrimination^[4]. However, comprehensive research is required in estimating their relative significance in identifying the attributes like age, gender^[5], handedness^[6], etc. The present study aims to emphasise the estimation of probable educational medium of writers, based on statistical validation of characteristics observed in 400 handwriting samples using chi square test.

2. METHODOLOGY

Handwriting samples of 400 writers (230 belonging to English medium and 170 belonging to Hindi medium educational background) of age range between 18 - 60 years were randomly collected from various regions of Uttar Pradesh, without letting them know about the purpose of the present study. All the participants were requested to copy the control passage in provided handwriting performa in their normal writing. The obtained samples were preserved for further analysis. The digital image processing of the scanned writing samples were performed in following steps:

2.1 Image Acquisition: High resolution HP LASERJET M1005 scanner was used to acquire the writing samples in the digital format at 1200 DPI and a dataset of 400 offline handwritings was prepared by compiling and storing the acquired images in memory of the system used.

2.2 Noise Removal: Range filters were used to remove the noise appeared during image acquisition due to a number of factors, thereby diminishing the intensity of actual image.

2.3 Feature Selection: The image is further segmented for feature extraction. The frequency of so obtained characteristics was computed statistically using chi square test and compared with the tabulated value to prove the alternate hypothesis that there is significant variation in Hindi Medium and English Medium handwriting as mentioned in Table 1.


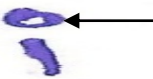


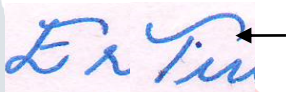


These characteristics were then framed to use in the Program designed to calculate the probable educational medium of writer mentioned in Table 2 and Figures 1(A) & 1(B).

Table 1: total characteristics studied for medium of education based classifier

S. No.	Characteristics Observed			English Medium	Hindi Medium	χ^2	Interpretation
1.	Loop formation in the lower part of “f”	P	Observed	182	65	73.29	Significant
			Expected	141	106		
		A	Observed	107	153		
			Expected	82	66		
2.	Correct use of punctuation marks	P	Observed	178	34	133.76	Significant
			Expected	121	91		
		A	Observed	50	138		
			Expected	107	81		
3.	Abbreviation	P	Observed	173	14	44.1028	Significant
			Expected	107	80		
		A	Observed	55	158		
			Expected	121	92		
4.	Placing over “i”	P	Observed	164	28	121.66	Significant
			Expected	109	83		
		A	Observed	64	144		
			Expected	119	89		
5.	Centrally aligned heading	P	Observed	38	160	90.65	Significant
			Expected	85	113		
		A	Observed	134	68		
			Expected	87	115		
6.	Loop formation in the stem of the letter “d”	P	Observed	160	41	20.63	Significant
			Expected	115	86		
		A	Observed	68	131		
			Expected	113	86		
7.	Cursivity in handwriting	P	Observed	28	155	105.58	Significant
			Expected	79	104		
		A	Observed	144	73		
			Expected	93	124		
8.	Embellishment	P	Observed	150	31	21.43	Significant
			Expected	103	78		
		A	Observed	78	141		
			Expected	125	94		
9.	Artistic ability	P	Observed	07	128	96.67	Significant
			Expected	58	77		
		A	Observed	165	100		
			Expected	114	146		
10.	Loop formation in the stem of letter “b”	P	Observed	119	14	18.65	Significant
			Expected	76	57		
		A	Observed	109	158		
			Expected	152	115		
11.	Spelling mistakes	P	Observed	141	27	47.27	Significant
			Expected	72	96		
		A	Observed	31	201		
			Expected	101	133		
12.	Round top formation in of letter “A”	P	Observed	114	23	30.34	Significant
			Expected	59	78		
		A	Observed	58	205		
			Expected	113	150		

At 5% Level of significance, Tabulated Value =3.84 ; P- Present , A-Absent

Table 2: characteristics selected for automated tool

S. No.	Characteristics Observed	Characteristics Demonstration	Probability Percentage of Writer belonging to English Medium Educational Background
1.	Loop formation in the lower part of letter "f"		19.13
2.	Placing over "i"		17.22
3.	Loop formation on the stem of letter "d"		16.74
4.	Cursive Handwriting		16.26
5.	Embellishments		15.78
6.	Loop formation on the stem of letter "b"		12.44
7.	Round top formation of letter "A"		2.39

```

Whether loop in lower part of f is present or not?Yes=1 No=0
1
p1=19.13
Whether circle over i is present or not?Yes=1 No=0
1
p2=17.22
Whether loop in stem of d is present or not?Yes=1 No=0
1
p3=16.74
Whether writing is cursive or not?Yes=1 No=0
1
p4=16.26
Whether embellishments are presents in writing or not?Yes=1 No=0
1
p5=15.78
Whether loop is present in stem of b or not?Yes=1 No=0
1
p6=12.44
Whether the top of A is rounded or not?Yes=1 No=0
1
p7=2.39
Sample Count is 1
Probability of writer being from English medium is 99.96
Press 1 to continue with another sample
    
```

Figure : 1 (A)

```

Whether loop in lower part of f is present or not?Yes=1 No=0
1
p1=19.13
Whether circle over i is present or not?Yes=1 No=0
0
p2=0
Whether loop in stem of d is present or not?Yes=1 No=0
1
p3=16.74
Whether writing is cursive or not?Yes=1 No=0
1
p4=16.26
Whether embellishments are presents in writing or not?Yes=1 No=0
0
p5=0
Whether loop is present in stem of b or not?Yes=1 No=0
0
p6=0
Whether the top of A is rounded or not?Yes=1 No=0
1
p7=2.39
Sample Count is 1
Probability of writer being from English medium is 54.52
Press 1 to continue with another sample

```

Figure : 1 (B)

Figure 1(a) & 1(b): screen shots of the developed tool to check probable medium of education of writer

3. RESULT

The proposed tool is based on the observation of the selected characteristics in writing samples, the user carefully examines the writing sample and enters the value as 0 - for the absence of the features; or 1- for the presence of the features in that sample. The tool then automatically computes the values and as output displays the overall probability of the writer belonging to English medium educational background. The tool was formulated on statistical observation of 400 samples and then validated by 50 unknown samples. The False Acceptance rate (FAR) was highest in cases of Rounded top formation of "A" 8% whereas the False Rejection Rate (FRR) was maximum in case Loop formation in the stem of "d" 8% as mentioned in Table 3. The tool has showed an accuracy of 87.85%.

Table 3: validation of tool

Characteristics	False Acceptance Rate	False Rejection Rate	Accuracy
<i>Loop formation in the lower part of letter "f"</i>	4	3	93%
<i>Placing over "i"</i>	7	4	89%
<i>Loop formation on the stem of letter "d"</i>	5	7	88%
<i>Cursive Handwriting</i>	6	6	88%
<i>Embellishments</i>	7	7	86%
<i>Loop formation on the stem of letter "b"</i>	7	8	85%
<i>Round top formation of letter "A"</i>	8	6	86%

4. CONCLUSION

The above observed characteristics proved to be highly significant in differentiating the writings belonging to English medium educational background. Along with this each characteristic is associated with its fixed probability value on the basis of their relative occurrence in the samples analyzed. Thus, a combination of different characteristics will give a different probability of the handwriting being written by a writer with English medium educational background. The tool proposed in present paper will assist the document examiners in estimating the educational background 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. (1999) *Handwriting Identification: Facts and Fundamentals*. CRC Press-Taylor & Francis, Boca Raton, Florida, USA, p.10-14.
- [2]. Hilton O. (1982) *Scientific Examination of Questioned Documents*. Elsevier, New York, USA, pp. 153-157.
- [3]. Nigam K., Saran V., Gupta A.K. and Sayeed A. (2014) Emerging Trends of Automated Handwriting Analysis. *International Research Journal of Management Science & Technology* 4(3):13-25.
- [4]. Srihari S.N., Cha S.H., Arora H. and Lee S. (2002) Individuality of Handwriting. *Journal of Forensic Sciences* 47: 856-872.
- [5]. Nigam K., Saran V., Srivastava A., Ahmed S. and Gupta A.K. (2018) Automated Tool Based on Statistical Study for Predicting Probability Dextrality of Writer. *Journal of Forensic Sciences & Criminal Investigation* 7(1):001-006.
- [6]. Nigam K., Saran V., Ahmed S. and Gupta A.K. (2013) Automated Tool Based on Statistical Study for Predicting Probability of Female Authorship. *International Journal of Engineering Research & Technology* 2(11):904-906.

