

# Exploring Gender-Specific Features from Handwriting for Devanagari Script

Omega Norton<sup>1</sup>, Beauty Arora<sup>2</sup>, Rajesh Kumar<sup>3</sup>

<sup>1,2,3</sup>Government Institute of Forensic Science, Chhatrapati Sambhajanagar, Maharashtra, INDIA

## ABSTRACT

Handwriting is a very important mode used for communication. It has been in use for a long time for various purposes. Handwriting is found in many criminal cases and its analysis is quite tedious work to do and even time-consuming. But it would be better if the gender of the writer is known; it would be helpful in narrowing the search of the suspects. In this work, we have tried to explore certain features which could be used for the identification of the gender of the writer. Various researches have been done to determine gender from handwriting in various languages. No such work has been done on any of the Indian languages yet. We have selected the Marathi language, the script used for writing is the Devanagari script, which is the script for some other languages used in India. The database was created and the features were extracted manually including line quality, pen pressure, formation of dots, and many more. However, few features are seen in males while some others are in females. The study is still not conclusive and looking for further work on it.

**KEYWORDS:** Handwriting, language, Devanagari script, gender, features

## INTRODUCTION

Handwriting is an acquired skill, generated by psycho-motor coordination. Handwriting has a great history (carvings on the inner wall of caves). Handwriting is a physical method of expressing thoughts, sharing information, and also for its storage. Handwriting is generally carried out by hand, but there are few people who use another body part as they are not able to use their hands. Handwriting is created by the combination of various signs and symbols and contains linguistics, cognitive, perceptive, and motor components that need to be coordinated in an Integral form. The symbols have it unique stroke formation, spatial relationship with other components, and the direction of the writing. Handwriting is unique for every individual.

The science of handwriting analysis is based on the premise that no two individuals can produce exactly the same handwriting and that an individual cannot exactly reproduce his own handwriting, otherwise known as natural variations [1], which can be used in handwriting examination. These variations could differentiate one writer from another; similarly, the handwriting of females can be differentiated from that of males. Determination of gender is helpful during the investigation. Though there are no certain reasons behind these variations, they could be useful in investigation.

The work done in the domain is not found to be satisfactory. The studies were conducted in two ways: online and offline method. In both methods, the features were somewhat similar to those of Forensic Document Examiners (used for the examination). The previous studies were carried out in languages like English, Roman, Bosnia, and Arabic. According to the literature, the features that were more prominent

in males were: hurried, uneven, messy, inconsistent, heavy, confident, and arrogant, and that of females were found to be neat, even, well organized, rounded, small, symmetrical, and lack of confidence. Even though all the features were helpful in determining the gender of the writers the accuracy in most of the cases ranges from 60% to 80%. The researchers have used various classifiers for the extraction of the features.

The aim of the study is to study the Marathi language (Indo-Aryan language), which is an Indian language (22 scheduled languages), and the official language and co-official language used in the States of Maharashtra and Goa respectively. The script used for writing is Devanagari, which is the base of languages, like Hindi, Kashmiri, Maithili, Nepali, Konkani, Bodo, Sindhi, and Sanskrit. A literature survey revealed that no work has been conducted for this script. In this study, we try to explore a set of features helpful in distinguishing the male handwriting from that of the female.

## LITERATURE REVIEW

Handwriting is a multi-component task. This implies that the production of writing strokes is the overt manifestation of divergent cognitive, psychomotor, and biophysical processes [2]. Gender identification is an important part of the biometric system. Attempts were made to differentiate the handwriting of males and females. The contribution made by various researchers in this domain has been summarized in the following:

Osborn stated that “*the writing of women is, as a rule, more delicate containing more superfluous peculiarities and mannerisms. It is generally more finished. Shading, if present in older writings, is likely to be in bunches, particularly at ends of words. Heavy shading of every stroke, and jabs of the pen, is more often masculine.*”[3]

In 1910, Downey conducted an experiment; he used 100 envelopes addressed by males and 100 by females. Thirteen examiners recorded their judgments regarding the gender of the writer in the form of four-point scale. The judgment was based on two methods: analogical approach and the reference to types or classes of handwriting. The analysis led to the belief that the female hand is colorless, conventional, neat, and usually small, while the male hand is bold or careless or experienced, above all individual. All researchers reported that an individual of one sex exhibits writing which seems more characteristics of the other sex which was termed as “inversion of sex sign”. These depend on the amount of writing done, age, or profession requirement. [4]

In 1926, Newhall conducted a similar experiment with 200 envelopes (half by both genders). 92 volunteers comprised of both males and females, were involved in the estimation of gender. The features observed in males are angles, irregularity, and slant while curves, regularity, and verticality were observed in the females. [5]

In 1929, Broom et.al conducted research on 40 samples randomly selected from the assortment of obtained samples. 18 samples of males and 22 samples of females were taken. Regularity, curves, conventional form, and uniformity of slant are characteristic of the samples judged most frequently as female handwriting. Irregularity, unconventional form, angles, and individual slants are the characteristics of the samples judged as male handwriting. There are sex differences in the handwriting and it is possible to determine the sex correctly roughly two times in every three judgments when the sex of the writer is unknown [6].

In 1931, Young carried out a similar experiment with samples collected from 25 males and 25 females. The accuracy rate was about 61% (mean performance). It helped in concluding that the untrained

judge could determine the gender with an average performance of 11% better than chance. Both males and females have an equal ability to identify the gender of the writer [7].

In 1934, Harry Tenwolde studied the difference in the quality of penmanship of 100 girls and 100 boys. Scoring of it was based on the Thorndike Handwriting scale. The finding was that girls write better than of boys but the difference is not entirely significant (reliable) [8].

In 2002, Vivien Burr conducted three experiments to provide evidence with respect to Hayes' finding [(a) gender of the writer was predicted at above chance levels even with very little stimulus material. (b) Men and women are equally accurate in their prediction] the gender of handwriting is discernible from minimal information and to identify cues used by adults in judging the gender of the writer. Experiment I: The cues for females are as follows: neatness, rounded and consistent style, and joined up. And that for males: scruffiness, spiky, style and not joined up. For experiment II the cues for females are: neat, even, well organized, well presented, rounded, small, ornate, symmetrical, and lack of confidence, and that for males are: hurried, uneven, messy, inconsistent, spiky, sloping, heavy, bold, confident, assured or arrogant. In the third experiment, the cues for females are: careful, neat, and regular/consistent. And that males are: hurried, irregular, and untidy/scruffy. The gender of the writer is discernible with 60- 70% accuracy under experimentally controlled conditions of minimal stimulus exposures [9].

In 2007, Marcus et.al conducted work on the IAM-On DB consisting of samples of Roman handwriting of 200 writers with 8 handwritten texts per writer (7 text lines on average). Out of this, he used 50 samples of each of the two genders for the training set and 25 each for the test set. They extracted 29 features to detect the gender of the writer. They were: Online features: (speed; writing direction; curvature; normalized x- and y-coordinate; speed in x- and y-direction; overall acceleration; acceleration in x- and y-direction; log curvature radius; vicinity aspect; vicinity curliness; vicinity linearity; vicinity slope) and Off-line features: (ascenders/ descenders; context map). Two approaches for the classification, namely, Support Vector Machines (SVMs) and Gaussian Mixture Models (GMMs) have been applied. The best accuracy was about 67.06% for GMM and 59.97% for SVM [10].

In 2011, Marcus Liwicki conducted an experiment on IAM-OnDB collected from the whiteboard from 200 writers. The classification method used was based on the Gaussian Mixture Model. The features extracted are: (a) Online Features such as, speed; writing direction; curvature; normalized x- and y-direction, overall acceleration; acceleration in x- and y- direction; log curvature radius which is the length of the circle with best approximates the curvature at the point p; vicinity aspect; vicinity curliness; vicinity linearity and vicinity slope. (b) Offline features such as the mean gray value of the pixels; center of gravity; second order vertical moment of the center of gravity; positions of the uppermost and lowermost black pixels; rate of change of these positions[with respect to the neighboring windows]; number of black-white transitions between the upper and lower pixels. Best reported accuracy is: 55.39 % (offline); 64.25 % (online); and 67.57 % (combination). Training has been performed with a modified version of EM based on the Maximum A posterior Principle (MAP) [11].

In 2012, Sokic et al. conducted a research on a database consisting 3766 entries, contributed by more than 300 students. The approaches used for classification are Generic Fourier Descriptor and Contour based descriptor. Feature analyzed include conventional features like arrangement, class of allograph, connections, design and construction of allograph, dimension [vertical and horizontal], slant, spacing, intra and inter word, abbreviations, baseline alignment, initial and terminal strokes, punctuations, embellishment, legibility, line continuity, line quality, pen control, writing movements, natural variations,

persistency, lateral expansion and word proportions and the others computational features like macro features which are extracted at the document level, like pen pressure, writing movement, stroke formation, slant and word proportion and the micro features which are extracted using a gradient, structural and concavity attributes. 20 parameters were used to semi-automatically detect gender with confidence slightly above the statistical significance. There was a qualitative difference in the spectral bands of the handwriting in the direction of movement, speed of direction change, curvature, and average bending energy [12].

In 2013, Youssef et.al conducted an experiment on 282 handwriting samples in Arabic and English languages. The classifier utilized was a Support Vector Machine (SVM). The features extracted from the image are direction, curvature, chain code, gradient direction, and Wavelet Domain Local Binary Patterns (WD-LBP). Classifiers that were trained separately produced an accuracy of 68.6% and 85.7% for Arabic and English language respectively. And the one trained with both showed an accuracy rate of 74% [13].

In 2014, Maadeed and Hassaine conducted research for Arabic and English languages (separately and combined) on the QUWI database, which consisted of handwriting samples from 200 writers. For the identification, they used Random Forest and Kernel discriminant analysis. The features proposed and described for this research include direction, curvature, tortuosities, chain code, and edge-based directional features. The features do not correspond to a single value, but are defined by a Probability Distribution Function (PDF) extracted from the handwriting images to characterize the writer's individuality. Classification rates were reported to reach about 74.05% for gender prediction. The complexity of Arabic script tends to help in better categorization of writers [14].

In 2015, Marzinotto et al. stated that the writing style is characterized by the appearance (inclination, curvature, etc) and the writing process dynamics (speed, acceleration, jerk). The two levels used for this experiment included: 1<sup>st</sup> level – allows generating writer-independent word clusters according to raw signal information. 2<sup>nd</sup> level – the handwritten words produced by the person is converted into a Bag of Prototype Words (BPW) by assigning each word to its closet cluster and subsequently generating the person's cluster frequency histogram. This information is then input into the second layer of clustering for generating of handwriting style based on: (a) raw spatiotemporal word similarities and (b) handwriting variability similarities between different writers. The two types of features include the dynamic features which gather local dynamic information, such as speed, acceleration, and jerk, and the spatial features which describe the static shape of local pen trajectories by measures such as local stroke angles and curvatures or inter-character spaces. To visualize clustering they used two dimensionality reduction techniques, Principal Component Analysis (PCA) and Stochastic Neighbor Embedding (SNE). They found that men tend to have a medium-speed writing style and tends to write by mixing “cursive” and “script” style. Whereas women tend to write either very fast or very slow and do not tend to mix “cursive” and “script” styles [15].

In 2016, Enric Sesa-Nogueras regarded sequencing of text as two separate subsequences as: (a) one of on surface trajectory or pen down stroke and (b) one consisting of in air trajectory or pen up stroke. Each word is given by 7 time sequence (features) which are as follows: The x co-ordinate, the y co-ordinate, time stamp value, the button status value, azimuth, the altitude and the pressure. Code book was generated using self-organizing map. The accuracy for pen down stroke obtained with a single word is 68% while for pen up stroke, it is 60.1%. The accuracy in combination is 68.5% [16].

In 2016, Boudjenek et al. conducted research for studying soft biometric traits such as writer's gender, handedness, ethnicity, and age. It has great significance in Forensics, for the identification of the

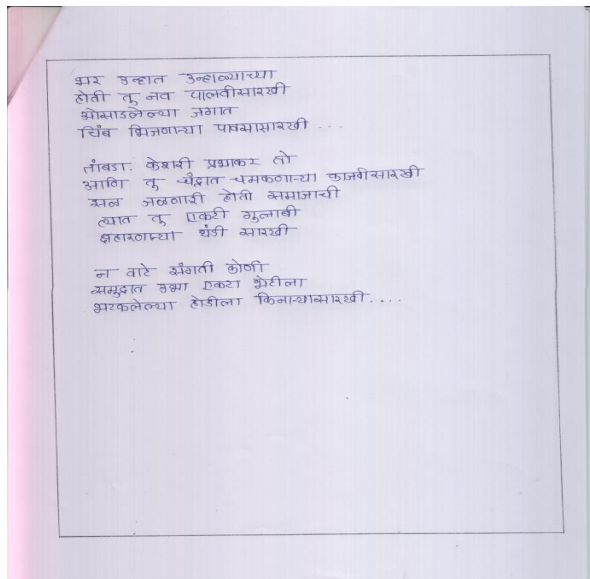
writer of the anonymous writing as well as historical documents. The classifier used for the same was the Support Vector Machine (SVM), which was trained on the basis of the pixel density, pixel distribution, and the Histogram of the Template (HOT). Two databases were used for this experiment; the IAM and KHATT, which contain handwritten sentences in English and Arabic language respectively. The output for pixel density, pixel distribution, and HOT resulted in an accuracy of 73.63%, 70.90%, and 71.82% respectively for the IAM database. And for the KHATT database, it is 71.11%, 73.33%, and 75.56% respectively [17].

## MATERIALS AND METHODS

A database of 200 handwritten documents (100 male and 100 females; Age group: 14 to 30) has been created from the students of the Marathwada region of the State of Maharashtra. The volunteers were asked to fill out a consent form before giving their handwriting samples. The volunteers write on an A-4 size white sheet with the help of a blue ball pen while sitting in a comfortable position on a chair. The volunteers were told to replicate the same text in Marathi (as specified earlier) three times. The sample handwritten text by a female and male writer has been depicted in Figure 1 and Figure 2 respectively.

For the present work, the conventional features (used by the Forensic Questioned Document Examiner) were utilized to find the authenticity or to identify the writer of the document. We have explored the efficacy of the following features for gender identification:

- Line quality (degree of regularity)
- Spacing (between letters, words, and lines)
- Size (Absolute size)
- Slant
- Embellishment (adds or extends strokes)
- Pen pressure (force applied by the fingers on writing surfaces)
- Curvature (curve used for format letter)
- Formation of dot (anuswar)
- Connecting Strokes (unions of two or more letters)
- Placement of 'matra'



**Figure 1: Handwriting exemplar of female**

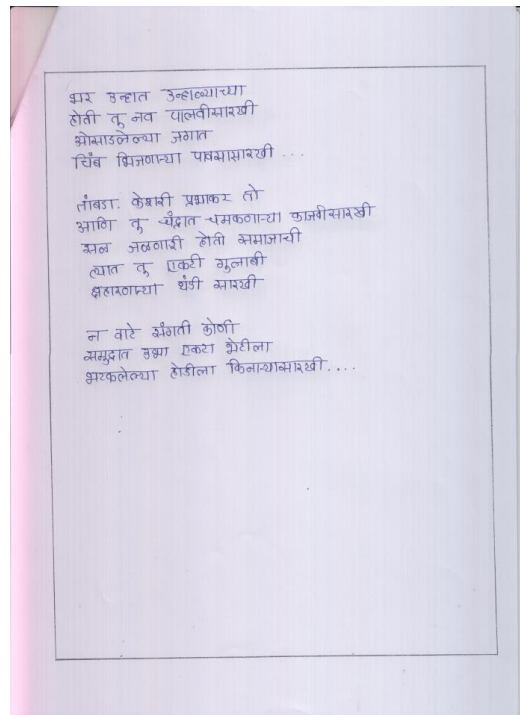


Figure 2: Handwriting exemplar of male

## RESULT

The results of the present work are shown in Figure 3 through Figure 15. Figure 3 depicts line quality. As shown in the figure, the line quality of females is average to good while the line quality of males is poor to average.

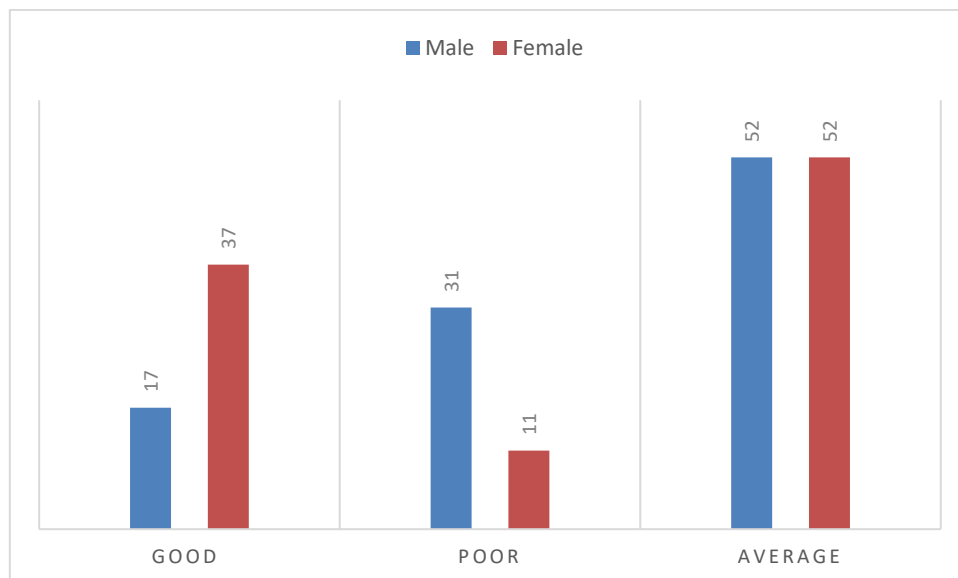
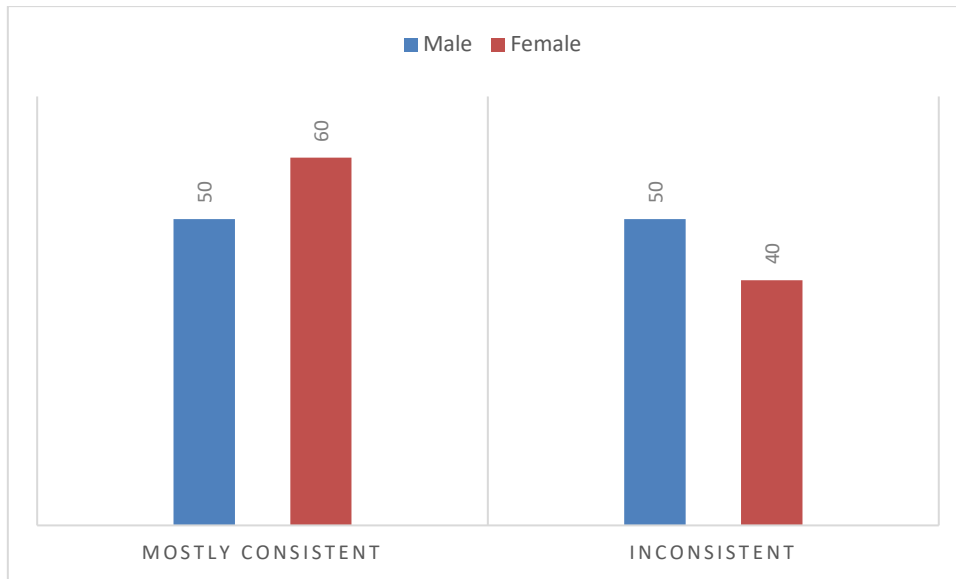


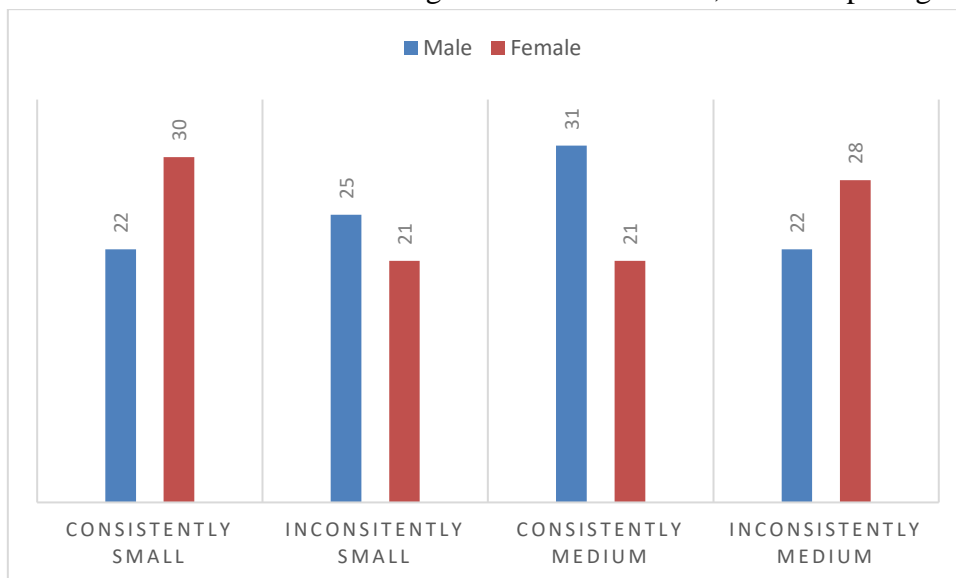
Figure 3: Line Quality of male and female writers measured in three scales: good, poor, and average

Figure 4 depicts the spacing between the letters in males and females. As shown in the figure, there is no significant difference between the handwriting of males and females, as far as spacing between the letters is concerned. Similar observations were also found in the case of spacing between words and sentences too.



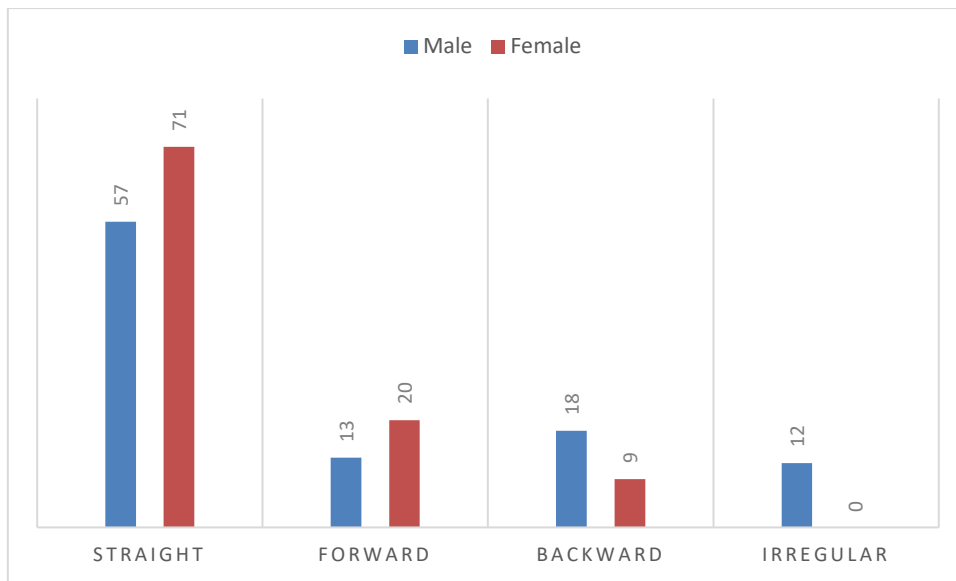
**Figure 4: Spacing between letters in male and female writers measured in terms of its consistency**

Figure 5 depicts the size of letters in handwriting for males and females. As shown in the figure, there is no significant difference between the handwriting of males and females, as far as spacing size is concerned.



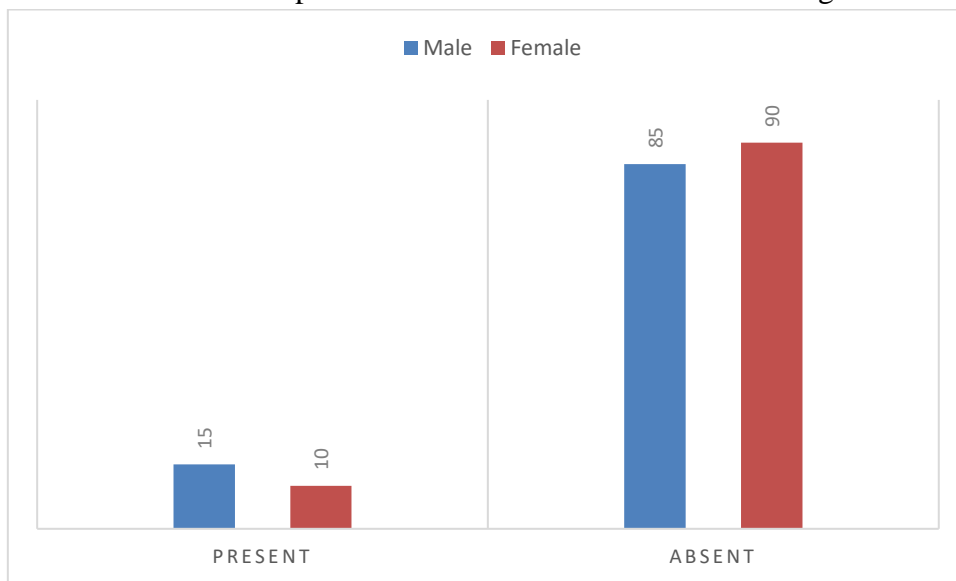
**Figure 5: size of letters in male and female writers measured in terms of consistency**

Figure 6 depicts slants of handwriting for males and females. As shown in the figure, males have a distributed slant, however, females are more inclined towards a straight slant.



**Figure 6: Slant of handwriting in male and female writers measured in four scales: straight, forward, backward, and irregular**

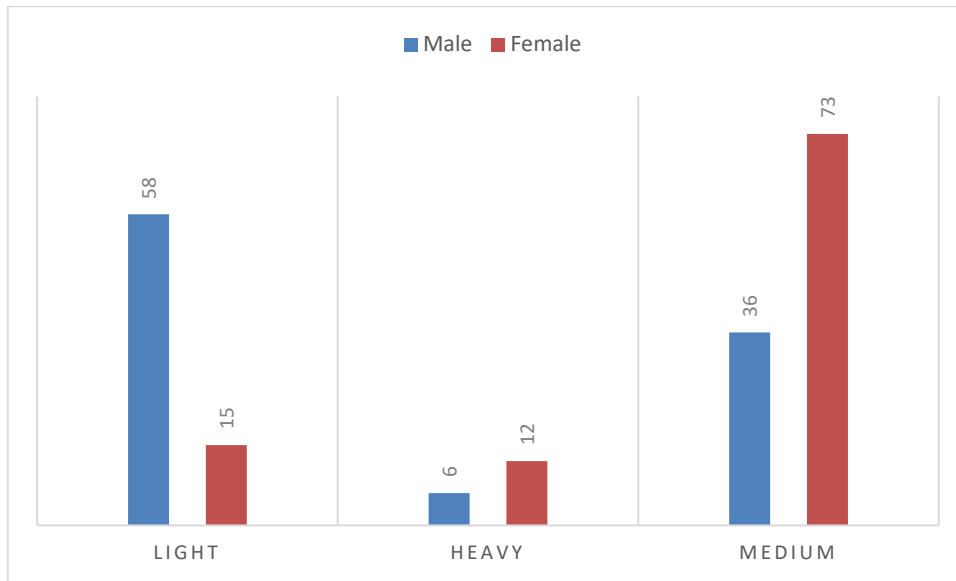
Figure 7 depicts embellishments in the handwriting of males and females. As shown in the figure, there is no significant difference as far as the presence of embellishment in handwriting is concerned.



**Figure 7: Embellishment of handwriting in male and female**

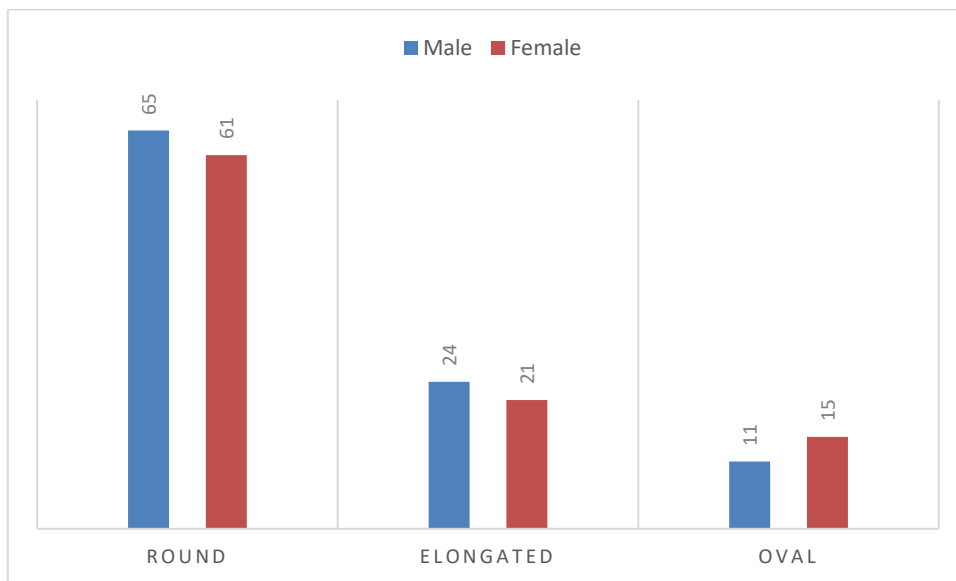
Figure 8 depicts pen pressure in the handwriting of males and females. As shown in the figure, males have light pen pressure on the other hand females are having medium pen pressure.





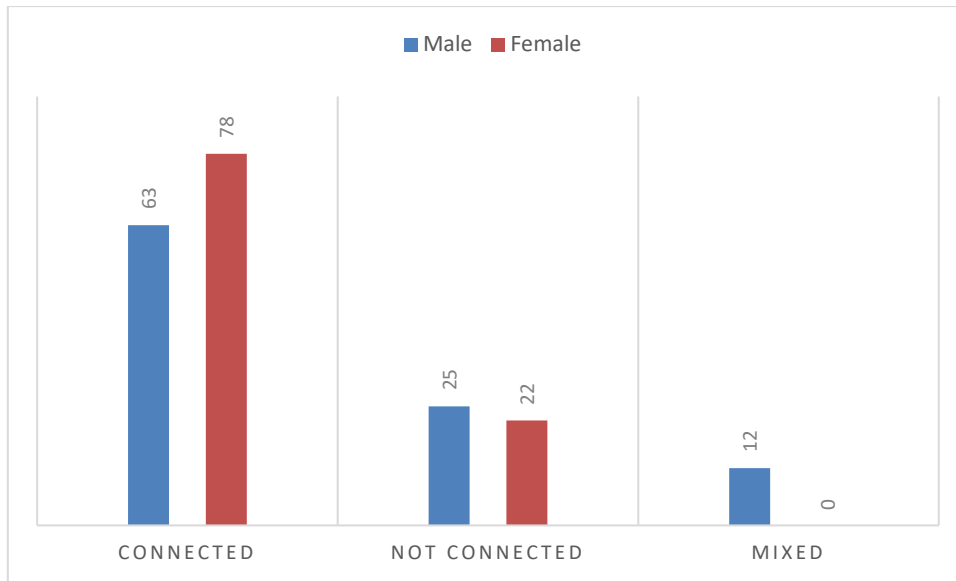
**Figure 8: Pen pressure in handwriting of male and female**

Figure 9 depicts types of curvature made by males and females in handwriting. As shown in the figure, there is no significant difference in curvature formation in the case of both males and females.



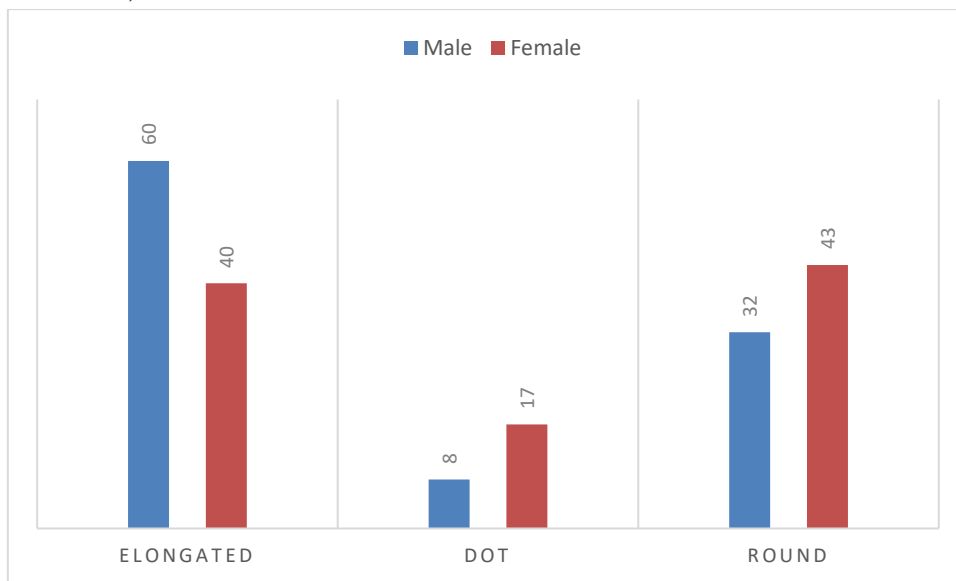
**Figure 9: Types of curvature made by male and female in handwriting**

Figure 10 depicts the connections between letters in handwriting by males and females. As shown in the figure, there is no significant difference, as far as connections between letters are concerned.



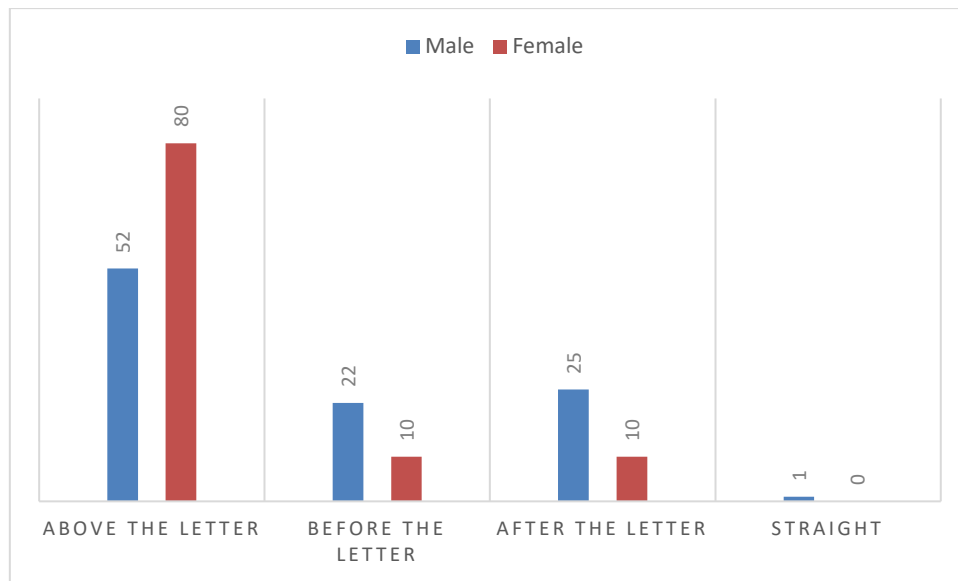
**Figure 10: Connection between letters by male and female in handwriting**

Figure 11 depicts the formation of dots (ansuwar) by males and females. As shown in the figure, there is no significant difference, as far as formations of dots are concerned.



**Figure 11: Formations of dots (ansuwar) by male and female in handwriting**

Figure 12 depicts the position of matra by males and females. As shown in the figure, there is no significant difference, as far as the position of matra is concerned.



**Figure 12: Position of matra by male and female in handwriting**

## CONCLUSION AND FUTURE WORK

In the present study, gender-specific features were aimed to explore the Devanagari script such that it can be later utilized for gender identification from handwriting. It can be concluded from the study that females possess the following characteristics: line quality to good, slant-straight and consistent, pen pressure-more, curvature-round to oval, formation of dot-round like zero, consistency in connecting strokes and placement of matra-just above the letter. While males possess the following characteristics: line quality-poor to average, alignment-highly inconsistent, slant-inconsistent, pen pressure-light, curvature-round to elongated, dot formation-elongated short line, spelling mistakes-more, connecting strokes-inconsistent, placement of matra-inconsistent. Based on the study, it is still hard to conclude if a particular feature belongs to a male or female. However, a few features are more tends toward males while others towards females. Exploration of other features, working on a larger database, and extension of the work in an automatic environment are a few future scopes of the present study.

## REFERENCE

1. [www.Robsonforensic.com/articles/forensic\\_handwriting\\_analysis\\_expert](http://www.Robsonforensic.com/articles/forensic_handwriting_analysis_expert)
2. Van Galen, G. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science*, 10, 165 – 191. Huber RA, Headrick AM. Handwriting identification: facts and fundamentals. Boca Raton (FL): CRC Press; 1999.
3. Osborn AS. Questioned Document. 1910.
4. Downey JE. Judgments on the sex of handwriting. *Psychol Rev.* 1910; 17(3):205–16.
5. Newhall SM. Sex differences in handwriting. *J Appl Psychol.* 1926; 10(2):151– 61.
6. Broom ME, Thompson B, Bouton MT. Sex differences in handwriting. *J Appl Psychol.* 1929; 13(2):159–66.
7. Young PT. Sex differences in handwriting. *J Appl Psychol.* 1931; 15(5):486– 98.
8. Tenwolde H. More on sex differences in handwriting. *J Appl Psychol.* 1934; 18(5):705–10.
9. Burr Vivien. Judging gender from samples of adult handwriting: Accuracy and use of cues. *The Journal of Social Psychology.* 2002; 142:691-700

10. Liwicki M, Schlapbach A, Loretan P, Bunke H. Automatic detection of gender and handedness from on-line handwriting. Proc. 13th Conf. Graphonomics Soc. 2007: 179–83.
11. Liwicki M, Schlapbach A, Bunke H. Automatic gender detection using online and off-line information. Pattern Anal Appl. 2011; 14(1):87–92.
12. Sokic E et al. Analysis of offline handwritten text samples of different gender using shape descriptor. XIth Conf. Telecommunications (BIHTEL). 2012.
13. Youssef A E et al. Automated gender identification for Arabic and English handwriting. 5th International Conf. Imaging for Crime Detection and Prevention (ICDP). 2013
14. Al Maadeed S, Hassaine A. Automatic prediction of age, gender, and nationality in offline handwriting. EURASIP J Image Video Process. 2014; 1:10.
15. Marzinooto G. Age and gender characterization through a two layer clustering of online handwriting. Advanced Concepts for Intelligent Vision Systems. 2015; 9386: 428–439.
16. Sesa-Nogueras E et al. Gender classification by means of online uppercase handwriting: A text-dependent allographic approach. 2016;8(1):15-29
17. Boudjenek N et al. Combination of topological and local shape features for writer's gender, handedness and age classification. 13th conf. ICIAR Image Analysis and Recognition. 2016.