Arab Journal of Forensic Sciences and Forensic Medicine 2026; Volume 8 Issue (1), 4-20

جامعة نايف العربية للعلوم الأمنية
NAIF ARAB UNIVERSITY FOR SECURITY SCIENCES
Est. 1978

Naif Arab University for Security Sciences
Arab Journal of Forensic Sciences and Forensic Medicine
المجلة العربية لعلوم الأدلة الجنائية والطب الشرعي
https://journals.nauss.edu.sa/index.php/AJFSFM

الجمعية العربية لعلوم الأدلة الجنائية والطب الشرعي
Arab Society for Forensic Sciences and Forensic Medicine

# Gender Estimation via Handwriting Dynamics: A Statistical and Machine Learning Approach

# تقدير الجنس عبر ديناميكيات خط اليد: نهج إحصائي وقائم على تعلم الآلة

CrossMark

Priya Sharma[1]*, Spriha Sharma[1], and Om Prakash Jasuja[2]
[1]Forensic Science Department, Chandigarh University, Punjab, India.
[2]Forensic Science Department, RIMT University, Punjab, India.

*Received 6 Jan. 2026; accepted 26 Apr. 2026; available online 7 Jun. 2026*


## Abstract

Gender estimation from handwriting has long been of interest in forensic document examination. However, most prior studies rely on static handwriting features, with limited exploration of dynamic writing behaviour. The advent of advanced technologies in recording handwriting made it possible to study the dynamic nature of the handwriting. The present study investigates gender-related differences in dynamic handwriting features derived from digitally captured natural style-handwriting, integrating statistical analysis with machine learning classification. Handwriting samples were collected from 200 participants (100 males, 100 females) using a pen-enabled digital tablet across three trials. Five primary dynamic features: Time-stamped X and Y coordinate data, pressure, azimuth, altitude and time, were recorded and used to derive 21 kinematic, spatial, pressure-based, and temporal features. The statistical analysis using Mann-Whitney U test, indicated significant gender-related differences in several dynamic features, including temporal parameters (pen-up duration, pen-down duration, total duration), pressure-related measures, spatial (handwriting width and height) and kinematic parameters like velocity and acceleration. Features demonstrating statistical significance were subsequently employed as inputs to supervised machine learning models, including Random Forest, Support


**Keywords:** forensic sciences, forensic document examination, dynamic handwriting, gender classification, statistical analysis, machine learning


## المستخلص

لطالما كان تقدير الجنس من خلال خط اليد محل اهتمام في مجال فحص المستندات الجنائية. ومع ذلك، اعتمدت معظم الدراسات السابقة على الخصائص «الاستاتيكية» (الثابتة) للخط، مع استكشاف محدود لسلوك الكتابة «الديناميكي» (الحركي). وقد أتاح ظهور التقنيات المتقدمة في تسجيل الكتابة اليدوية إمكانية دراسة الطبيعة الديناميكية للخط.

تبحث الدراسة الحالية في الفروق المتعلقة بالجنس في سمات خط اليد الديناميكية المستمدة من الكتابة اليدوية بالأسلوب الطبيعي والمصورة رقمياً، مع دمج التحليل الإحصائي مع تصنيف تعلم الآلة. وقد جُمعت عينات الخط من 200 مشارك (100 ذكر، 100 أنثى) باستخدام جهاز لوحي رقمي يدعم القلم عبر ثلاث تجارب. وقد سُجلت خمس سمات ديناميكية أولية هي: (بيانات إحداثيات X وY المحددة زمنياً، الضغط، السمت «Azimuth»، الارتفاع «Altitude»، والوقت)، واستُخدمت لاشتقاق 21 سمة كينماتيكية (حركية)، ومكانية، وزمنية، وأخرى قائمة على الضغط.

وأشار التحليل الإحصائي باستخدام اختبار «مان - ويتني يو» (Mann-Whitney U) إلى وجود فروق جوهرية ذات دلالة إحصائية بين الجنسين في عدة سمات ديناميكية، بما في ذلك: المعايير الزمنية (مدة رفع القلم، مدة وضع القلم، المدة الإجمالية)، ومقاييس الضغط، والسمات المكانية (عرض وارتفاع الخط)، والمعايير الكينماتيكية مثل: السرعة والتسارع.

وعقِب ذلك، استُخدمت السمات التي أظهرت دلالة إحصائية كمدخلات لنماذج تعلم الآلة الخاضعة للإشراف (Supervised Machine Learning)، بما في ذلك مصنفات: الغابة العشوائية (Random Forest)، وآلة ناقل


**الكلمات المفتاحية:** علوم الأدلة الجنائية، فحص المستندات الجنائية، خط اليد الديناميكي، تصنيف الجنس، التحليل الإحصائي، تعلم الآلة

Production and hosting by NAUSS

* Corresponding Author: Priya Sharma
Email: priyasharmafsc@gmail.com
doi: 10.26735/AEYH5217

Vector Machine, and Gradient Boosting classifiers. The Random Forest achieved the highest classification accuracy (86.7%) followed by Support Vector Machine (82.7%) and Gradient Boosting (80%). The findings demonstrate the potential of dynamic handwriting features in forensic document examination for gender estimation and preliminary profiling of writers

الدعم (SVM)، وتعزيز التدرج (Gradient Boosting). حقق نموذج «الغابة العشوائية» أعلى دقة تصنيف بنسبة (86.7%)، يليه «آلة ناقل الدعم» بنسبة (82.7%)، ثم «تعزيز التدرج» بنسبة (80%).

كما تثبت هذه النتائج إمكانية الاستفادة من سمات خط اليد الديناميكية في فحص المستندات الجنائية لتقدير الجنس والتحليل الأولي لشخصية وخصائص الكاتب (Profiling).

## 1. Introduction

Handwriting is a highly skilled and complex neuromuscular task that involves precise motor coordination of the arm, hands, and fingers. It is the most advanced motor skill in humans, executed through continuous fluid motions controlled by precise neural timings. With practice, these movements become automatic and unique to each individual [1-4]. Handwriting has numerous applications across various domains, including forensic document examination, biometric authentication, healthcare, and education. It is used in forensic examinations for writer identification. In biometrics, handwriting supports secure identity verification [5], while in healthcare, it assists in detecting neurodegenerative disorders [6-8]. Educational technologies also employ handwriting tools for assessment of students' writing skills and to identify brain development issues [9-10]. Handwriting is influenced by multiple intrinsic and extrinsic factors, such as age, gender, handedness, fatigue, and emotional state. Among these factors, gender hold greater gravity in influencing the handwriting characteristics. If handwriting reflects certain personality traits, it is believed that one of the most notable distinctions is between male and female writers [4].

The prospect of handwriting estimating a writer's gender from handwriting characteristics has long intrigued forensic document examiners and behavioural scientists. Early researchers have reported stylistic differences in male and female handwriting. Handwriting features like heavier shading, firmer pen strokes, and clumping at word endings are observed in male handwriting [11]. However, features like neatness, delicacy, fluency and more decorative has been associated with female handwriting [12-13]. Specific tendencies, such as backward slants, rounded counters, and consistent crossbar angles, are more commonly associated with feminine energy [14]; conversely, forward slants, angular strokes, and pressure variations are more typical of masculine handwriting. As per earlier study, females generally exhibit larger, more rounded, and upright handwriting compared to males [15]. Despite the observed differences, it is important to note that there may be an overlapping distribution in handwriting styles, spanning from complete masculinity to complete femininity. Furthermore, several forensically relevant studies using pen-and-paper writing-based static features with statistical analysis have identified gender-related difference in letter forms like loops, diacritics, initial and terminals strokes [16-20].

More recently, the use of dynamic handwriting, captured using digital pens or tablets, have opened new avenues for handwriting examination. Unlike static handwriting (pen-paper writing), dynamic handwriting records temporal and spatial attributes such as horizontal and vertical position, pressure, azimuth, altitude, and stroke sequence. The dynamic handwriting data provides precise insights into writing behaviour and motor control, offering objective and reproducible metrics [21-22]. Few forensic studies have explored dynamic features to examine forgery and motor disorders. For instance, prior studies explored features like jerk, pressure, pen-up and down durations across genuine and disguised signatures, reporting significant differences [23-24]. Although handwriting on a digital surface differs slightly from that on paper, the differences are not very significant [25]. Thus, dynamic features may offer useful insights in handwriting examination. There is plethora of exploratory studies on digitally captured handwriting as biometric tool for personal and gender classification. These studies have employed both static and dynamic features to estimate gender from handwriting. Liwicki et al. [26] proposed the first automatic gender classification method based on both static and dynamic handwriting features using Gaussian Mixture Models. The model based on dynamic features outperformed (~64%) the static features model (~55%) and the combined accuracy was ~68%. Another study focused on dynamic features in pen-up and pen-down strokes reported an accuracy rate of 72% for pen-down strokes and 64% for pen-up strokes with an overall accuracy of 74%. Additionally, the study highlighted the importance of longer text sample than single word inputs in gender classification [27]. Subsequent, empirical studies have utilised statistical feature selection with machine learning models and achieved a low accuracy of around 65% for gender classification on particular population which highlighted the need for further exploration of dynamic features in different population dataset [28]. Other studies identified significant gender differences in features such as stroke count, writing time, pressure and acceleration, although these studies were not validated through classification models[29-30]. A more recent study reported high accuracies of 94% (SVM) and 97% (ANN) using digitized (scanned) handwriting, where static features like margins, spacing and irregularity of strokes outperformed dynamic feature. The dynamic feature, that is, pressure alone achieved 73% accuracy. In this study, pressure was estimated based on shading produced by writing pen rather than quantitatively captured on digital tablet which implies further exploration of pressure and other dynamic features for gender estimation. Another empirical study utilizing dynamic handwriting features derived from controlled drawing patterns achieved 88% accuracy in adults and 90% in children; however, the study was limited by task specificity (zig-zag pattern drawings) and relatively small sample size which further warrants exploration of dynamic features [32].

As per earlier psychological and stylistic feature related studies, the gender of writer can be predicted better than a chance. These studies could achieve accuracy from 60-70% with stylistic features[12,14-16]. The forensic relevant studies have primarily explored the static handwriting features using statistical methods; however, many studies have not evaluated effect sizes or validated findings through classification models[18-19]. Where ever such validations have been performed, only moderate classification accuracy has been reported and the potential of dynamic features remains unexplored [20]. The biomedical factors or neuromuscular factors have been less explored as these did not give very promising results regarding gender estimation [27-30]. Recent studies based on the computational methods have explored gender estimation from handwriting as biometric tool with limited applicability in forensic investigations. The findings of the studies have focused on specific tasks, particular cultural and geographical background [31-32]. Forensic examinations require methodologies that not only classify gender accurately but also align with established scientific conduct, including transparent protocols, error rate estimation, and robustness to real-world conditions.

The present study is motivated by the need to bridge the gap between technical advancements in handwriting

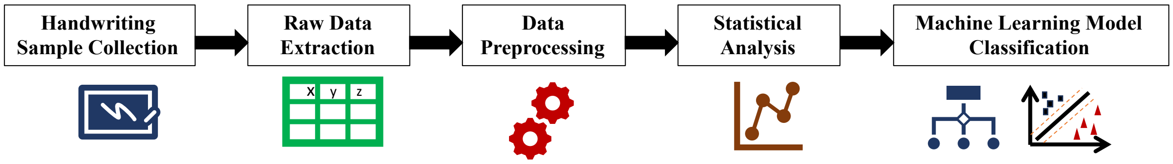



**Figure 1-** *Flowchart of the proposed methodology for gender classification.*

examination and their forensic applicability in gender classification. By exploring the dynamic features and their relationship with the gender of the writer, this work aims to strengthen the handwriting examination for investigative or preliminary profiling of suspects. By utilizing digital tablet with movement analysis software, the dataset in the form of natural handwriting of the writer has been collected. The statistical analysis is performed on the numerical data obtained from digitally captured handwriting followed by deriving specific features to differentiate male and female handwriting. The statistically significant features are subsequently employed as input for supervised machine learning classifiers to evaluate their discriminative strength. The findings of the study will not only expand the horizon for gender estimation from handwriting examination but also pave the way for the broader application of dynamic features in forensic investigation.

## 2. Materials and Methods

The present study proposes a systematic framework for gender classification using dynamic features of handwriting (Figure 1). The methodology involves handwriting data acquisition through digital tablet, preprocessing of raw handwriting signals, derivation of dynamic features, statistical analysis of dynamic features and application of machine learning models for classification.

### 2.1. Participants and Data Collection

Handwriting samples are collected from 200 volunteers (100 males and 100 females, aged 18-45 years, from the Punjab Region of India. All participants are graduates and digitally literate. Written informed consent is obtained prior to participation. Demographic characteristics, including age, education, handedness, and prior exposure to digital writing devices, are recorded. The anonymous CSV files containing raw handwriting feature data (excluding demographic details) are available from the corresponding author upon request.

### 2.2. Experimental Setup

Handwriting samples are collected using the Wacom One Display Model DTC133W0C, a digital tablet with sampling rate of 100 Hz. This tablet offered real-time visual feedback and a matte surface that mimicked paper-like friction. A stylus is used to capture precise handwriting data. For recording and analyzing the samples, MovalyZeR software (Version 6.1, NeuroScript LLC) is employed [33]. Participants were briefed on the objective and procedure of the study. During the experiment, participants sat comfortably on adjustable chairs, with the digital tablet placed on an optimum height table for easy access. They were allowed to do practice trials to familiarize themselves with the interface of the digital tablet. Each participant could adjust the placement of both the tablet and the chair to suit their ergonomic needs. A standardized paragraph (Figure 2) containing all alphabet letters was provided to the writers and instructed to copy in natural handwriting style. The handwriting samples were collected under uniform experimental conditions for each participant in three separate trials. Participants were asked to copy the paragraph placed adjacent to the digitizing tablet. Dictation was provided only as supporting

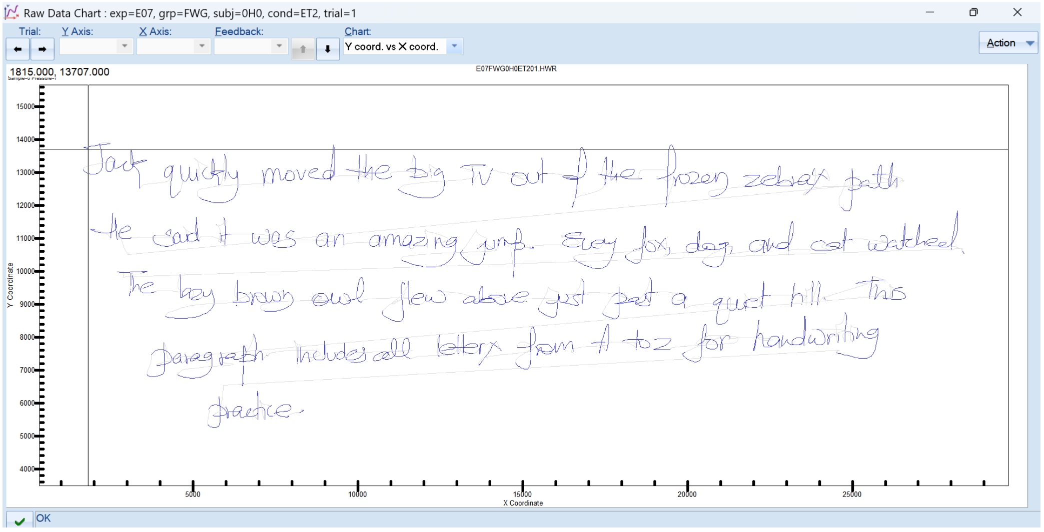



**Figure 2-** *Sample paragraph consisting of all the alphabet A to Z.*

**Table 1-** *Raw Dynamic Features extracted from digitally captured handwriting*

| S. No. | Name of Feature | Definition |
|---|---|---|
| 1 | X-coordinate | The horizontal movement component of the pen tip recorded within a stroke along the x-axis. |
| 2 | Y-coordinate | The vertical movement component of the pen tip recorded within a stroke along the y-axis. |
| 3 | Pressure (N) | The force applied by the pen tip on the tablet surface, recorded in normalized units ranging from 0 to 1024. |
| 4 | Azimuth (Horizontal angle) | The angular orientation of the pen in the horizontal plane, measured in degrees from 0° to 360°. |
| 5 | Altitude (Vertical angle) | The vertical inclination of the pen relative to the tablet surface, measured in degrees from 0° (parallel to the surface) to 90° (perpendicular to the surface). |

aid; however participants were free to rely solely on visual copying if preferred, particularly to maintain writing fluency and spelling accuracy. If any errors occurred during the trials, the samples were re-recorded to maintain the consistency of the data.

### 2.3. Feature Extraction and Preprocessing

Raw handwriting data consisted of time-stamped horizontal and vertical pen positions, pressure, azimuth, and altitude angles (Table 1). The raw data was not subjected to any additional segmentation or filtering. From these raw signals values, 21 dynamic features are derived (Table 2) , encompassing spatial, kinematic, pressure-based, and temporal characteristics. Pressure values are normalized following ISO/IEC 19794-7 standards [34]. Feature computation is performed consistently across trials; average values are used for to reduce intra-writer variations.

**Pressure values are normalized by using the Equation (1):**

$$\text{Normalized Pressure} = \left(\frac{\text{Raw Value}}{1024}\right) \times 100 \qquad (1)$$

**Table 2-** *Features derived from raw data*

| No. | Feature | Description | Formula |
|---|---|---|---|
| 1 | X-maximum | Maximum value in the X-coordinate | $XMax = max(X)$ |
| 2 | X-minimum | Minimum value in the X-coordinate | $XMin = min(X)$ |
| 3 | Y-maximum | Maximum value in the Y-coordinate | $YMax = max(Y)$ |
| 4 | Y-minimum | Minimum value in the Y-coordinate | $YMin = min(Y)$ |
| 5 | Maximum X-velocity | Maximum velocity along the X-axis | $VXMax = max(X/t)$ |
| 6 | Minimum X-velocity | Minimum velocity along the X-axis | $VXMin = min(X/t)$ |
| 7 | Maximum Y-velocity | Maximum velocity along the Y-axis | $VYMax = max(Y/t)$ |
| 8 | Minimum Y-velocity | Minimum velocity along the Y-axis | $VYMin = min(Y/t)$ |
| 9 | Maximum X-acceleration | Maximum acceleration along the X-axis | $AXMax = max\ (VX/t)$ |
| 10 | Minimum X-acceleration | Minimum acceleration along the X-axis | $AXMin = min\ (VX/t)$ |
| 11 | Maximum Y-acceleration | Maximum acceleration along the Y-axis | $AYMax = max(VY/t)$ |
| 12 | Minimum Y-acceleration | Minimum acceleration along the Y-axis | $AYMin = min(VY/t)$ |
| 13 | Average pressure | Average pressure applied | $Pavg = n\sum P$ |
| 14 | Peak pressure | Maximum pressure applied | $PMax = max(P)$ |
| 15 | Pen-up duration | Duration when the pen is not in contact with the surface (Pressure = 0) | $Tup = \sum(tup)$ |
| 16 | Pen-down duration | Duration when the pen in contact with the surface (Pressure = 1024) | $Tdown = \sum(tdown)$ |
| 17 | Total duration | Total time taken from start to finish | $Ttotal\ = tend - tstart$ |
| 18 | Handwriting width | Width of the handwriting (difference between max and min X-coordinates) | $HW = XMax - XMin$ |
| 19 | Handwriting height | Height of the handwriting (difference between max and min Y-coordinates) | $HH = YMax - YMin$ |
| 20 | Average horizontal Angle | Average horizontal angle of the strokes | $n\theta Xavg = n\sum\theta X$ |
| 21 | Average vertical Angle | The average vertical angle of the strokes | $n\theta Yavg = n\sum\theta Y$ |

Here, 1024 represents the maximum pressure value recorded by the device using software.

### 2.4. Statistical Analysis

To explore gender-based differences in dynamic handwriting parameters, at first the normality of the data was tested using the Shapiro-Wilk approach. The results indicated a non-normal distribution for both male and female groups. Consequently, the non-parametric Mann-Whitney U test was utilized to evaluate statistical differences in handwriting features between the two groups. This test is effective for comparing the distribution of two independent samples and is based on ranking observations [35-36]. Mann-Whitney U test was applied to investigate the significant differences in dynamic handwriting features of both genders. The analysis yielded the rank sums for both male and female groups, with

W-statistics and p-values calculated for each feature. A significant threshold of $p < 0.05$ was adopted, with results further categorized by significance levels: (***$p<0.001$, **$p<0.01$, and *$p<0.05$). In addition to statistical significance, effect sizes were quantified using Cliff's delta (Δ) to assess the magnitude of gender-related differences. Effect size interpretation followed established thresholds: Negligible ($|\Delta| < 0.147$), small ($0.147 \leq |\Delta| < 0.33$), medium ($0.33 \leq |\Delta| < 0.474$) and large ($|\Delta| \geq 0.474$) [37].

### 2.5. Machine Learning Classification

Features showing significant differences between male and female groups in the statistical analysis were selected as inputs for the machine learning models. Feature selection was solely performed based on statistical significance derived from the dataset to enhance interpretability rather than to optimize predictive performance.

Machine learning (ML), a subset of artificial intelligence, enables computers to learn from data and make predictions. In this study, a supervised learning approach was adopted, where algorithms learn from labelled data. Specifically, classification models were used to distinguish between two classes, male and female. These models predict discrete outcomes by recognizing patterns in the training data and applying them to unseen data [38]. Three most commonly utilized classification models, Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting(GB) are implemented. SVM identifies an optimal hyperplane that separates the two groups, while RF constructs an ensemble of decision trees and combines their output through majority voting [39-40]. GB is an ensemble technique that sequentially combines weak learners by minimizing classification error through gradient based optimization [41]. The machine learning models were build based on significant features only. The feature selection was done on complete dataset prior to model training to retain significant features and further enhance interpretability. While this approach may introduce a degree of data leakage, it was adopted to explore the significant handwriting features rather than optimizing classification performance. Machine learning was therefore applied as supportive validation tool. To assess model performance, the dataset is split into training and testing sets in a 75:25 ratio using stratified sampling to preserve class balance. Additionally, stratified five-fold cross-validation was applied on testing set to ensure a reliable and

**Table 3-** *Hyperparameter settings of the classification models*

| Model | Parameter | Setting |
|---|---|---|
| Random Forest | Number of trees | 400 |
| | Random state | 42 |
| Support Vector Machine | Kernel | Radial basis function (RBF) |
| | Regularization parameter (C) | 1.0 |
| | Gamma | Scale |
| | Probability estimates | Enabled |
| Gradient Boosting | Number of estimators | 100 |
| | Learning rate | 0.1 |
| | Maximum tree depth | 3 |
| | Random state | 42 |

unbiased performance evaluation. The hyper parameter settings are selected to balance accuracy and computational efficiency (Table 3). The number of trees in RF model was empirically evaluated, and performance was stabilized beyond 400 trees. Therefore, the number of trees was fixed at 400 to balance classification performance. Similarly, for GB model the number of estimators is fixed at 100 as further increase did not yield noticeable improvements in classification accuracy. The RBF kernel in SVM is selected due to its effectiveness in modelling non-linear decisions boundaries in handwriting data. Hyperparameters did not cause any data leakage as these were fixed prior to data evaluation based on empirical observations. Therefore, the test data remain completely unseen during both model training and hyperparameter selection.

Model performance was assessed using various performance metrics to ensure a comprehensive assessment. Accuracy, defined as the proportion of correct predictions across all classes, was one of the key metrics used. Additionally, the confusion matrix was employed as a tabulated summary to show the distribution of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

Sensitivity (Recall), which measures the number of True positive cases divided by the number of all true positive cases, and False negatives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

Specificity was the number of True negative cases divided by the total number of True negative cases and False positives.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

Precision was calculated as the number of true positive cases divided by the number of cases predicted to be positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

The F1 score, which balances precision and recall, was also included to provide a single metric that reflects both the model's ability to correctly identify positive cases and minimize false positives.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precison} + \text{Recall}} \tag{6}$$

Lastly, the Area Under the Curve (AUC) was determined, reflecting the model's ability to distinguish between the classes. These combined metrics comprehensively evaluated each model's performance [42]. The proposed classification models have been implemented using Python in Jupyter Notebook environment with libraries such as Pandas, NumPy, Scikit-learn, Seaborn, and Matplotlib [43].

## 3. Results and Discussion

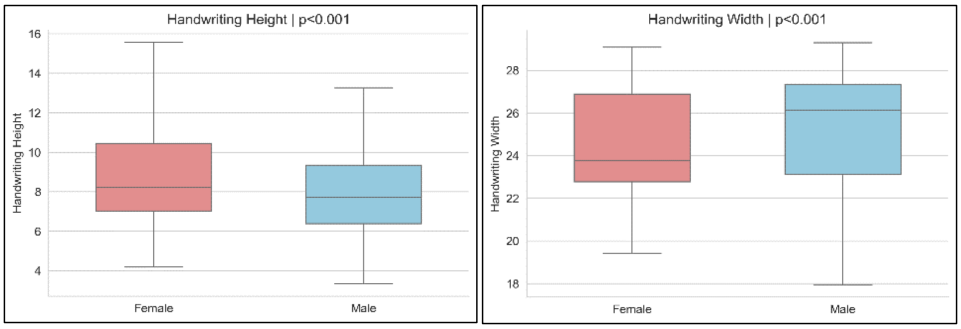
### 3.1. Mann-Whitney U Test Results

Several features demonstrated statistically significant differences between male and female writers (Table 4). Notably, the handwriting width and handwriting height demonstrated highly significant differences ($p<0.001$) between genders. Males exhibited higher values in handwriting width indicating broader horizontal strokes in their handwriting. The broader horizontal strokes observed in males may be attributed to differences in grip strength and motor control. In contrast, handwriting height was significantly higher in female writers, with $p<0.001$, respectively (Figure 3 a and b). This suggests that female handwriting tends to involve taller strokes, possibly depicting a more refined motor control and stylistic preference for vertically extended letterforms. The traditional studies have long described male handwriting as wider and more forceful whereas female handwriting tend to be more vertically extended and refined [13-15]. The X-maximum and

**Table 4-** *Mann-Whitney U Test result*

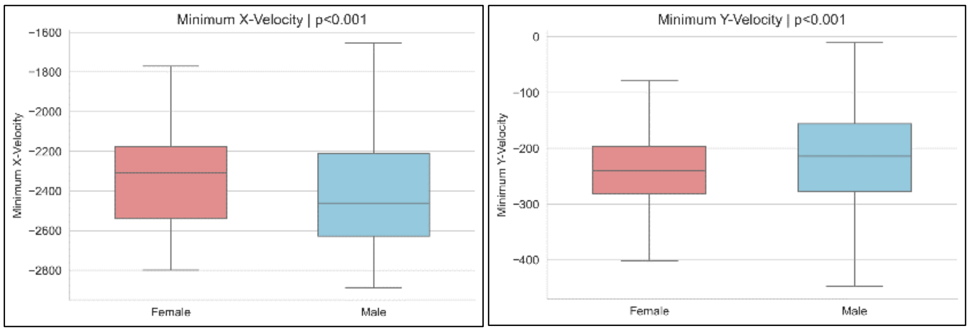
| Feature | Rank Sum Female | Rank Sum Male | W-Statistic | p-value | Cliff's Delta | Effect size |
|---|---|---|---|---|---|---|
| X-Maximum | 81611.500 | 76591.500 | 81611.500 | 0.1923 | 0.0636 | N |
| X-Minimum | 96601.000 | 83699.000 | 96601.000 | 0.0024* | 0.1434 | N |
| Y-Maximum | 86510.000 | 93790.000 | 86510.000 | 0.0865 | -0.0809 | N |
| Y-Minimum | 77638.500 | 100267.500 | 77638.500 | 0.0000*** | -0.2414 | S |
| Maximum X-Velocity | 85843.500 | 74617.500 | 85843.500 | 0.0213* | 0.1119 | N |
| Minimum X-Velocity | 96765.000 | 79950.000 | 96765.000 | 0.0002** | 0.1771 | S |
| Maximum Y-Velocity | 81057.500 | 85695.500 | 81057.500 | 0.2187 | -0.0592 | N |
| Minimum Y-Velocity | 78485.000 | 94681.000 | 78485.000 | 0.0005** | -0.1670 | S |
| Maximum X-Acceleration | 81136.500 | 95578.500 | 81136.500 | 0.0002** | -0.1772 | S |
| Minimum X-Acceleration | 96783.000 | 79932.000 | 96783.000 | 0.0002** | 0.1776 | S |
| Maximum Y-Acceleration | 94233.000 | 80703.000 | 94233.000 | 0.0014** | 0.1516 | S |
| Minimum Y-Acceleration | 80998.500 | 93937.500 | 80998.500 | 0.0014** | -0.1516 | S |
| Average Pressure | 82085.000 | 97615.000 | 82085.000 | 0.0002** | -0.1765 | S |
| Peak Pressure | 77625.500 | 102674.500 | 77625.500 | 0.0000*** | -0.2783 | S |
| Pen-up Duration | 62474.000 | 107762.000 | 62474.000 | 0.0000*** | -0.5364 | L |
| Pen-down Duration | 100739.500 | 75381.500 | 100739.500 | 0.0000*** | 0.2783 | S |
| Total Duration | 79962.000 | 93204.000 | 79962.000 | 0.0003** | -0.1737 | S |
| Handwriting Width | 82145.500 | 97554.500 | 82145.500 | 0.0002** | -0.1751 | S |
| Handwriting Height | 97376.000 | 80530.000 | 97376.000 | 0.0002** | 0.1763 | S |
| Average Horizontal Angle | 82692.000 | 90474.000 | 82692.000 | 0.1443 | -0.0696 | N |
| Average Vertical Angle | 83700.000 | 85371.000 | 83700.000 | 0.9576 | -0.0026 | N |

Note: Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Effect size (Cliff's delta): N = Negligible ($|\Delta| < 0.147$), S = small ($0.147 \leq |\Delta| < 0.33$), M = medium ($0.33 \leq |\Delta| < 0.474$), L = large ($|\Delta| \geq 0.474$).



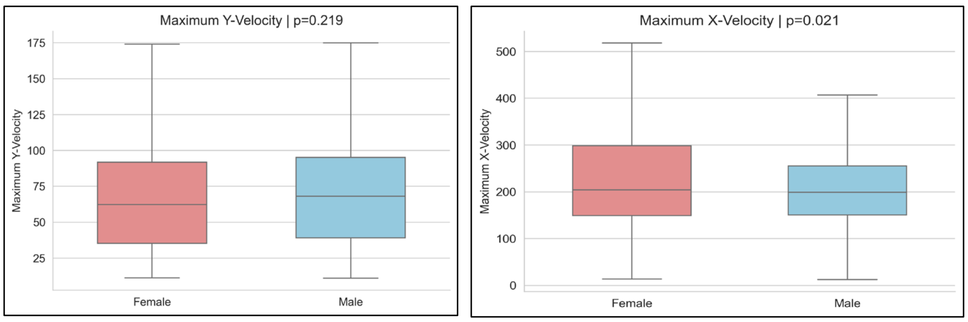

(a) (b)

**Figure 3-** *Boxplots showing significant differences in Handwriting Height (a) and Handwriting Width (b) in Males and Females*




(a) (b)

**Figure 4-** *Boxplots showing significant differences in Minimum X-Velocity (a), Minimum Y-Velocity (b) in Males and Females*

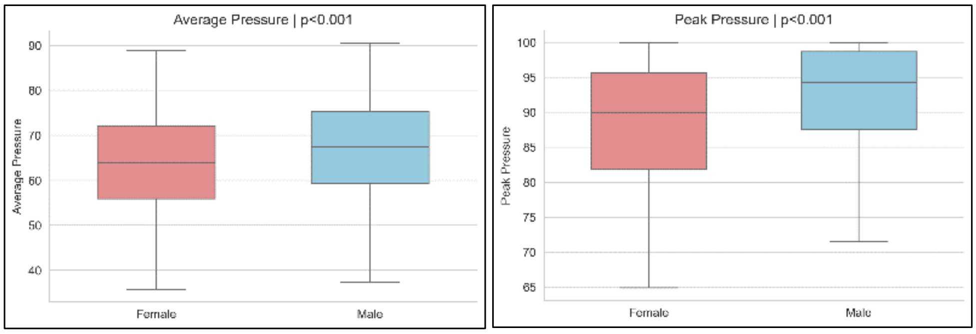



(a) (b)

**Figure 5-** *Boxplots showing significant differences in Maximum Y-Velocity (a), Maximum X-Velocity (b) in Males and Females*

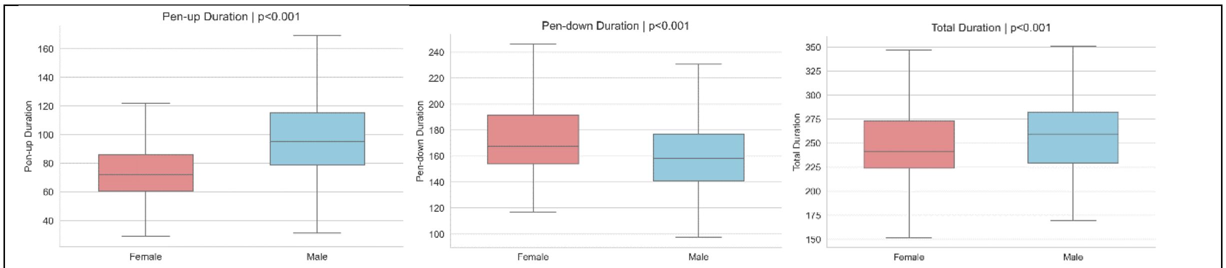
Y-maximum parameters did not differ significantly. However, the spatial coordinates, X-minimum and Y-minimum showed statistically significant differences with $p=0.002$ and $p<0.001$, respectively. These parameters reflect the gender differences in sub-baseline stroke formation behaviours. These findings highlight the discriminative value of spatial features in dynamic handwriting and underscore the need to consider both coordinates, horizontal and vertical, for gender differentiation. A prior study also identified horizontal and vertical coordinates as important discriminators in online handwriting [44]. However, in present study, effect size analysis indicated that these spatial features exhibited small or negligible effect sizes which suggest that these features are statistically significant but have limited discriminative strength.

Significant differences were observed in minimum X-velocity and minimum Y-velocity with $p<0.001$(Figure 4 a and b), indicating clear divergence in the deceleration phase of handwriting movement. Maximum X-velocity also showed a moderate but significant difference ($p=0.021$), maximum Y-velocity, however, did not show a significant difference with $p>0.05$ (Figure 5 a and b). This suggests a similarity in peak vertical movement execution. These findings demonstrate that gender differences are more pronounced in slow or deceleration phases of handwriting movements rather than in fast or maximum velocity movements. All acceleration based features, including maximum and minimum X-acceleration, and maximum and minimum Y-acceleration showed statistically significant differences between male and female




(a) (b)

**Figure 6-** *Boxplots showing significant differences in Average (a) and peak pressure (b) for males and females*




(a) (b) (c)

**Figure 7-** *Boxplots showing significant differences in Pen-up (a), Pen-down (b) and total duration (c) for male and female gender*

writers ($p<0.001$). Female showed more controlled velocity and acceleration profile than male writers. Prior online handwriting studies have reported that female generally exhibited smoother and more stable handwriting movements, whereas males tend to show greater variability during initial and terminal strokes [9]. The present study observed significant differences, however, the effect size showed small differences in both genders.

Pressure-related feature, average and peak pressure showed significant ($p<0.001$) gender-based differences. Male writers exhibited significantly higher average and peak pressure as compared female writers indicating greater applied pen force during handwriting execution (Figure 6 a and b). These findings support the earlier studies [6,30,45], which reported higher pressure in male handwriting. While pressure has been considered as controversial feature due to inconsistent findings across some offline studies [46], dynamic handwriting allows more precise measurement of force modulations. The extremely significant p-value for peak pressure further highlight the gender difference in force modulation. Effect size analysis indicated that pressure related parameters have small yet consistent gender-related differences.

Temporal parameters demonstrated the strongest gender differences. Pen-up duration, pen-down duration and total duration were all highly significant with $p<0.001$ (Figure 7 a and b). Male writers exhibited significantly longer pen-up duration and total duration indicating increased pause time and longer overall execution time. In contrast, female writers indicated significantly longer pen-down duration suggesting more continuous, smoother and consistent stroke production. This aligns with the previous studies, who identified duration and pressure as key traits in dynamic handwriting analysis [28]. Similarly, other studies [24,47] highlighted duration as a vital indicator in distinguishing genuine signatures from forgeries and identifying simulated writing. These findings reinforce the relevance of time-based features not only in gender estimation but also in forensic examination of forged or simulated handwriting. Neither average horizontal angle nor average vertical angle showed statistically significant differences ($p>0.05$). This suggests that overall writing orientation and stroke inclination are stable across genders and not discriminative parameters but still possess investigative value. The effect size analysis using Cliff's delta provided insight into the practical relevance of these differences. Among all features, pen-up duration demonstrated a large effect size indicating substantial gender differences. Other temporal parameters, pen-down and overall duration exhibited small effect sizes suggesting modest differences between male and female writers. These findings indicate that temporal features, particularly pen-up duration is most discriminating feature.

### 3.2. Machine Learning Models

The performance of all classifiers on the test dataset was assessed through performance metrics. All models demonstrated high sensitivity (Recall) for negative class (female) than positive class (male). The recall values were high for RF model (0.947), SVM (0.947) and GB (0.867) indicating strong capability of models to correctly identify female instances. In contrast, the precision values for the positive class (male) were consistently high across all models, with RF (0.937) and SVM (0.930) showing strong performance. This suggests that these models predict male instances with high degree of reliability. The F1-score, which balances precision and recall, further highlights the superiority of RF model (Male=0.855; Female=0.877) over SVM and GB models. Overall, the Random Forest model exhibited the most balanced performance across both classes, although difference relative to SVM and GB were subtle. Among the evaluated classifiers, the Random Forest model achieved the highest overall performance metrics followed closely by SVM and Gradient Boosting. The models showed better performance for females as compared to male (Table 5).

The confusion matrices further provided a detailed insight into the classification performance of each model. For the RF model, the confusion matrix shows 71 true negatives (female correctly classified) and 59 true positives (male correctly classified), with only 4 false positives and 16 false negatives (Figure 8a). In SVM model, 71 female samples were correctly classified, while 53 male samples were correctly identified (Figure 8b). However, the number of false negatives (22) is higher than that observed in RF, resulting in lower recall value. Despite this, the relatively low number of false positives (4) contributes to the high precision achieved by the SVM model. The GB model, correctly classified 65 female and 55 male samples, but exhibited higher number of misclassification, particularly false negatives (20) and false positives (10), which accounts for its slightly lower recall and precision values compared to other models (Figure 8c). Across all models, misclassifications are predominantly associated with false negatives, indicating that some male handwriting samples share overlapping dynamic

**Table 5-** *Performance Metrics for classification models.*

| Criteria | RF | SVM | GB |
|---|---|---|---|
| Test Accuracy | 0.867 | 0.827 | 0.800 |
| Precision (Female = 0) | 0.816 | 0.763 | 0.765 |
| Recall (Female = 0) | 0.947 | 0.947 | 0.867 |
| F1-Score (Female = 0) | 0.877 | 0.845 | 0.813 |
| Precision (Male = 1) | 0.937 | 0.930 | 0.846 |
| Recall (Male = 1) | 0.787 | 0.707 | 0.733 |
| F1-Score (Male = 1) | 0.855 | 0.803 | 0.786 |
| Balanced Accuracy | 0.867 | 0.827 | 0.800 |
| ROC AUC | 0.910 | 0.905 | 0.899 |
| CV Accuracy (Mean) | 0.820 | 0.811 | 0.791 |
| CV Accuracy (SD) | 0.041 | 0.040 | 0.051 |

*Gender labels are encoded as Female=0 and Male=1, Precision, Recall and F1 score are computed with respect to the positive and negative class.

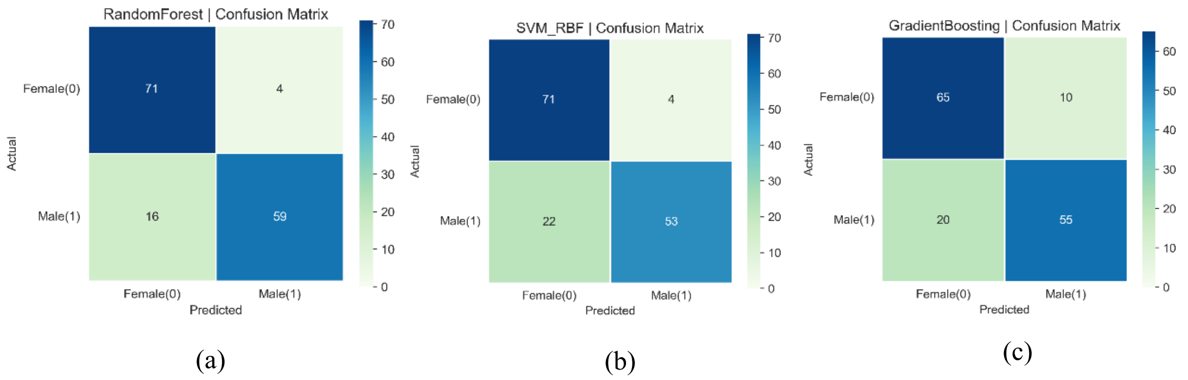
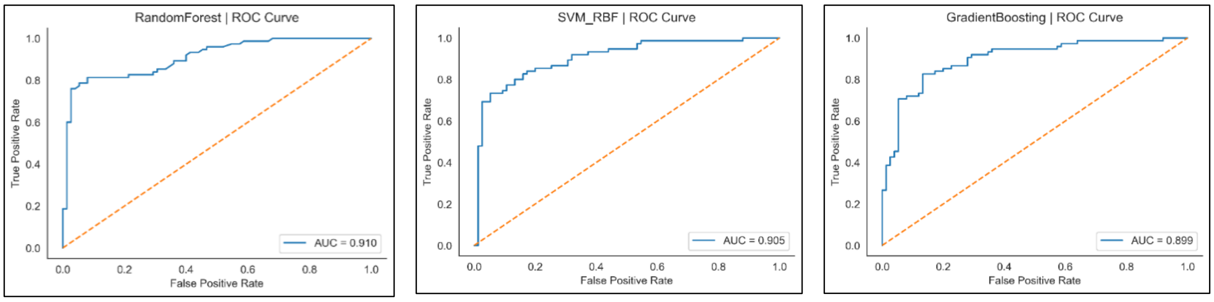



**Figure 8-** *Confusion Matrices for RF (a), SVM (b) and GB(c) classification models*

characteristics with female samples. The similarity in misclassification patterns across classifiers suggests intrinsic overlapping of features rather than model specific limitations. Handwriting is influenced by neuromuscular coordination, individual writing characteristics and natural variations, which may contribute to overlapping feature distribution between gender groups. Although, handwriting samples were collected in three trials to ensure consistency and reduce chance variations, natural variations persisted across samples.

The ROC-AUC values (Figure 9 a, b and c) further supports these findings. All three models exhibited strong discriminatory performance as indicated by high ROC-AUC values (RF=0.910, SVM-0.905, GB= 0.8999). The ROC curve models consistently lies closer to the top left corner of the plot, suggesting that dynamic features contain substantial class separating capability, enabling effective discrimination between the two gender groups beyond random classification. The slightly higher ROC-AUC achieved by RF model further underscores its robustness and superior overall




(a) (b) (c)

**Figure 9-** *ROC-AUC of for RF (a), SVM (b) and GB(c) classification models*

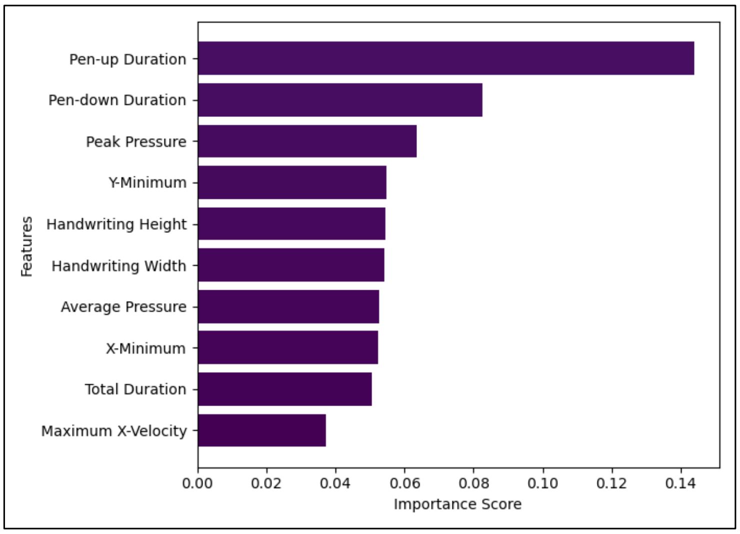



**Figure 10-** *Feature Importance score in RF Model*

performance. However, Paired-T test analysis conducted on five-fold cross-validated ROC-AUC scores revealed no statistically significant difference between top performing models (p=0.295), indicating comparable classification performance. Further, the feature importance analysis was performed using RF model and the top 10 features that contributed predominantly to the classification performance (Figure 10). Pen-up and pen-down duration followed by peak pressure contributed majorly in classification performance. The findings of the present study are in agreement with and in some cases outperform prior work on gender classification from handwriting using digital devices [26,28,29,32]. The findings demonstrated that while certain features demonstrated statistically significant features, it is important to note that findings should not be interpreted as standalone indicators of gender. The classification accuracy of 86.7% indicates promising discrimination ability. However, there is presence of overlap in feature distribution which limits the reliability of model in actual casework.

## 4. Conclusions

This study investigated the potential of dynamic handwriting features in predicting the gender of writers using digitally captured handwriting. Statistical analysis using the Mann-Whitney U test revealed significant gender-based differences across several dynamic features, including spatial characteristics (handwriting width and height), kinematic parameters velocity and acceleration extremes), pressure related measures (average and peak pressure and temporal features (pen-up, pen-down and total duration). Female writers exhibited greater overall writing duration, relatively more controlled minimum velocity and acceleration profiles, indicating slower execution, frequent pen pauses and smoother motor control. They also demonstrated, greater handwriting height indicating more vertically elongated stroke formation. In contrast, male writers showed higher peak and average pen pressure, broader handwriting width, more extreme velocity and acceleration measures in both horizontal and vertical directions, suggesting more forceful, faster, and dynamically variable stroke execution. These observations support the hypothesis that dynamic handwriting behaviour differs systematically between genders. The significant features identified were further evaluated using machine learning classifiers, including Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB) classification models. Among these, the RF model achieved the highest classification accuracy (86.7%) demonstrating the potential effectiveness of the ensemble method (RF) in handwriting-based gender classification. The time- related features and pressure contributed predominantly to classify accuracy. Despite these promising results, some degree of misclassification was observed across all models, primarily due to the inherent overlap of dynamic handwriting features between male and female writers. The study has limited sample size and consists of participants from relatively homogeneous population which causes limitations on generalizability. The participants shared similar educational exposure which may influence writing habits and reduce inter-writer variability. Additionally, handwriting can vary across different cultural and linguistic backgrounds. Hence, the derived results may not directly be applicable to population of different cultural and linguistic backgrounds. The study primarily includes young adult population which limits the wider applicability in diverse age groups. Therefore. validation on larger and more diverse datasets is necessary to improve generalizability and classification performance. Overall, the findings of the study provide a useful foundation for understanding the role of dynamic features in gender estimation, although it should be interpreted with caution. The outcomes may be viewed as exploratory and supportive in nature rather than conclusive. These results may contribute to preliminary profiling and future studies can be conducted for further validation.

## Conflict of interest

The authors declare no conflicts of interest.


## Source of funding

The authors received no financial support for the research, authorship or publication of this paper.


## References


1. Hilton, O. (1993). Scientific examination of questioned documents (Rev. ed.). CRC Press.
2. Thomassen, A. J. W. M., & Teulings, H. L. (1983). The development of handwriting. In The psychology of written language (pp. 179–213).
3. Bisesi, M. S. (2006). Scientific examination of questioned documents. In J. S. Kelly & B. S. Lindblom (Eds.), Scientific examination of questioned documents (2nd ed.). CRC Press.
4. Harralson, H. H., & Miller, L. S. (2017). Huber and Headrick's handwriting identification: Facts and fundamentals. CRC Press.
5. Faundez-Zanuy, M., Fierrez, J., Ferrer, M. A., Díaz,

M., Tolosana, R., & Plamondon, R. (2020). Handwriting biometrics: Applications and future trends in e-security and e-health. Cognitive Computation, 12, 940–953.

6. Mergl, R., Tigges, P., Schröter, A., Möller, H. J., & Hegerl, U. (1999). Digitized analysis of handwriting and drawing movements in healthy subjects. Journal of Neuroscience Methods, 90(2), 157–169.
7. Harralson, H. H., Teulings, H. L., & Farley, B. G. (2007). Comparison of handwriting kinematics in movement disorders and forgery. In Proceedings of the 13th Conference of the International Graphonomics Society (pp. 143–148).
8. Faundez-Zanuy, M., Mekyska, J., & Impedovo, D. (2021). Online handwriting, signature and touch dynamics: Tasks and potential applications in the field of security and health. Cognitive Computation, 13, 940–954.
9. Accardo, A. P., Genna, M., & Borean, M. (2013). Development, maturation, and learning influence on handwriting kinematics. Human Movement Science, 32(1), 136–146.
10. Hopcan, S., & Tokel, S. T. (2021). Exploring the effectiveness of a mobile writing application for students with dysgraphia. Education and Information Technologies, 26, 5561–5584.
11. Osborn, A. S. (1929). Questioned documents (2nd ed.). Boyd Printing (p.140).
12. Hodgins, J. H. (1971). Determination of sex from handwriting. Canadian Society of Forensic Science Journal, 4(4), 124–132.
13. Hamid, S., & Loewenthal, K. M. (1996). Inferring gender from handwriting in Urdu and English. Journal of Social Psychology, 136(6), 778–782.
14. Young, P. T. (1931). Sex differences in handwriting. Journal of Applied Psychology, 15(5), 486–498.
15. Hecker, M. R. (1996). The scientific examination of sex differences in handwriting. Paper presented at the American Society of Questioned Document Examiners Meeting.
16. Lester, D., Werling, N., & Heinle, N. H. (1982). Graphoanalytic differences by sex and handedness. Perceptual and Motor Skills, 55(3), 1190.
17. Totty, R. N., Hardcastle, R. A., & Dempsey, J. (1983). The dependence of slope of handwriting upon the sex and handedness of the writer. Journal of the Forensic Science Society, 23(3), 237–240.
18. Kumar, S., Saran, V., Vaid, B. A., & Gupta, A. K. (2013). Handwriting and gender: A statistical study. Z Zagadnień Nauk Sądowych, 95, 620–626.
19. Upadhyay, S., & Singh, J. (2017). Determination of sex through handwriting characteristics. International Journal of Current Research and Review, 9(13), 11–18.
20. Sharma, V., Bains, M., Verma, R., Verma, N., & Kumar, R. (2023). Novel use of logistic regression and likelihood ratios for estimation of gender from handwriting features. Australian Journal of Forensic Sciences, 55(1), 89–106.
21. Harralson, H. H. (2013). Developments in handwriting and signature identification in the digital age. Anderson Publishing.
22. Angel, M., & Kelly, J. S. (Eds.). (2020). Forensic document examination in the 21st century. CRC Press.
23. Mohammed, L. A., Found, B., Caligiuri, M., & Rogers, D. (2011). The dynamic character of disguise behavior for text-based, mixed, and stylized signatures. Journal of Forensic Sciences, 56, S136–S141.
24. Ahmad, S. M. S., Ling, L. Y., Anwar, R. M., Faudzi, M. A., & Shakil, A. (2013). Analysis of the effects and relationship of perceived handwritten signature size, graphical complexity, and legibility with dynamic parameters for forged and genuine samples. Journal of Forensic Sciences, 58(3), 724–731.
25. Sharma, P., Singh, M., & Jasuja, O. P. (2021). Forensic examination of electronic signatures: A comparative study. Nowa Kodyfikacja Prawa Karnego, 59, 149–149.
26. Liwicki, M., Schlapbach, A., & Bunke, H. (2011). Automatic gender detection using online and offline in-

formation. Pattern Analysis and Applications, 14(1), 87–92.

27. Sesa-Nogueras, E., Faundez-Zanuy, M., & Roure-Alcobé, J. (2016). Gender classification by means of online uppercase handwriting: A text-dependent allographic approach. Cognitive Computation, 8(1), 15–29.
28. Likforman-Sulem, L., Cordasco, G., & Esposito, A. (2022). Is online handwriting gender-sensitive? In Pattern Recognition and Artificial Intelligence (LNCS Vol. 13363, pp. 300–314). Springer.
29. Cordasco, G., Buonanno, M., Faundez-Zanuy, M., Riviello, M. T., Likforman-Sulem, L., & Esposito, A. (2020). Gender identification through handwriting: An online approach. In Proceedings of the 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 1–6). IEEE.
30. Faundez-Zanuy, M., & Mekyska, J. (2023). Analysis of gender differences in online handwriting signals for enhancing e-health and e-security applications. Cognitive Computation, 15, 208–219.
31. Al-Qawasmeh, N., Khayyat, M., & Suen, C. Y. (2023). Novel features to detect gender from handwritten documents. Pattern Recognition Letters, 171, 201–208.
32. Shin, J., Uchida, Y., Maniruzzaman, M., Hirooka, K., Megumi, A., & Yasumura, A. (2024). Online handwriting-based gender recognition using statistical and machine learning approaches. IEEE Access, 12, 1–10.
33. NeuroScript. (n.d.). NeuroScript website. https://neuroscript.net/
34. International Organization for Standardization. (2014). ISO/IEC 19794-7: Information Technology-Biometric data interchange Formats-Part 7: Signature/sign time series data. ISO/IEC.
35. Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18, 50–60.
36. Alkahtani, A. A., & Platt, A. W. (2011). The influence of gender on the ability to simulate handwritten signatures: A study of Arabic writers. Journal of Forensic Sciences, 56(4), 950–953.
37. Meissel, K., & Yao, E. S. (2024). Using Cliff's delta as a non-parametric effect size measure: an accessible web app and R tutorial. Practical Assessment, Research, and Evaluation, 29(1).
38. Zhou, Z.H. (2021). Machine learning. Springer Nature.
39. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273–297.
40. Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.
41. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.
42. Berrar, D. (2019). Performance measures for binary classification. In Encyclopaedia of bioinformatics and computational biology (Vol. 1, pp. 546–560). Elsevier.
43. Scikit-learn developers. (2007). Scikit-learn user guide. https://scikit-learn.org
44. Lei, H., & Govindaraju, V. A. (2005). A comparative study on the consistency of features in on-line signature verification. Pattern Recognition Letters, 26(15), 2483–2489.
45. Pascal, G. R. (1943). Handwriting pressure: Its measurement and significance. Journal of Personality, 11(3), 235–254.
46. Linden, J., Marquis, R., Bozza, S., & Taroni, F. (2018). Dynamic signatures: A review of dynamic feature variation and forensic methodology. Forensic Science International, 291, 216–229.
47. Dewhurst, T. N., Ballantyne, K. N., & Found, B. (2016). Empirical investigation of biometric, non-visible, intra-signature features in known and simulated signatures. Australian Journal of Forensic Sciences, 48(6), 659–675.