Article

A Structured Framework for Formalized and Quantitative Handwriting Examination

Yury Chernov * and Claudia Caspers

Zurich Institute for Handwriting Sciences (IHS), 8046 Zurich, Switzerland

* Corresponding author. E-mail: yc@ihs-sgg.ch (Y.C.)

ABSTRACT: The demand for a formalized and transparent approach to handwriting assessment has long been recognized within forensic and legal contexts. A structured methodology not only reduces interpretative subjectivity but also enables quantifiable measurement and ensures greater consistency in evaluations. This article presents a practical framework that models the degree of similarity between handwriting samples—texts and signatures—through a two-stage process: feature-based evaluation and congruence analysis. Both stages produce quantitative markers that are integrated into a unified similarity score, forming the foundation for more complex comparisons involving multiple questions and known texts. The proposed procedure, which is the major result of the paper, is not merely theoretical; it has been applied in real forensic casework, yielding preliminary statistical outcomes. In particular, it demonstrates the discriminative power of different handwriting features. The paper also discusses future directions for development, with a focus on the integration of artificial intelligence (AI) to enhance specific components of the assessment process.

1. Introduction

Handwriting examination is a forensic and scientific discipline focused on analyzing handwritten documents to determine authorship, detect forgery, and glean insights into an individual's cognitive and motor functions. It plays a crucial role in legal investigations, where expert assessments can significantly influence judicial outcomes. Over time, the field has evolved from relying primarily on subjective expert judgment to adopting more structured, scientifically grounded approaches, including the integration of digital tools. Despite these advancements, handwriting examination continues to face key challenges—such as limited standardization, lingering subjectivity, and ongoing skepticism regarding its reliability in forensic and legal contexts.

To address these concerns, the expert community has been pursuing several avenues of improvement:

1. Development of general guidelines [1,2]:
   They are intended to "provide a framework of procedures, quality principles, training processes and approaches to the forensic examination of handwriting" [1] (p. 3). These guidelines aim to unify practices and foster collaboration across the forensic science community. However, they are often tailored to large forensic laboratories and remain less accessible or applicable to private practitioners and smaller agencies.

2. Emphasis on professional training and peer review:
   A recurring theme is the importance of comprehensive training not only in forensic methodology but specifically in handwriting examination. Studies have shown that properly trained professionals tend to produce more accurate and reliable results [3,4], while non-professionals or laypeople typically demonstrate the famous Dunning–Kruger effect [5]. However, hands-on experience remains equally, if not more, critical. Another recommended practice is "independent, blinded peer review of the examination", which serves as a key mechanism for reducing error and enhancing objectivity [6].

3. Integration of graphometric analysis:

Support of graphometric analysis, *i.e.*, the quantitative evaluation of handwriting features [7]. In this context, it is essential to apply qualitative evaluation not only to the handwriting features that can be physically measured, like size, width, spaces or slant, but to all involved features. The quantitative evaluation of all handwriting characteristics is a central component of the examination procedure proposed in the present study.

4. Application of Statistical Modeling:

Various statistical methods have been explored to model the results of handwriting examinations [8–10]. These approaches provide quantitative assessments that are often more digestible for legal stakeholders. Nevertheless, the practical implementation of such methods is limited by the scarcity of comprehensive data. In many cases, experts resort to pseudo-probabilistic assessments, which can be misleading if not grounded in rigorous statistical reasoning. As a result, standardization bodies discourage the use of numerical probabilities unless they are backed by validated statistical models, favoring the use of calibrated verbal scales instead.

5. Development of Computer-Aided Tools:

Software solutions have been developed to assist with digitizing and analyzing handwriting samples [11,12]. However, these tools face significant limitations [13]. They typically automate only a small subset of handwriting features—insufficient for comprehensive analysis, e.g., [14]—and often perform unreliably, especially on complex or varied handwriting. Consequently, these tools are currently better suited for research rather than operational forensic use.

6. Exploration of Artificial Intelligence (AI):

AI models have shown promise in comparing handwriting samples and identifying anomalies [15–17]. While experimental results appear encouraging, no robust, practical applications have yet emerged. This is not surprising, as studies show that over 80% of AI projects fail [18]. Challenges include the lack of tailored AI models and insufficient, high-quality training data. Developers frequently rely on generic AI architectures trained on datasets that do not reflect real forensic scenarios. Moreover, most AI tools are designed for pairwise comparison, which does not align with typical forensic tasks involving multiple known samples of varying quality. As such, AI currently holds more potential for auxiliary tasks—such as report generation or advanced text searches based on natural language queries—than for core handwriting examination.

The procedure described in the following section integrates key elements from these emerging directions. Rather than a purely theoretical construct, it is a practical framework that has been applied in real forensic cases. Given the current limitations of computer-aided tools, the method relies on manual assessment. It is focused on comparing handwriting samples, not evaluating the authenticity of documents or detecting technical forgeries, which are beyond the scope of this study.

2. Materials and Methods

The aim of the proposed procedure is to maximize the objectivity and reliability of handwriting assessment by minimizing subjective influence and quantifying the evaluation process. This quantification facilitates both more substantiated probabilistic conclusions and the possibility of statistical research into the significance of individual handwriting characteristics.

The procedure incorporates the five main principles defined in [1], which are considered individually in every examination:

- "No two people write exactly alike.
- No one person writes exactly the same way twice, and no two naturally written signatures are exactly the same.
- The significance of any feature, as evidence of identity or non-identity, and the problem of comparison becomes one of considering its rarity, complexity, the relative speed and naturalness with which it is written, and its agreement or disagreement with comparable features.
- No one is able to imitate all of the features of another person's handwriting and simultaneously write at the same relative speed and skill as the writer that he/she is seeking to imitate.
- In those cases where the writer disguises their normal handwriting or imitates the handwriting of another".

In handwriting examination, two types of handwriting are typically distinguished: textual handwriting and signatures. Textual handwriting constitutes the main body of written content and is generally more stable and consistent. In contrast, signatures are identifiers often in stylized or abbreviated form (paraph), which exhibit greater variability. While the proposed procedure is suitable for analyzing both types, each presents its specific challenges and considerations.

The procedural steps outlined herein are familiar to forensic experts; however, our contribution lies in their comprehensive integration and systematic formalization at every stage. The general framework assumes a typical

forensic scenario in which the expert is provided with one or more questioned documents and several known documents—*i.e.*, samples of authenticated handwriting attributed to the presumed author(s). In some cases, known documents may be grouped according to different potential authors. This does not alter the core procedure, which is simply applied to each group independently.

The proposed algorithm of handwriting examination follows a two-stage structure: feature evaluation to compare general characteristics of handwriting and congruence analysis, a detailed letter-by-letter comparison to assess graphical and kinematic consistency. Both stages yield quantitative results that are combined into a final similarity score. The full procedure consists of the following steps:

1. Pre-assessment—preliminary review of all materials to ensure suitability for examination.
2. Feature evaluation of known documents—a systematic analysis of handwriting features in each known sample.
3. Determination of variation ranges—establishing the range of variation for each feature across known samples.
4. Feature evaluation of the questioned document—assessing the same set of features in the questioned handwriting.
5. Similarity grading for features—comparing the questioned features to known variation ranges and assigning similarity grades.
6. Evaluation of so-called handwriting elements (defined below).
7. Calculation of feature-based similarity score—aggregating individual handwriting element comparisons into a cumulative score.
8. Congruence analysis of letterforms—a detailed examination of each letter and its allographic (variant) forms in both questioned and known samples, including specific letter-pair combinations where necessary.
9. Evaluation of congruence score—quantitative assessment of consistency between corresponding letters and letter pairs.
10. Calculation of total similarity score as a function of the feature-based score and congruence score.
11. Expert conclusion—formulating the final expert opinion based on the total similarity score and contextual case information. Each of these steps is discussed in detail in the following sections.

2.1. Pre-Assessment

Pre-assessment is conducted prior to any detailed, side-by-side comparison of handwriting samples. Its purpose is threefold. The first objective is to assess whether the provided handwritten materials are suitable for comparison. For a meaningful analysis, the documents must be written in the same general style. For instance, it is not appropriate to compare cursive writing with print-style writing, as the differences in form and structure make such comparisons unreliable.

The second objective is to define the limitations of the analysis, which depends on the quality, quantity, and type of documents available. Increasingly, forensic handwriting experts must work with copies instead of original documents. This presents inherent limitations—many signs of forgery or document manipulation can only be reliably detected on original materials. However, due to various reasons, clients often simply do not have access to the originals. Additionally, the following aspects must be evaluated:

- Legibility of the documents;
- Verification that known samples are genuinely representative of the purported author;
- Assessment of whether known samples are contemporaneous with the questioned writing;
- Determination of whether sufficient material is available to assess natural handwriting variation, including the presence of multiple letterforms and their usage in various word positions.

These factors significantly influence the scope, depth, and reliability of the subsequent examination.

Finally, understanding the broader context of the case is essential. This includes the circumstances under which the questioned and known documents were created, the parties involved, and any relevant background that may inform the interpretation of handwriting characteristics—such as potential motivations for forgery, attempts at disguise, or psychological conditions that may affect writing style.

2.2. Evaluation of Handwriting Features

All known documents, one by one, are evaluated to assess every handwriting feature. The set of these characteristics covers all aspects of handwriting, including spatial structure, shape and dynamics. The selection of features is informed by numerous authoritative sources. To the standard sources traditionally belong, for instance, refs. [19–21] in the English-speaking countries or [22–24] in the German-speaking regions.

The main purpose of this stage is the quantitative assessment of each handwriting feature. The framework supports two primary approaches. The primary and most widely used method is based on ordinal scales. In our practice, we predominantly utilize the eight-point scale proposed in [23,24], which we have found to be the most practical and reliable after testing several alternatives. An example of such a scale for the letter size is given in Table 1. It should be mentioned that the letter size is evaluated using inner middle-zone letters—that is, letters located in the center of words, excluding the first and last letters, to reduce positional bias. Although some experts may disagree with specific scale definitions, clarity and consistency in classification are paramount in forensic handwriting comparison. In this context, unambiguous specification is more critical than universal agreement.

Table 1. Assessment of the letter size.

Value	Meaning	Remarks
(0)	Evaluation not applicable/meaningful	Generally, letter size should always be assessable, however, sometimes the scale of a copy is not known and only proportions, but not the size, can be assessed.
(1)	Very small letter size	At least 50% of letters are very small (<1 mm) and the rest are small
(2)	Small letter size	80% of letters have small size
(3)	Rather small letter size (tendency to the small size)	At least 50% of letters are small and the rest are medium
(4)	Indifferent or medium letter size	At least 80% of letters have medium size (2.0–3.5 mm) or different sizes are present, and it is not possible to identify the small or large sizes
(5)	Rather large letter size (tendency to the large size)	At least 50% of letters are large and the rest are medium
(6)	Large letter size	80% of letters have large size
(7)	Very large letter size	At least 50% of letters are very large (>5.5 mm) and the rest are large

The second method involves categorizing handwriting into five size groups—very small, small, medium, large, and very large—and then measuring the proportion of letters falling into each category on a scale from 0 to 1, such that the total sums to 1.0. This approach is more suitable for specialized analyses, such as investigating potential cognitive decline (e.g., in neurodegenerative conditions) [25].

A like definition exists for the majority of handwriting features. They are called evaluative features. However, some more descriptive features have a nominal scale. For example, the definition of the connection form (Table 2). In this case, several values can be chosen since several connection forms can be present in the analyzed handwriting at the same time.

Table 2. Assessment of the connection form.

Value	Meaning	Remarks
(0)	Evaluation not applicable/meaningful	None of the specific forms dominates or clearly present
(1)	Angular connections	
(2)	Soft angular connections	
(3)	Garlands	
(4)	Garlands with a loop	
(5)	Arcades	
(6)	Arcades with a loop	
(7)	Threads	
(8)	Double-curve connections	
(9)	Shorten connections	
(10)	Direct, linear connections	
(11)	School-like form	
(12)	Special, original form	

A formal and unambiguous definition of all handwriting features is the prerequisite for the objectivity and reliability of their evaluation. Even complicated and indirect ones like speed can then be correctly modeled [25].

A formal and unambiguous definition of all handwriting features is a prerequisite for achieving objectivity and reliability in their evaluation. Clear definitions ensure consistency in feature identification and scoring across different examiners and cases. This works well also for complex or indirect features— such as writing speed—which can be effectively and accurately modeled (see, for example, [26]).

All features (especially descriptive ones) are most important for author identification when they differ from the "norm". However, the natural variability of handwriting can lead to the random occurrence of such a feature in the sample. To evaluate such features, the "three-or-half" rule is used. This means that in order to say that a feature is present in the sample, it should appear at least three times or in half of the cases in which it could potentially be assessed. For example, if we are analyzing the shape of a lower loop and there are only four of them, then two occurrences are sufficient.

For textual handwriting, a total of 83 features are defined. For signatures, 45 features are considered, some of which overlap with textual handwriting while others are specific to the structural and geometric characteristics unique to signatures (e.g., overall layout or spatial configuration).

The results of these evaluations are compiled into a feature matrix, where each row corresponds to a handwriting feature, and each column represents one of the known documents (see Table 3).

Table 3. Evaluation of known samples.

Handwriting Feature	Vmin	Vmax	V1	V2	V3	V4
Letter size	3	4	4	3	4	3
Size regularity	2	4	2	4	4	0
Letter zone proportion	5	5	5	5	5	5
Letter width	2	3	2	3	3	2
Regularity of letter width	4	6	5	4	6	0
Inter-letter intervals	3	5	3	5	4	4
……………………						

2.3. Assessment of Range of Variation for Handwriting Features

Once all known documents have been evaluated, the next step is to define the range of variation for each handwriting feature. This range captures the natural variability in an individual's writing and serves as a reference for comparing the questioned sample. A special table is used to record this information. In practice, this table is a Microsoft Excel sheet—which offers ease of use and flexibility and simple o integration into a structured database for more advanced applications. The table structure is illustrated in Table 3.

In this Table, V1, V2, *etc.*, denote different known documents. Vmin and Vmax are correspondingly minimal and maximal values of the feature among the known samples. It is important to note that a feature value of 0 does not serve as a valid boundary for defining the range. A zero value typically indicates that the feature could not be evaluated—for example, due to insufficient data or the absence of relevant instances in the sample. Such cases are classified as "missing features". Missing features cannot be used in comparative analysis and are excluded from assessments of similarity or difference.

2.4. Evaluation of Handwriting Features of the Questioned Sample

The evaluation of the questioned handwriting does not differ from the evaluation of known documents. The result is a vector of evaluated handwriting features.

2.5. Assessment of Similarity Grad for Handwriting Features

The similarity grade for a specific handwriting feature quantifies the extent to which the value observed in the questioned sample (X value) aligns with the values obtained from the known samples (V values). In practical terms, this involves determining whether the X value falls within the variation range defined by the interval [Vmin–Vmax].

In [1], they propose a five-point categorical scale for this evaluation, comprising the following descriptors: clear similar (++), similar (+), inconclusive (~), different (−), clear different (−−), and missing feature/not comparable (N/C). While we follow the conceptual framework of this scale, we implement it in a numerical format to support further quantitative analysis. The similarity grade is expressed as a value ranging from 0 (complete dissimilarity) to 1.0 (complete similarity), using the following discrete levels: 0, 0,25, 0.50, 0,75, 1.0. These values are assigned based on formal comparison rules, which are outlined as follows:

- Similarity grad equals 0 if the X-value is outside the variation range (Vmin–Vmax).
- Similarity grad equals 1 if the X-value is strictly inside the variation range (Vmin–Vmax) or X-value equals Vmin or Vmax when the range is only 2 points, *i.e.*, Vmax − Vmin = 1.
- Similarity grad equals 0.5 or 0.75 if X-value equals Vmin or Vmax and the range is over 2 points, *i.e.*, Vmax − Vmin ≥ 2.
- Similarity grad gets no value (n/c) if X-value = 0 or V-value = 0 (feature cannot be evaluated).

Some evaluation examples are shown in Table 4.

Table 4. Examples for similarity grad assigning.

Feature Similarity Grad	X-Value	Vmin	Vmax	V1	V2	V3	V4
0	2	3	4	4	3	4	3
1	3	3	4	4	3	4	3
1	3	2	4	4	2	4	3
0.75	2	2	4	4	2	4	3
0.50	2	2	4	4	2	4	4
n/c	0	3	5	3	5	4	4
n/c	4	0	0	0	0	0	0

This approach is applicable primarily to evaluative features, where values can be placed on a defined ordinal or continuous scale. In contrast, descriptive features are represented not by a single value but by a set of observed characteristics. For these features, the range of variation in the known samples and the value(s) observed in the questioned sample are each expressed as sets. The assessment of similarity in this case is based on the degree of intersection between these two sets:

- If all elements of the questioned sample's feature set (X) are present in the known set (V), the similarity grade is 1.0.
- If none of the elements intersect—*i.e.*, the sets are completely disjoint—the similarity grade is 0.0.
- In partial overlaps, intermediate values (e.g., 0.25, 0.50, 0.75) may be used to represent the relative proportion of matching elements, depending on the number and relevance of shared features.

This approach ensures that both quantitative and qualitative handwriting characteristics are systematically assessed, providing a unified framework for feature-level similarity evaluation.

2.6. Assessment of Similarity Grad for Handwriting Elements

As previously mentioned, the handwriting analysis encompasses 83 distinct features, each contributing differently to the identification process. While efforts are ongoing to define the identification significance of individual features formally, it remains a complex and context-dependent task. Therefore, a more effective and practically relevant strategy is to assess the similarity of handwriting elements. A handwriting element refers to a specific structural or stylistic aspect of handwriting—such as margins, baseline alignment, slant, or letter proportions—that represents a higher-order construct composed of multiple individual features. Each handwriting element is considered to have equal evidentiary weight in the context of writer identification. The principal handwriting elements are listed in Table 5.

Table 5. Handwriting elements.

Handwriting Element	Corresponding Handwriting Features
Document	Particular parts of a document (titles, dates, signatures, *etc.*)
Handwritten text	Overall organization, line spacing and its regularity, word spacing and its regularity
Paragraphs	Presence of paragraphs, paragraph indentations, paragraph spacing
Lines	Direction of lines, line shape, regularity of lines
Words	Initial emphasize/attenuation, final emphasize/attenuation
Syllables	Hyphenation
Size	Letter size (small letters), size regularity
Size proportions	Difference in letter length (relation of the middle zone to the upper and lower zones), division of letter length (relation of the lower zone to the upper zone)
Width	Letter width, regularity of letter width, inter-letter intervals, regularity of inter-letter intervals
Slant	Slant, slant regularity
Writing pressure	Writing pressure, regularity of pressure, pressure flow, pressure rhythm
Writing speed	Writing speed, regularity of speed
Connectivity	Connectivity, dexterity in linking
Connection forms	Connection forms
Small letters	Overall letter shape, shape stability, shape enrichment/simplification
Fullness	Fullness, regularity of fullness
Lower zone	Writing pressure, fullness, shape, movement pattern, regularity
Upper zone	Writing pressure, fullness, shape, movement pattern, regularity
Capital letters	Size, width, shape enrichment/simplification
Ovals	Size, shape, regularity

Diacritics	Vertical arrangement, horizontal arrangement, form, writing pressure
Punctuation marks	Vertical arrangement, horizontal arrangement, form, writing pressure
Strokes	Stroke tenseness, stroke quality, movement suggestions/setbacks, stroke disturbances
Corrections	Type of corrections
General characteristics	Legibility, orderliness, handwriting maturity

The similarity score for a handwriting element is derived from the similarity grades of its associated features, following these rules:

- If all features of a given element have a similarity grade of 1.0, the element score is also 1.0.
- If any feature within the element has a similarity grade of 0.0, the element score is likewise 0.0, regardless of other feature scores. For example, if the shape of the left margin in the questioned document differs from that in all known samples, the entire "margins" element is considered dissimilar, and its score is set to 0.0.
- In all other cases (*i.e.*, partial similarity), the score for the handwriting element is determined based on the specific constellation of feature-level scores. This may involve weighted or contextual interpretation depending on the nature of the differences observed.

The outcome of this step is a vector of similarity scores, one for each handwriting element. This vector provides a structured, intermediate representation of the similarity between the questioned and known handwriting samples and forms the basis for the final similarity assessment.

2.7. Evaluation of Feature-Based Similarity Score

The total feature-based similarity score is calculated as the average of the similarity grades assigned to all evaluated handwriting elements. This aggregated value provides an overall measure of how closely the questioned sample aligns with the known samples in terms of observable handwriting characteristics.

However, since the completeness of the evaluation directly affects its reliability, the raw average is adjusted using a reliability coefficient. This coefficient reflects the proportion of handwriting features that were successfully evaluated and accounts for cases where missing or inconclusive data might reduce the robustness of the analysis. To model this adjustment, a sigmoid (S-shaped) function is applied:

- The reliability coefficient equals 1.0 when more than 80% of all defined features were evaluated.
- It equals 0 when fewer than 20% of the features were assessed.
- For intermediate values, the coefficient increases non-linearly with the proportion of evaluated features, reflecting a gradual improvement in confidence as more data becomes available.

This method ensures that the final feature-based similarity score not only reflects the degree of similarity observed but also accounts for the strength and completeness of the underlying data. In most practical cases, where a sufficient number of features can be assessed, the reliability coefficient tends to be close to 1.0, thus minimally affecting the uncorrected similarity score.

2.8. Congruence Analysis

The congruence analysis is a systematic, step-by-step comparison of all allographic variations of each letter present in the questioned document with their corresponding forms in the known documents. This analysis constitutes a core component of any handwriting examination, allowing for detailed scrutiny of letterforms at both graphical and dynamic levels. Where necessary, the scope of the analysis extends beyond isolated letters to include letter pairs or combinations. This is particularly relevant when certain handwriting features, such as connecting strokes or contextual variations, depend on adjacent letter interactions.

To ensure a consistent and objective approach, each comparative unit (*i.e.*, letter or letter pair) is analyzed according to a standardized set of characteristics, as outlined in Table 6.

Table 6. Characteristics for the congruence analysis.

Characteristic	Explanation
Form construction/shaping	In shaping, various aspects are considered, including the structure and proportions of the letters, the slant, the curvature of the lines, the way individual letters are connected, and other characteristic features.
Letter complexity	Letter complexity refers to the variety of strokes, curves, and details in a letter.
Movement execution for letters	Refers to the way in which the writing movements are performed to form letters. Movement execution includes various aspects, such as speed, pressure, writing direction, and continuity.
Movement execution for connections	Refers to the way in which the writing movements are performed to form connections.
Stroke length	Stroke length refers to the distance or length of individual pen strokes within letters or words. Key aspects of stroke length include letter proportions and line consistency.
Number of movements (strokes)	Number of movements refers to the total count of distinct pen strokes or movements used to form letters or parts of a signature.
Pressure distribution	Pressure distribution refers to the way writing pressure is applied and varies across different parts of a stroke, letter, or word.
Starting points (mainly refer to first letters in connected handwriting)	Position of the starting point, type of the starting point (sharp or less sharp), and angle of the starting point.
Endpoints (mainly refer to the last letters in connected handwriting)	Position of the endpoint, type of the endpoint, and angle of the endpoint.
Joins	Type of joins and their position.
Turning points	Number of turning points, positions, angles, and continuity.
Lifts	Number and position.
Stops	Number and position.
Overlapping/covering strokes	Number and location.

The output of this process is a vector of congruence scores, one for each comparative unit. Each congruence score is a quantitative value ranging from 0 to 1, reflecting the degree of graphical and kinetic correspondence between the questioned and known samples based on the defined characteristics. This vector provides a granular representation of the letter-by-letter compatibility, forming the basis for the final integration of the congruence analysis into the overall similarity assessment.

Signatures are analyzed as common comparative units. If a signature is transcribed alphabetically, each individual letter is also subject to congruence analysis.

2.9. Evaluation of Congruence Score

The total congruence score is calculated as the mean value of the individual congruence scores assigned to each comparative unit. Each comparative unit has already been evaluated on a scale from 0 to 1, reflecting the degree of graphical and kinematic compatibility between the questioned and known samples. The total congruence score thus represents the overall consistency of letterform execution across all units examined. This score provides a quantitative measure of structural similarity at the micro-level of handwriting, complementing the feature-based similarity score obtained in previous steps. Together, they contribute to a more robust and balanced final evaluation of authorship likelihood.

2.10. Calculation of Total Similarity Score

The total similarity score is derived as a weighted combination of the two core components of handwriting comparison: the feature-based similarity score and the congruence score. It provides an integrated, quantitative assessment of the degree of similarity between the questioned and known handwriting samples. The weights are assigned based on the nature of the handwriting under examination. For textual handwriting, the feature-based score is given greater emphasis, reflecting the broader structural aspects of writing style. In contrast, for signatures, where individual letterforms carry more weight and variation is often more pronounced, the congruence score is prioritized. In all cases, the total similarity score lies within the interval [0,1], ensuring interpretability and consistency across different examination contexts. This score serves as the primary quantitative indicator guiding the expert's conclusion regarding authorship.

2.11. Expert Decision

The total similarity score serves as the primary basis for the expert's decision regarding authorship. While the score provides a quantitative measure of similarity between the questioned and known handwriting samples, it should not be interpreted as a direct probability. Such an interpretation, especially at higher score levels, risks oversimplifying the nuanced nature of handwriting comparison. Instead, the similarity score functions as a threshold indicator within a broader evaluative framework. Based on extensive empirical observations from casework, it can be stated with reasonable confidence that a score below 0.8 generally supports a rejection of the hypothesis that the questioned handwriting was produced by the known author. In other words, such a score indicates insufficient similarity to justify identification.

However, reaching a positive conclusion of authorship requires careful consideration of contextual factors identified during the pre-assessment stage. Thus, the expert decision should not be solely score-driven but rather result from a holistic integration of quantitative analysis and qualitative judgment within the forensic context. The total similarity score strengthens the objectivity and reproducibility of the process and supports expert interpretation.

2.12. Databank

All acquired, processed, and calculated data is stored in the system's databank. This information is also utilized for statistical research in an anonymized form—real cases, personal names, and authentic handwriting samples are not accessible for this purpose. The databank serves as a valuable resource for the ongoing refinement of the examination procedures and the development of new methodological adaptations.

3. Theory/Calculation

To formalize the presentation of the described handwriting examination procedure, let us denote:

- S as the total similarity score,
- F as the feature-based similarity score,
- C as the congruence score.

The total similarity score is then expressed as a weighted sum of the feature score and the congruence score:

$$S = a \cdot F + b \cdot C \tag{1}$$

Here, a and b are weighting coefficients reflecting the relative importance of each component. Empirical values used in practical casework are: for textual handwriting: $a = 0.5$, $b = 0.5$; for signatures: $a = 0.4$, $b = 0.6$.

The feature-based similarity score F is computed as the average similarity of handwriting elements, adjusted by a reliability coefficient r:

$$F = \frac{\sum f_i}{m} \cdot r \tag{2}$$

where:

- f_i is the similarity score for the i-th handwriting element,
- m is the number of handwriting elements evaluated,
- r is the reliability function, depending on the number of features assessed.

The reliability coefficient r is modeled using a sigmoid function, capturing how the confidence in the similarity score increases with the number of features evaluated:

$$r = \frac{1}{1 + e^{-10(1.88\frac{n}{N}-0.83)}} \tag{3}$$

where:

- N is the total number of defined handwriting features ($N = 83$ for textual handwriting, $N = 45$ for signatures),
- n—the number of evaluated handwriting features

This function ensures that the reliability approaches 1 as the evaluated features approach completeness and approaches 0 when the evaluation is based on insufficient data. Figure 1 illustrates the behavior of the reliability function.
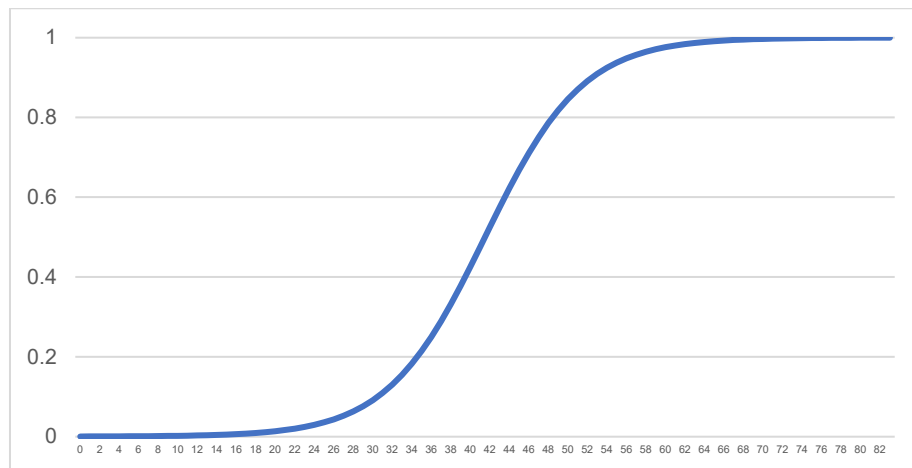
Figure 1. Reliably as a function of number of handwriting features.

The congruence score *C* is calculated as the mean value of the congruence scores across all comparative units:

$$C = \frac{\sum c_j}{k} \tag{4}$$

where:
- c_j is the congruence score for the *j*-th comparative unit (individual letters or letter pairs),
- k is the total number of comparative units assessed.

The rules for evaluating individual feature similarity scores f_i and congruence scores c_j are described in the preceding sections and are therefore not repeated here in a formalized form.

4. Results

The examined handwriting cases typically exhibit significant variability in terms of both structure and complexity. This variability is often influenced by differences in the number and quality of known samples, as well as the availability of relevant background information for each case. Despite these differences, the use of a standardized comparison procedure allowed for certain statistical generalizations to be made.

Notably, some handwriting features, while essential for comprehensive handwriting analysis, do not significantly contribute to the comparative evaluation between questioned and known samples. For example, characteristics such as letter size and line spacing are almost never decisive in distinguishing between writers. These features, though routinely observed, rarely showed sufficient variation to impact authorship attribution.

Table 7 presents our preliminary statistical findings, indicating the relative discriminative power of various handwriting features. The "level of discrimination" refers to the proportion of cases in which a given feature displayed observable differences between the questioned and known samples.

Table 7. Most discriminating handwriting features and elements.

Handwriting Feature/Element	Level of Discrimination	Remarks
Pressure flow and pressure rhythm	0.60	Pressure fluctuations and its certain patterns.
Stroke quality and security	0.50	Observed degree of confidence, control, fluency, automaticity and consistency within handwriting.
Stroke tension	0.35	Shows the degree of stiffness or suppleness in the way the written trail (ductus) progresses along the letter/word or signature. It relates to contraction and release in the pattern of writing.
Fullness	0.35	Mainly assessed on the ovals and loops.
Connection forms	0.30	Compared are connections in the same letters and letter pairs.
Specific stroke disturbances	0.25	Trembling, tremor, doubling of strokes, stroke breaks, deformations, trailing, breakpoints, *etc.*
Writing speed	0.25	Evaluated indirectly.

Inter-letter intervals	0.25	Handwriting text and signature. Interval width and its regularity.
Letter form	0.25	Including form enrichment/simplification.
Writing pressure strength	0.20	The pressure itself can be properly analyzed only in original documents, but the comparison can be made on copies if they are of good quality.
Fluency of connections	0.20	Smoothness, rhythm, and continuity with which strokes transition between letters or parts of letters. Reflects how effortlessly the writer moves from one element to the next.
Fluidity	0.20	Prevalence of linear (straight) versus curved (rounded) stroke forms.
Capital letters	0.20	Here, size and width are analyzed. More often, the width is different.
Letter zone proportions	0.15	Relation of the upper and lower zones to the middle.
Diacritics	0.15	Location and pressure.
Number of strokes or movements	0.15	Mainly applicable for signatures.
Letter width	0.10	Mainly, the regularity is different.
Geometric form	0.10	Applicable only for signatures.
Graphic complexity	0.10	Applicable only for signatures.
Ovals	0.05	Shape and size.

The results of the congruence analysis are, unsurprisingly, highly individual and depend heavily on the specific handwriting samples examined. Nevertheless, some general observations can be made, even though they do not claim statistical significance. Discrepancies were most frequently observed in approximately 20% of cases for the following letters: d, f, k, m, w, and S. A slightly lower discrepancy rate, around 15%, was noted for a, b, e, g, r, t, E, and V. The majority of handwriting samples analyzed were written in German and English.

5. Discussion

The presented procedure has demonstrated both reliability and robustness in practice. Nonetheless, there remain several areas that warrant further refinement and development.

Firstly, at present, no explicit weighting scheme is applied to individual features; instead, this limitation is partially addressed through the introduction of handwriting elements. However, the absence of a formal methodology and sufficient data precludes statistically valid evaluations. Additionally, it is worth considering that feature weights may not be static but rather should vary depending on contextual or background information—an informal consideration that already plays a role in expert practice. A formalized, data-driven weighting model would therefore, be highly beneficial to enhance objectivity and consistency.

Secondly, as mentioned earlier, the assessment of handwriting samples is still conducted manually. Existing software tools, including those incorporating AI techniques, have yet to meet the rigorous demands of professional handwriting examination. While data scarcity remains a major obstacle for AI training, there is also a more fundamental methodological issue. Most AI applications in this field are designed to provide a binary decision regarding the authenticity of a questioned document. However, this framing may represent a false target. As noted in [18], one of the primary reasons for failure in AI-based projects stems from flawed project objectives, which are often determined by managerial decisions misaligned with real needs. In handwriting examination, this issue is reflected in the fact most existing AI applications in handwriting analysis have been developed by computer scientists, often without domain expertise in forensic handwriting. Consequently, the design of such tools may prioritize computational elegance over practical relevance. To address this, handwriting experts must assume a leading role in guiding AI development in this domain. Rather than attempting to replicate the final decision-making process, AI could be more effectively applied to the automated evaluation of individual handwriting features. These assessments could then be integrated using traditional software solutions, with the final judgment left to human experts. In this way, AI acts as a supportive analytical tool, not a replacement for expert interpretation.

An additional area where AI could prove particularly useful is in the detection of computer-generated handwriting forgeries, which are becoming increasingly sophisticated. Such forgeries often pose significant challenges even for seasoned experts, highlighting the need for technological support.

Finally, by narrowing the scope of AI applications to specific subtasks—such as feature recognition or forgery detection—the requirements for training data become more manageable. In fact, training data for these subtasks could be generated artificially in large quantities by dedicated AI programs, effectively enabling an "AI trains AI" paradigm.

6. Conclusions

The procedure presented in this study offers a more formalized and objective alternative to traditional handwriting examination practices. Having been actively employed by the authors in real-world forensic contexts, the approach has demonstrated both practical relevance and operational feasibility. Consistent results in practice further support confidence in its applicability.

Future enhancements, particularly those addressing the limitations discussed earlier, can further strengthen the procedure. As more comprehensive datasets become available, the potential for deeper statistical verification will allow for further refinement.

Author Contributions

Both authors contributed equally.

Ethics Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Funding

This research received no external funding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. ENFSI. *Best Practice Manual for the Forensic Handwriting Examination*, 4th ed.; ENFSI: Wiesbaden, Germany 2022.
2. ANSI/ASB. *Standard for Examination of Handwritten Items*, 1st ed.; ANSI/ASB: Colorado Springs, CO, USA, 2022.
3. Hicklin RA, Eisenhart L, Richetelli N, Miller MD, Belcastro P, Burkes TM, et al. Accuracy and reliability of forensic handwriting comparisons. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2119944119. doi:10.1073/pnas.2119944119.
4. Crot S, Marquis R. A comparative review of error rates in forensic handwriting examination. *J. Forensic Sci.* **2024**, *69*, 2127–2138. doi:10.1111/1556-4029.15589.
5. Kruger J, Dunning D. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Personal. Soc. Psychol.* **1999**, *77*, 1121–1134. doi:10.1037/0022-3514.77.6.1121.
6. Crown N, Marquis R, Kupferschmid E, Dziedzic T, Belic D, Kerzan D. Error mitigation in forensic handwriting examination: The examiner's perspective. *Forensic Sci. Res.* **2024**, *9*, owae065. doi:10.1093/fsr/owae065.
7. Vasquez JL, Ravelo-García AG, Alonso JB, Dutta MK, Travieso CM. Writer identification approach by holistic graphometric features using off-line handwritten words. *Neural Comput. Appl.* **2020**, *32*, 15733–15746. doi:10.1007/s00521-018-3461-x.
8. Crawford AM, Berry NS, Carriquiry AL. A clustering method for graphical handwriting components and statistical writership analysis. *Stat. Anal. Data Min.* **2021**, *14*, 41–60. doi:10.1002/sam.11488.
9. Crawford AM, Ommen DM, Carriquiry AL. A statistical approach to aid examiners in the forensic analysis of handwriting. *J. Forensic Sci.* **2023**, *68*, 1768–1779. doi:10.1111/1556-4029.15337.
10. Johnson MQ, Ommen DM. Handwriting identification using random forests and score-based likelihood ratios. *Stat. Anal. Data Min.* **2022**, *15*, 357–375. doi:10.1002/sam.11566.

11. Srihari SN, Srinivasan B, Desai R. Questioned Document Examination Using CEDAR-FOX. *J. Forensic Doc. Exam.* **2018**, *28*, 15–26. doi:10.31974/jfde28-15-26.

12. Van Erp M, Vuurpijt L, Franke K, Schomaker L. The Wanda Measurement Tool for Forensic Document Examination. *J. Forensic Doc. Exam.* **2018**, *28*, 5–14. doi:10.31974/jfde28-5-14.

13. Chernov Y. *Компьютерные методы анализа почерка [Computer Methods of Handwriting Analysis]*; IHS Books: Zurich, Switzerland, 2021.

14. Miller JJ, Patterson RB, Gantz DT, Saunders CP, Walch MA, Arch M, et al. A Set of Handwriting Features for Use in Automated Writer Identification. *J. Forensic Sci.* **2017**, *62*, 722–734. doi:10.1111/1556-4029.13345.

15. Geradts Z, Franke K. (Eds.) *Artificial Intelligence (AI) in Forensic Sciences*; Wiley: Hoboken, NJ, USA, 2024.

16. Zhau H, Li H. Handwriting identification and verification using artificial intelligence-assisted textural features. *Sci. Rep.* **2023**, *13*, 21739. doi:10.1038/s41598-023-48789-9.

17. Marcinowski M. Top interpretable neural network for handwriting identification. *J. Forensic Sci.* **2022**, *67*, 1140–1148. doi:10.1111/1556-4029.14978.

18. Ryseef J, De Bruhl B, Newberry J. The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed. RAND Research Report, 13 August 2024. Available online: https://www.rand.org/pubs/research_reports/RRA2680-1.html (accessed on 12 April 2025).

19. Huber RA, Headrick AM. *Handwriting Identification: Facts and Fundamentals*; CRC Press: Boca Raton, FL, USA, 1999.

20. Harralson HH, Miller LS. *Huber and Headrick's Handwriting Identification. Facts and Fundamentals*; CRC Press: Boca Raton, FL, USA, 2018.

21. Morris R. *Forensic Handwriting Identification. Fundamental Concepts and Principles*, 2nd ed.; Elsevier Academic Press; London, UK, 2021.

22. Michel L. *Gerichtliche Schriftvergleichung. Eine Einführung in Grundlagen, Methoden und Praxis*; Walter de Gruyter & Co: Berlin, Germany, 1982.

23. Seibt A. *Forensische Schriftgutachten: Einführung in Methoden und Praxis der forensischen Handschriftuntersuchungen*; Verlag C.H. Beck: München, Germany, 1999.

24. Seibt A. *Unterschriften und Testamente. Praxis der Forensischen Schriftuntersuchung*; Verlag C.H. Beck: München, Germany, 2008.

25. Chernov Y. Handwriting Markers for the Onset of Alzheimer's Disease. *Curr. Alzheimer Res.* **2023**, *20*, 791–801. doi:10.2174/0115672050299338240222051023.

26. Chernov Y, Nauer MA. (Eds.) Reliability of Evaluation of Handwriting Signs. In *Handwriting Research: Forensics & Legal*; IHS Books: Zurich, Switzerland, 2023; pp. 37–56. doi:10.61246/ihs2/ycman037056.