

Article

Assessment of the Identification Value of Handwriting Features and Individual Graphical Elements in Forensic Handwriting Examination

Yury Chernov *

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

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

Received: 10 August 2026; Revised: 7 September 2026; Accepted: 18 September 2026; Available online: 30 September 2026

ABSTRACT: Forensic handwriting examination requires not only the assessment of similarities and differences but also an evaluation of their relative importance for writer identification. This methodological study proposes a formalized framework for quantifying the identification value of handwriting features and individual graphical elements. The model distinguishes identification value from similarity grade and combines three components: objective discrimination power based on feature prevalence, relative discrimination power derived from forensic comparison data, and case-specific presence and representation. Identification values are subsequently transformed into weights for use in a formalized handwriting comparison model. Data from the HSDetect handwriting database and the HSForen forensic examination database are used to illustrate the calculation and practical behavior of the proposed components. The numerical values obtained from the current databases are treated as preliminary and illustrative rather than as validated population parameters or established measures of forensic performance. The proposed framework provides a transparent and empirically adjustable method for differentiating the contribution of handwriting findings while preserving expert interpretation. Further research using larger datasets, independently established ground truth, reliability testing, and external validation is required before the resulting weights can be considered validated for forensic casework.

Keywords: Forensic handwriting examination; Writer identification; Handwriting comparison; Questioned documents; Handwriting features; Individual letters; Discriminative power; Identification value; Quantitative evaluation; Feature weighting

1. Introduction

Forensic handwriting examination is based on the comparison of characteristics observed in questioned and known writings. The examiner identifies similarities and differences, considers the writer's natural variation, and evaluates whether the overall pattern of findings supports common or different authorship [1–8]. However, not all findings have the same significance. Handwriting characteristics differ in how common they are, how consistently the same writer reproduces them, and how effectively they distinguish between different writers.

Treating all similarities and differences as equally important may therefore lead to an oversimplified assessment. Numerous agreements involving common or weakly discriminative characteristics do not necessarily provide stronger support for common authorship than a smaller number of agreements involving unusual and stable writing habits. Similarly, a difference in a highly variable characteristic may be less meaningful than a difference affecting a consistent and distinctive element of the writing.

Handwriting examination has long recognized that the significance of a characteristic depends on its individuality, consistency, and occurrence in the available material [9]. In practice, however, these factors are still assessed largely through qualitative expert judgment. Characteristics may be described as common, unusual, significant, or highly individual, but the basis for assigning these levels of importance is not always stated explicitly. This can make the weighting of handwriting evidence difficult to reproduce, compare, and validate.

A formalized approach requires a clear distinction between the similarity grade of a finding and its identification value [10]. The similarity grade expresses how closely a handwriting feature or individual letter in the questioned writing corresponds to that in the known writing. The identification value describes how important this similarity or difference is for distinguishing between writers. These two quantities, therefore, serve different purposes. A characteristic may show a high degree of similarity but have little identification value because it is common or only weakly discriminative. Agreement in normal line spacing or a horizontal baseline, for example, may be almost complete but still provide limited support for common authorship. Conversely, a similarity or difference involving an uncommon and highly discriminative characteristic may have a much greater influence on the overall assessment. Similarity describes the result of the comparison; identification value determines the weight assigned to that result.

The present study considers two main types of analytical units: general handwriting features and individual letters. For simplicity, the term “letters” is used here in a broad sense to include lowercase and uppercase letters, numerals, punctuation marks, diacritical strokes, and other graphical signs occurring in handwriting. General handwriting features describe broader properties of the writing system, including letter size, proportions, slant, spacing, baseline direction, connectivity, pressure, rhythm, stroke tension, movement form, and spatial organization. A handwriting sign represents the general object of assessment, whereas a handwriting feature describes its particular manifestation. For example, letter size is a handwriting sign, while very small, small, medium, large, and very large are its possible features. The importance of considering handwriting complexity, feature structure, and the relationships between individual elements in forensic comparison has also been emphasized in [11,12]

A letter is not treated merely as a static visual form. It is regarded as an integrated structural and dynamic graphic-motor unit. Its examination may include the number and sequence of strokes, movement direction, starting and terminal points, turning points, proportions, loops, crossings, retracing, connections with neighboring letters, rhythm, pressure distribution, and indications of fluency, hesitation, or automation. This distinction is important because letters with a similar general appearance may have been produced through different movement sequences. Conversely, somewhat different visible forms may arise from closely related habitual movements. The examination of individual letters should therefore consider not only their outline but also the organization of the movement through which they were produced.

Not every letter provides the same amount of potentially identifying information. Some letters have a simple construction and allow relatively few variations. Others contain several strokes, loops, crossings, changes of direction, or separately executed elements and therefore offer more possibilities for writer-specific execution. Their identification value may also depend on their position within a word, their connections with neighboring letters, and the number of evaluable occurrences in the available material.

For general handwriting features, population distinctiveness can be assessed relatively directly. A defined feature, such as very small handwriting, strong leftward slant, absence of interletter spacing, or increased pressure in diacritical strokes, can be recorded across a reference database. Features that occur in only a small proportion of writers have greater objective discrimination power than common manifestations. Rarity alone,

however, is not sufficient. An uncommon feature may vary considerably within the handwriting of the same person or may be strongly affected by writing conditions. General pressure strength, for example, may depend on the pen, paper, writing surface, posture, and quality of the available reproduction.

A characteristic must therefore be evaluated not only according to its prevalence but also according to its performance in same-writer and different-writer comparisons. A useful handwriting characteristic should remain sufficiently consistent within the writing of the same person while showing differences between different writers. This relationship is represented in the proposed framework by relative discrimination power. A feature that is stable but common will have limited discriminative value, as will a rare feature that varies unpredictably within the same writer.

The effective contribution of a characteristic also depends on whether it can be assessed reliably in the particular case. A feature or letter with high general discriminative potential cannot contribute meaningfully when it is absent, poorly reproduced, insufficiently represented, or not comparable between the questioned and known writings. This case-specific component is particularly important for individual letters, since short documents often contain only part of the alphabet and may provide only one or two examples of certain characters.

In the present study, identification value is defined as the relative capacity of a handwriting feature or individual letter to contribute to discrimination between same-writer and different-writer propositions. For handwriting features, the determination is based on objective discrimination power, relative discrimination power, and case-specific presence and evaluability. For individual letters, the model primarily considers their relative discrimination power and their representation in the particular examination, while expected linguistic frequency may provide an additional objective indication of their anticipated availability in written text.

The calculated identification values are transformed into weights that determine the contribution of each feature or letter to the overall comparison score. The observed similarity remains a separate quantity. Thus, the identification value determines how strongly a finding is weighted, whereas the similarity or difference observed between the questioned and known writings determines the direction and degree of its contribution.

The proposed framework is intended to make the weighting of handwriting evidence more explicit, differentiated, and empirically testable. It does not replace expert interpretation. The examiner must still consider natural variation, document comparability, technical limitations, disguise, imitation, and the consistency of the findings as a whole. Formalization provides a structured way of explaining why some observations receive greater importance than others.

The aim of the present methodological study is to develop and formalize a quantitative framework for assessing the identification value of general handwriting features and individual graphical elements. Data from the HSDetect and HSForen databases are used to illustrate the calculation and behavior of the proposed model components rather than to establish definitive population parameters. Particular attention is given to the distinction between similarity and identification value, the different treatment of handwriting features and graphical elements, and the influence of case-specific representation on the contribution of individual findings.

2. Materials and Methods

2.1. Study Design

The study was designed primarily as a methodological investigation aimed at developing a formalized framework for assessing the identification value of handwriting features and individual letters. Available data from the HSDetect and HSForen databases were used to illustrate the calculation of the proposed model components and their practical application. The empirical values presented in this study should therefore be regarded as illustrative rather than as definitive estimates of population parameters. Their

purpose is to demonstrate the behavior of the proposed framework and the relative weighting of different analytical units.

The empirical analysis was conducted at two complementary levels. First, the distribution of defined handwriting characteristics in a reference population was examined to estimate their prevalence and population distinctiveness. Second, the same characteristics were evaluated in same-writer and different-writer comparisons to determine the extent to which they remain stable or vary within the handwriting of one person and distinguish between different writers.

The study considered both general handwriting features and individual letters. Individual letters were not treated simply as static visual forms but as structural and dynamic graphic-motor units. Their assessment, therefore, included both the visible organization of the written character and the observable or inferable characteristics of its execution.

The empirical parameters obtained from these analyses were subsequently incorporated into the formal model of identification value presented in Section 3.

2.2. Handwriting Databases

The empirical analyses were based on two complementary databases: the HSDetect handwriting database, which contains systematically evaluated handwriting samples collected for research and methodological purposes, and the HSForen database, which contains material from forensic handwriting examinations. Together, the databases comprise evaluated handwriting material from several hundred writers.

HSDetect and HSForen were used because both databases were developed using the same operational definitions and assessment protocol as the proposed identification-value model. The handwriting signs, features, and individual graphical elements are coded according to predefined criteria and comparable measurement scales, which allows population prevalence, same-writer consistency, different-writer discrimination, and case-specific evaluability to be calculated directly. Publicly available handwriting databases, for example [13–15], are generally not structured according to this assessment protocol and often contain handwriting images or trajectories rather than systematically coded forensic features. Their direct inclusion would therefore require retrospective re-evaluation of the samples and could introduce differences caused by coding procedures rather than genuine handwriting variation. Using HSDetect and HSForen ensures methodological consistency between the reference data, forensic comparison data, and the calculations applied in the present study.

The available material includes naturally produced free texts as well as handwriting samples collected under controlled research conditions. Only samples of sufficient technical quality and legibility were included in the analysis. More specifically, a sample was considered suitable when:

- The document quality permitted a reliable assessment of the relevant handwriting characteristics;
- The material contained a sufficient amount of writing for the evaluation of general handwriting features;
- letters included in the letter-level analysis were represented by at least one technically evaluable occurrence;
- questioned and known writings used in pairwise comparisons were sufficiently comparable with regard to writing system and document type.

Where both original documents and electronic reproductions were available, the originals were given preference. Characteristics that depend on fine line quality, pressure variation, stroke sequence, pen lifts, retracing, or other dynamic aspects of execution were assessed from copies only when the image quality allowed these features to be evaluated with sufficient reliability.

For each sample, available metadata were recorded, including writer identity, document type, language and writing system, writing instrument, document format, date of production, and whether the material was available as an original document or a reproduction. Demographic information was used only when relevant to the definition of the reference population.

For estimating population prevalence, each writer was counted only once in a given calculation, even when multiple samples from the same person were available. This procedure prevented writers with multiple documents from having a disproportionate influence on the resulting frequency estimates. In comparison analyses, more than one sample from the same writer could be included, provided that any resulting dependence between observations was accounted for in the organization and interpretation of the data.

2.3. Operational Definitions and Coding

Each handwriting feature and letter characteristic was defined operationally. The definitions specified the object of assessment, the applicable measurement or classification scale, the criteria for identifying the characteristic, and the conditions under which two observations were considered to agree or differ. They also established the acceptable range of variation within a category, the minimum technical quality required for evaluation, and, where relevant, the minimum number of occurrences needed for a sufficiently stable case-specific assessment.

In the present study, the unit of analysis depended on the component being estimated. For objective discrimination power D , the unit of analysis was the handwriting sample, and the frequency of each predefined feature was determined across the available HSDetect samples. For relative discrimination power G of handwriting signs, the unit of analysis was the forensic comparison stored in HSForen. For individual letters, the analysis was performed at the level of the corresponding letter within an evaluable forensic comparison. Case-specific presence and evaluability C were determined from the questioned and known handwriting material available in the individual examination. The handwriting signs and feature manifestations included in the present calculations are those reported in Tables 1–3, while the individual letters included in the letter-level analysis are presented in Table 4.

Because handwriting signs differ substantially in their nature, they cannot all be represented adequately using a single type of scale. Depending on the characteristic and the quality of the available material, observations were coded using binary, categorical, ordinal, or continuous measures. Binary coding was applied when a characteristic could be classified reliably as present or absent. Categorical coding was used for mutually exclusive manifestations, such as predefined categories of letter form. Ordinal scales were employed when a characteristic varied in degree but could not be measured directly with sufficient precision, or when such measurement would provide little practical benefit. Continuous scales were used when a property such as angle, proportion, spacing, frequency, or similarity could be expressed numerically. In HSDetect, continuous values are generally represented on a scale from 0 to 1.

For example, the handwriting sign *letter size* is defined as the height of letters within the middle zone, measured along the slant of the letter axis rather than vertically. It represents the distance from the baseline to the upper boundary of the middle zone. The upper and lower zones (ascenders and descenders) are not included in this measurement. Only internal letters within words are considered; word-initial and word-final letters are excluded. The sign includes five features: very small (<1 mm), small (1–2 mm), medium (2.0–3.5 mm), large (3.5–5.5 mm), and very large (>5.5 mm). The proportion of letters belonging to each category provides a quantitative value for that category on a continuous scale.

Letter size is considered stable when at least 80% of the evaluated letters in a sample fall within the same category. *Size stability* is therefore represented as a binary feature. In principle, stability could also be expressed on a continuous scale. However, our practical experience indicates that this would complicate the evaluation without providing substantial additional value.

An example of operational coding for the letter size is shown in Figure 1.

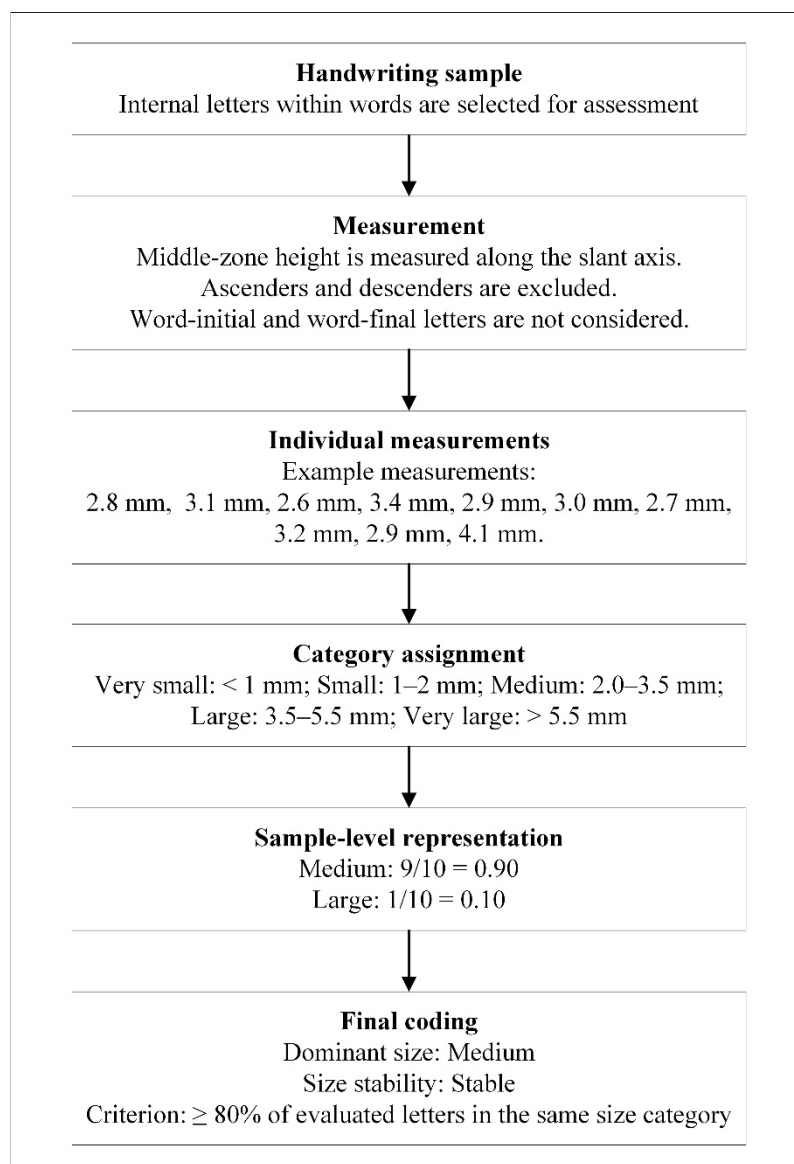


Figure 1. Example of operational assessment of letter size.

In HSForen, where forensic handwriting examination cases are stored, the same sign is represented on a seven-category scale that is better suited to direct comparison between questioned and known writings: very small, small, rather small, medium, rather large, large, and very large. Seven-category scales are commonly used within the present examination framework for many handwriting signs.

Whenever direct and reproducible measurements were possible, quantitative values were used. This applied particularly to geometric and spatial characteristics, including letter height, letter width, spacing, zone proportions, slant angles, stroke-length ratios, and the relative positions of graphical elements. Such measurements reduce dependence on subjective classification and make the resulting data more suitable for statistical analysis.

Other handwriting characteristics describe aspects of movement organization and execution that cannot currently be measured directly with comparable forensic reliability. Examples include letter and connection forms, pressure rhythm, movement fluency, stroke tension, and specific movement patterns. These characteristics were assessed using standardized categorical or ordinal scales with clearly defined categories and reference examples. For example, dominant connection forms include angular, garland, arcade, thread, school-like, direct, shortened garland, shortened arcade, soft angular, looped garland, looped arcade, and

double-curve forms. A non-evaluable category was treated as missing information rather than as a substantive feature level.

Continuous scales allow handwriting variation to be represented more gradually, reduce information loss due to rigid category boundaries, and simplify subsequent mathematical modeling. They also provide a direct basis for calculating distances and similarity scores between questioned and known writings. The framework, therefore, supports both discrete and continuous representations. The choice of scale depends on the nature of the handwriting characteristic, the reliability with which it can be observed or measured, the quality of the available material, and the level of precision required.

An individual letter is not assessed solely according to its visible outline. Its analysis may include the number and sequence of strokes, movement directions, starting and terminal points, changes of direction, turning points, proportions between component parts, loops, crossings, retracing, overwriting, connections with adjacent letters, movement continuity, rhythm, pressure distribution, speed-related characteristics, pauses, and indications of automation or hesitation. The term individual letter, therefore, refers to the complete observable organization of the letter and its inferred execution process.

Handwriting samples were assessed using a standardized coding framework [16,17]. Examiners worked according to operational definitions, coding instructions, measurement thresholds, and reference examples for the relevant features. Where a feature required measurement, digital tools were used to determine angles, proportions, distances, or relative stroke dimensions. Measurements were performed on standardized digital images or original documents under controlled examination conditions.

To reduce the influence of expectation, database coding was separated, as far as practicable, from the subsequent calculation of identification values. Knowledge of whether a comparison involved the same or different writers was not used to modify the operational coding of the observed characteristics.

The present databases were not created as formal inter-examiner reliability studies. Some samples and characteristics were assessed independently by more than one examiner during database development, but systematic reliability statistics are not available for all variables and samples. Accordingly, the present study does not provide a comprehensive quantitative assessment of inter-examiner reliability. Standardized operational definitions, measurement rules, coding instructions, thresholds, and reference examples instead supported reproducibility. A dedicated reliability study, using independent examiners and scale-appropriate agreement statistics, remains necessary for further validation of the framework.

2.4. Statistical Analysis

Statistical analysis was performed separately for population prevalence, same-writer consistency, different-writer discrimination, and case-specific availability. Population prevalence was calculated as the number of independent writers displaying a defined feature divided by the total number of writers. Multiple samples from one writer were consolidated at the writer level for this analysis.

Same-writer consistency was estimated from evaluable comparisons between samples produced by the same person, whereas different-writer discrimination was estimated from evaluable comparisons between samples produced by different persons.

Potential overlap between handwriting characteristics was considered conceptually when defining the analytical units. Characteristics representing substantially the same handwriting property were avoided or combined where appropriate. However, statistical dependencies among the variables have not yet been systematically evaluated. Such dependencies will require investigation during empirical validation of the framework.

2.5. Ethical and Data-Protection Considerations

The samples from the databases were analyzed in an anonymized form. Information identifying individual writers was not included in the analytical dataset. Handwriting samples were used in accordance with the applicable consent, institutional, legal, and data-protection requirements.

The analyses were performed at group and feature levels. No conclusions regarding the identity, personality, health, or other personal characteristics of individual database participants were drawn in the present study.

3. Theory/Calculation

The concept of identification value is relevant to different handwriting examination procedures, but it is particularly useful in formalized approaches. In such methods, the two main components of handwriting comparison—assessment of similarity and evaluation of the identification value of the observed findings—are integrated within a single formal model. The overall similarity score between two analyzed handwritings is a composition of handwriting feature similarity values and the congruence value of individual letters and other graphical signs:

$$S = \alpha \cdot H + \beta \cdot R \quad (1)$$

where H is the weighted feature similarity score, R is the congruence score, and α and β are coefficients. In the absence of empirical calibration, equal coefficients ($\alpha = 0.5$, $\beta = 0.5$) were used as a provisional neutral weighting of feature-level and letter-level evidence.

Feature similarity is calculated as:

$$H = \frac{\sum w_i \cdot h_i}{\sum w_i} \quad (2)$$

where h_i is the similarity of the i -th handwriting feature and w_i its weight.

The congruence score is calculated analogously:

$$R = \frac{\sum v_j \cdot r_j}{\sum v_j} \quad (3)$$

where r_j denotes the congruence of the j -th graphical element and v_j its weight.

The weights w_i and v_j are calculated based on identification values, which are described below.

3.1. Concept of Identification Value

Handwriting features and individual letters differ substantially in their ability to discriminate between writers. Some characteristics are widespread in the relevant writing population and therefore provide only limited evidence of common or different authorship. Others are uncommon, structurally complex, consistently reproduced by the same writer, and rarely shared by different writers. Such characteristics may make a substantially greater contribution to writer identification.

The evidential importance of a characteristic cannot, however, be determined from rarity alone. A rare feature may have limited value when it varies considerably within the handwriting of the same person. Conversely, a highly stable characteristic may have little discriminative value when it is common among many writers. In addition, even a theoretically informative characteristic cannot contribute to a particular examination when it is absent, insufficiently represented, poorly reproduced, or otherwise not reliably comparable in the questioned and known writings.

To account for these factors, the framework introduces an identification value Z as a quantitative measure of the relative importance of a handwriting feature or an individual letter in writer discrimination.

Identification value is defined as the relative capacity of a handwriting feature or individual letter, considered in its structural and dynamic execution, to contribute to discrimination between same-writer and different-writer propositions in a particular forensic examination.

The model combines three components:

$$Z = D \cdot G \cdot C \quad (4)$$

where D is objective discrimination power, G is relative discrimination power, and C is case-specific presence and evaluability.

A feature receives a high identification value only when it is distinctive in the population, performs well in writer discrimination, and is sufficiently available for comparison in the examined writings.

The general principle applies to both handwriting features and individual letters, although the individual components are calculated somewhat differently for these two types of analytical units. The components are combined multiplicatively because a serious limitation in any one component should reduce the effective contribution of the analytical unit. A rare characteristic that is unstable within one writer should not receive a high weight. Similarly, a highly discriminative letter cannot contribute fully when only insufficient comparable occurrences are available.

3.2. Objective Discrimination Power

Objective discrimination power describes the population distinctiveness of a handwriting feature [18,19]. It reflects how uncommon a defined feature is within an appropriate reference population. The underlying principle is that a characteristic observed in only a small proportion of writers generally has greater potential to distinguish between writers than a characteristic occurring frequently.

For a handwriting feature, empirical prevalence (E) is calculated as:

$$E = \frac{k}{n} \quad (5)$$

where (k) is the number of handwriting samples containing the feature and (n) is the total number of samples in the reference database.

Objective discrimination power is then derived from empirical prevalence:

$$D = -\log(E) \quad (6)$$

Thus, each tenfold reduction in prevalence increases objective discrimination power by one unit. The logarithmic transformation is preferable to a simple inverse transformation because it moderates the influence of extremely rare characteristics. It increases the relative weight of uncommon features while compressing very large differences at the rare end of the distribution. The resulting continuous measure is therefore more suitable for integration into a broader weighting model.

The value of D depends, in principle, on the population from which prevalence is estimated. Handwriting features may vary according to writing system, language, educational background, writing style, age group, historical period, document type, and other contextual factors. A feature that is rare in one population may be more common in another. HSDetect, however, contains heterogeneous groups, and prevalence was therefore calculated for the entire reference population. In practical terms, the observed differences between subgroups were relatively small, whereas subdivision of the database would substantially reduce sample sizes and the reliability of the estimates.

In the proposed framework, objective discrimination power is calculated directly for handwriting features and is not assigned in the same way to individual letters. This distinction follows from the different analytical nature of these units. A handwriting feature represents a clearly defined manifestation that can be assessed across a population. An individual letter, in contrast, is a complex graphic-motor unit composed

of several interrelated structural and dynamic characteristics. The frequency of the letter itself is largely determined by language and does not directly reflect the writer's individuality.

The value of D is predefined for each feature based on reference data. In cases of disagreement between the questioned and known writings, the D value corresponding to the feature observed in the questioned document is applied, while the degree of disagreement is represented separately by the similarity grade h_i .

3.3. Relative Discrimination Power

Objective discrimination power reflects the rarity of a handwriting feature in a reference population, but rarity alone does not establish how effectively that feature distinguishes writers in actual comparisons. A handwriting characteristic may be uncommon but vary substantially within the handwriting of the same person. Conversely, a characteristic may be highly stable within one writer but occur in a similar form among many different writers. Neither characteristic would provide strong discrimination on its own.

This practical performance is described by relative discrimination power. It reflects the extent to which a handwriting sign or individual letter satisfies two complementary requirements: it differs between writings produced by different writers and remains sufficiently consistent across writings produced by the same writer.

An important distinction is that relative discrimination power is defined at the level of the handwriting sign, rather than the individual feature, whereas objective discrimination power is feature-specific. The same relative discrimination value is therefore assigned to all features belonging to the same handwriting sign.

Relative discrimination power is calculated as:

$$G = U \cdot M \quad (7)$$

where U is the discrimination component, and M is the consistency component.

The discrimination component U describes how frequently a handwriting sign or letter differs in comparisons between writings produced by different persons:

$$U = \frac{n_u}{N_{dw}} \quad (8)$$

where n_u denotes the number of non-matching comparisons showing a difference, and N_{dw} is the total number of evaluable non-matching comparisons.

A non-matching comparison is a comparison classified in the forensic expert report as involving different writers. The value of U ranges from 0 to 1. A value close to 0 indicates limited capacity to differentiate between writers, whereas a value close to 1 indicates that the characteristic differs in most different-writer comparisons.

Only comparisons in which the characteristic could be assessed reliably in both samples were included in the denominator. Comparisons involving insufficient material, poor reproduction quality, illegibility, or absence of the required letter were treated as non-evaluable and excluded rather than classified as agreements.

The consistency component M describes how frequently the analytical unit remains in agreement across handwriting samples produced by the same person:

$$M = \frac{n_m}{N_{sw}} \quad (9)$$

where n_m is the number of matching comparisons showing agreement, and N_{sw} is the total number of matching comparisons. A matching comparison is a comparison classified in the forensic expert report as involving the same writer.

A value of M close to 1 indicates that the characteristic is reproduced with a high degree of consistency within the writer's natural range of variation. A lower value indicates greater intra-writer variability and, therefore, a lower potential weight for writer identification.

Consistency does not imply exact repetition. Natural handwriting variation means that successive executions of the same feature or letter are never completely identical. Agreement must therefore be defined as correspondence within an operationally or empirically established tolerance.

For individual letters, a different empirical estimator of relative discrimination power is used. G is calculated as the proportion of all evaluable forensic cases in the database in which the execution of a given letter differed between the questioned and known writings:

$$G = \frac{n_l}{N_{dw} + N_{sw}}, \quad (10)$$

where n_l is the number of cases where the letter execution was different.

3.4. Case-Specific Presence and Representation

A handwriting feature or individual letter can contribute to writer identification only when it is adequately represented and can be assessed reliably in both the questioned and known writings. This requirement is expressed by the case-specific presence component C .

The term presence is used here in a broad methodological sense. It does not mean that the same manifestation of the feature must occur in both writings. Rather, it indicates that the relevant handwriting sign or letter is available for comparison and can be evaluated with sufficient reliability on both sides. For example, slant may be fully evaluable even when the questioned writing shows a rightward slant and the known writing a vertical slant. The difference between these manifestations constitutes the comparison result, whereas C indicates the extent to which the sign can be evaluated.

For handwriting features, case-specific presence is calculated as:

$$C = 1 - \frac{|q_q - q_n|}{\max(q_q, q_n)} \quad (11)$$

$$q_q = \frac{k_q}{N_q}; \quad q_n = \frac{k_n}{N_n}$$

where k_q and k_n denote the numbers of observed occurrences of the feature in the questioned and known writings, while N_q and N_n represent the corresponding numbers of opportunities for evaluation. This formulation assigns higher values when the feature occurs with similar relative frequency in both questioned and known writings.

For many general features, such as size, width, spacing, and baseline direction, q_q and q_n are effectively equal to 1; therefore, $C = 1$. The value of C becomes more relevant for features that may occur only in particular graphical elements, such as upper- and lower-zone formations or specific zone relationships.

$$C = \sqrt{k_q \cdot k_n} \quad (12)$$

In practice, Formula (12) is relevant when the number of available letter occurrences is limited. A provisional threshold of five occurrences was adopted as a practical criterion for sufficient representation. Accordingly, k_q and k_n are transformed as follows:

$$K_q = \min\left(\frac{k_q}{5}, 1\right) \text{ and } K_n = \min\left(\frac{k_n}{5}, 1\right) \quad (13)$$

and

$$C = \sqrt{K_q \cdot K_n} \quad (14)$$

Thus, the maximum value of C is 1. The use of a capped value prevents highly frequent letters from receiving an excessive advantage once a sufficient number of occurrences is available for reliable comparison.

The inclusion of case-specific presence distinguishes the proposed framework from purely population-based weighting schemes. Identification value is therefore determined not only by the general properties of handwriting features and letters but also by their actual availability within the specific forensic examination.

3.5. Weights Based on Identification Values

Identification value measures the relative evidential importance of a handwriting feature or an individual letter. To incorporate this information into the overall handwriting comparison, the raw identification values are transformed into normalized weights.

The weight assigned to an analytical unit determines how strongly its observed similarity or difference contributes to the combined comparison result. A handwriting feature with a comparatively high identification value, therefore, receives a greater share of the total weight than a common, weakly discriminative, unstable, or insufficiently represented feature.

The weight of the i -th feature in Equation (2) is calculated as:

$$w_i = \frac{Z_i}{\sum Z_\xi} \quad (15)$$

The denominator contains the sum of the identification values of all evaluated features, ensuring that the weights sum to 1.

The weights of individual letters, v_j , in Equation (3) are calculated in the same way.

4. Results

Objective discrimination power was calculated using data from 550 handwriting samples in the HSDetect database. The values presented in Table 1 are intended as illustrative estimates for selected basic handwriting features and are used to demonstrate the proposed calculation procedure. Given the current sample size and composition of the database, these values should not be interpreted as definitive population parameters.

Table 1. Objective discrimination power of handwriting features.

Handwriting Sign	Handwriting Feature	k	E	CI	D
Letter size	very small	11	0.02	0.01–0.04	1.70
	small	77	0.14	0.11–0.17	0.85
	medium	198	0.36	0.32–0.4	0.44
	large	121	0.22	0.19–0.26	0.66
	very large	22	0.04	0.03–0.06	1.40
Letter zone proportions	all zones are proportional	176	0.32	0.28–0.36	0.49
	larger middle zone	72	0.13	0.1–0.16	0.89
	smaller middle zone	94	0.17	0.14–0.2	0.77
	larger lower zone	77	0.14	0.11–0.17	0.85
	smaller lower zone	28	0.05	0.03–0.07	1.30
	large upper zone	33	0.06	0.04–0.08	1.22
	smaller upper zone	72	0.13	0.1–0.16	0.89
Letter width	narrow	154	0.28	0.24–0.32	0.55
	medium	204	0.37	0.33–0.41	0.43
	wide	193	0.35	0.31–0.39	0.46
Letter spacing	normal	231	0.42	0.38–0.46	0.38
	large	149	0.27	0.23–0.31	0.57
	small	149	0.27	0.23–0.31	0.57
	absent	22	0.04	0.03–0.06	1.40
Word spacing	narrow	83	0.15	0.12–0.18	0.82
	medium	209	0.38	0.34–0.42	0.42

	wide	253	0.46	0.42–0.50	0.34
	no intervals (connected words)	6	0.01	0.00–0.02	2.00
Line spacing	narrow	160	0.29	0.25–0.33	0.54
	normal	259	0.47	0.43–0.51	0.33
Slant	wide	132	0.24	0.21–0.28	0.62
	slight right	39	0.07	0.05–0.09	1.15
	normal right	226	0.41	0.37–0.45	0.39
	strong right	94	0.17	0.14–0.20	0.77
	upright, vertical	105	0.19	0.16–0.22	0.72
	strong left	11	0.02	0.01–0.04	1.70
	normal left	33	0.06	0.04–0.08	1.22
	slight left	44	0.08	0.06–0.11	1.10
	horizontal	237	0.43	0.39–0.47	0.37
Baseline direction	ascending, rising	193	0.35	0.31–0.39	0.46
	descending, falling	77	0.14	0.11–0.17	0.85
Left margin width	irregular or wavy	44	0.08	0.06–0.11	1.10
	normal, medium	143	0.26	0.23–0.30	0.59
	wide	110	0.20	0.17–0.24	0.70
	narrow	248	0.45	0.41–0.49	0.35
Left margin form	absent	50	0.09	0.07–0.12	1.05
	straight	242	0.44	0.40–0.48	0.36
	narrowing	61	0.11	0.09–0.14	0.96
	widening	116	0.21	0.18–0.25	0.68
	convex	6	0.01	0.00–0.02	2.00
	concave	11	0.02	0.01–0.04	1.70
	uneven	110	0.20	0.17–0.24	0.70
	stair-like	6	0.01	0.00–0.02	2.00
Connectiveness	connected	149	0.27	0.23–0.31	0.57
	partly connected	264	0.48	0.44–0.52	0.32
	disconnected	138	0.25	0.22–0.29	0.60
Pressure strength	average	292	0.53	0.49–0.57	0.28
	strong	143	0.26	0.23–0.30	0.59
	weak	116	0.21	0.18–0.25	0.68
Stroke tension	low, looseness	116	0.21	0.18–0.25	0.68
	normal, elastic stroke	292	0.53	0.49–0.57	0.28
	high, rigidity	143	0.26	0.23–0.30	0.59
Diacritic pressure	normal	121	0.22	0.19–0.26	0.66
	stronger	28	0.05	0.03–0.07	1.30
	weaker	165	0.30	0.26–0.34	0.52

In this dataset, objective discrimination power ranged from 0.28 for normal stroke tension, observed in 53% of writers, to 2.00 for features observed in only 1% of writers. The mean value was 0.83. Prevalence ratios may sum to less than 1.00 because non-evaluable samples were not assigned to any of the listed feature categories. The prevalence of graded feature categories does not necessarily follow a monotonic pattern, as the observed frequencies depend on the empirical distribution of handwriting and the operational category boundaries.

Table 1 presents only a subset of the handwriting features used in forensic examination. Additional characteristics include form-related signs such as dominant letter form, connection form, baseline form, and upper- and lower-zone formations; dynamic characteristics such as writing tempo, pressure rhythm, and stroke quality; and features describing the regularity and stability of different aspects of handwriting.

When insufficient data were available to calculate reliable values of E and D for a particular feature, the mean value of the evaluated features was used provisionally.

Relative discrimination power was calculated from the HSForen data. The volume of data in HSForen is currently smaller and less representative than that of HSDetect, and the database consists mainly of German-language cases. In Table 2, the number of non-matching cases $N_{nm} = 36$; the number of matching cases $N_{mt} = 56$. The database is being continuously expanded, and the retrospective transfer of additional forensic cases from traditional, non-formalized examination records into the standardized HSForen structure is ongoing.

In the present study, the classification of comparisons as matching or non-matching was based on the conclusion recorded in the corresponding forensic expert report rather than on independently verified ground truth. Accordingly, N_{mt} denotes the number of evaluable cases concluded by the examiner to involve the same writer, whereas N_{nm} denotes the number of evaluable cases concluded to involve different writers. These categories, therefore, represent the outcomes of forensic examinations stored in HSForen and should not be interpreted as externally confirmed authorship status.

For practical calculations, Jeffreys smoothing was applied to the discrimination and consistency components to avoid zero estimates:

$$U = \frac{n_u + 0.5}{N_{nm} + 1}; M = \frac{n_m + 0.5}{N_{mt} + 1} \quad (16)$$

Table 2. Illustrative values of relative discrimination.

Handwriting Sign	n_u	U	$U\ CI$	n_m	M	$M\ CI$	G
Letter size	9	0.26	0.14–0.42	48	0.85	0.73–0.92	0.22
Letter zone proportions	9	0.26	0.14–0.42	52	0.92	0.82–0.97	0.24
Letter width	5	0.15	0.07–0.30	56	0.99	0.94–1.00	0.15
Letter spacing	12	0.34	0.21–0.50	48	0.85	0.73–0.92	0.29
Word spacing	5	0.15	0.07–0.30	56	0.99	0.94–1.00	0.15
Line spacing	5	0.15	0.07–0.30	52	0.92	0.82–0.97	0.14
Slant	9	0.26	0.14–0.42	56	0.99	0.94–1.00	0.26
Baseline direction	5	0.15	0.07–0.30	39	0.69	0.56–0.80	0.10
Left margin width	9	0.26	0.14–0.42	52	0.92	0.82–0.97	0.24
Left margin form	1	0.04	0.01–0.16	56	0.99	0.94–1.00	0.04
Connectiveness	12	0.34	0.21–0.50	52	0.92	0.82–0.97	0.31
Pressure strength	1	0.04	0.01–0.16	48	0.85	0.73–0.92	0.04
Stroke tension	20	0.55	0.39–0.70	30	0.54	0.41–0.66	0.30
Diacritic pressure	9	0.26	0.14–0.42	56	0.99	0.94–1.00	0.26

Based on the illustrative values presented in Table 2, preliminary case-independent identification values Z_b were calculated to demonstrate how objective and relative discrimination components interact within the proposed framework (Table 3).

Table 3. Illustrative case-independent identification values.

Handwriting Sign	Handwriting Feature	D	G	Z_b
Letter size	very small	1.70	0.21	0.36
	small	0.85	0.21	0.18
	medium	0.44	0.21	0.09
	large	0.66	0.21	0.14
	very large	1.40	0.21	0.29
Letter zone proportions	all zones are proportional	0.49	0.23	0.11

	larger middle zone	0.89	0.23	0.20
	smaller middle zone	0.77	0.23	0.18
	larger lower zone	0.85	0.23	0.20
	smaller lower zone	1.30	0.23	0.30
	large upper zone	1.22	0.23	0.28
	smaller upper zone	0.89	0.23	0.20
Letter width	narrow	0.55	0.15	0.08
	medium	0.43	0.15	0.06
	wide	0.46	0.15	0.07
Letter spacing	normal	0.38	0.30	0.11
	large	0.57	0.30	0.17
	small	0.57	0.30	0.17
	absent	1.40	0.30	0.42
Word spacing	narrow	0.82	0.15	0.12
	medium	0.42	0.15	0.06
	wide	0.34	0.15	0.05
	no intervals (connected words)	2.00	0.15	0.30
Line spacing	narrow	0.54	0.14	0.08
	normal	0.33	0.14	0.05
	wide	0.62	0.14	0.09
Slant	slight right	1.15	0.25	0.29
	normal right	0.39	0.25	0.10
	strong right	0.77	0.25	0.19
	upright, vertical	0.72	0.25	0.18
	strong left	1.70	0.25	0.43
	normal left	1.22	0.25	0.31
	slight left	1.10	0.25	0.28
Baseline direction	horizontal	0.37	0.10	0.04
	ascending, rising	0.46	0.10	0.05
	descending, falling	0.85	0.10	0.09
	irregular or wavy	1.10	0.10	0.11
Left margin width	normal, medium	0.59	0.23	0.14
	wide	0.70	0.23	0.16
	narrow	0.35	0.23	0.08
	absent	1.05	0.23	0.24
Left margin form	straight	0.36	0.05	0.02
	narrowing	0.96	0.05	0.05
	widening	0.68	0.05	0.03
	convex	2.00	0.05	0.10
	concave	1.70	0.05	0.09
	uneven	0.70	0.05	0.04
	stair-like	2.00	0.05	0.10
Connectiveness	connected	0.57	0.32	0.18
	partly connected	0.32	0.32	0.10
	disconnected	0.60	0.32	0.19
Pressure strength	average	0.28	0.04	0.01
	strong	0.59	0.04	0.02
	weak	0.68	0.04	0.03
Stroke tension	low, looseness	0.68	0.30	0.20
	normal, elastic stroke	0.28	0.30	0.08
	high, rigidity	0.59	0.30	0.18
Diacritic pressure	normal	0.66	0.25	0.17

stronger	1.30	0.25	0.33
weaker	0.52	0.25	0.13

Within the illustrative calculations presented in Table 3, comparatively low Z_b values were obtained for normal line spacing, horizontal baseline direction, left-margin form, and pressure strength. These examples demonstrate how common feature manifestations or handwriting signs with limited relative discrimination receive lower weights within the proposed model.

Conversely, the illustrative calculations produced comparatively high Z_b values for very small handwriting, absence of interletter spacing, strong leftward slant, and increased pressure in diacritical marks. These examples demonstrate how low prevalence combined with higher relative discrimination can increase the resulting identification value. The numerical values should not, however, be interpreted as established rankings of forensic importance.

HSForen also contains results from the congruence analysis of individual letters. Relative discrimination power for German-language letter data is presented in Table 4.

Table 4. Letter relative discriminative power for the German language and case-specific evaluability.

Lowercase Letter	G	K_q	K_n	C	Z	Uppercase Letter	G	K_q	K_n	C	Z
a	0.21	1	1	1	0.21	A	0.07	0.40	1	0.63	0.04
b	0.11	1	1	1	0.11	B	0.07	0.60	1	0.77	0.05
c	0.04	1	1	1	0.04	C	0.04	0	1	0	0
d	0.11	1	1	1	0.11	D	0.18	0.20	1	0.44	0.08
e	0.11	1	1	1	0.11	E	0.04	0	1	0	0
f	0.14	1	1	1	0.14	F	0.07	0.60	1	0.77	0.05
g	0.25	1	1	1	0.25	G	0.07	0.20	1	0.44	0.03
h	0.14	1	1	1	0.14	H	0.18	0	1	0	0
i	0.18	1	1	1	0.18	I	0.11	0.20	1	0.44	0.05
j	0.00	0	1	0	0	J	0.11	0	1	0	0
k	0.11	0.4	1	0.63	0.07	K	0.21	0.20	1	0.44	0.09
l	0.04	1	1	1	0.04	L	0.04	0	1	0	0
m	0.18	1	1	1	0.18	M	0.21	0.20	1	0.44	0.09
n	0.14	1	1	1	0.14	N	0.04	0	1	0	0
o	0.18	1	1	1	0.18	O	0.04	0	1	0	0
p	0.07	1	1	1	0.07	P	0.11	0.20	1	0.44	0.05
q	0.00	0	1	0	0.00	Q	0.00	0	1	0	0
r	0.14	1	1	1	0.14	R	0.07	0.20	1	0.44	0.03
s	0.14	1	1	1	0.14	S	0.11	0.60	1	0.77	0.08
t	0.21	1	1	1	0.21	T	0.07	0.20	1	0.44	0.03
u	0.11	1	1	1	0.11	U	0.07	0	0.40	0	0
v	0.07	0.4	1	0.63	0.04	V	0.14	0	1	0	0
w	0.11	0.4	1	0.63	0.07	W	0.04	0.40	1	0.63	0.03
x	0.00	0	1	0	0	X	0.00	0	0.20	0	0
y	0.00	0	1	0	0	Y	0.00	0	0.20	0	0
z	0.11	0.8	1	0.89	0.10	Z	0.07	0.20	0.60	0.35	0.02
ä	0.04	0.4	1	0.63	0.03	Ä	0.00	0	1	0	0
ö	0.14	0.2	1	0.45	0.06	Ö	0.00	0	1	0	0
ü	0.04	0.4	1	0.63	0.03	Ü	0.00	0	1	0	0
ß	0.11	0	0	0	0	ß	0.00	0	0	0	0

Table 4 combines two levels of information for illustrative purposes. The relative discrimination value G is derived from the available forensic database and is independent of the particular case. In contrast, K_q ,

K_n , and C are determined from the handwriting material examined in the specific case. The resulting Z value is therefore case-specific and should not be interpreted as a general identification value of the letter.

Zero values of G indicate that HSForen currently contains no cases in which a difference in the execution of the corresponding letter was recorded between questioned and known writings. These values should therefore not be interpreted as evidence that the letters have no discriminative potential. In the current illustrative dataset, some letters received higher relative discrimination values than others. These differences are presented only to demonstrate the letter-level calculation and should not be interpreted as a validated ranking of the forensic discriminative value of individual letters.

To illustrate the calculation of case-specific evaluability for individual letters, a short questioned-writing sample is used in Figure 2. The sample is not part of a real forensic case and is included solely for demonstration purposes.

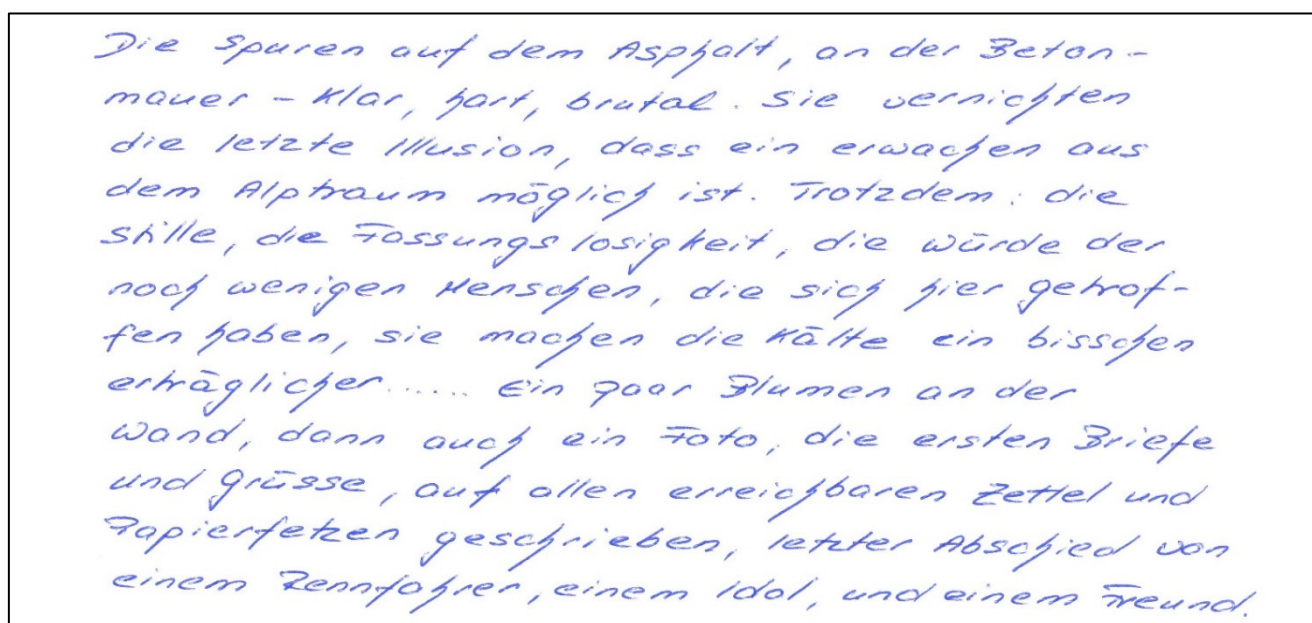


Figure 2. Questioned document.

The sample contains sufficient handwriting material to evaluate nearly all general handwriting features, with the possible exception of the upper and lower margins, which are frequently non-evaluable in practical casework. Other potentially problematic characteristics, such as the location and form of diacritical marks, punctuation, and upper- and lower-zone formations, are represented sufficiently. In this framework, i-dots and t-crosses are also classified as diacritical marks.

The relevant characteristics are also sufficiently stable to permit clear classification. For example, the diacritical marks are placed consistently in the horizontal and vertical directions and show weaker pressure, while the lower-zone formation of the letter g is open.

In this example, case-specific evaluability therefore does not reduce the identification value of the general handwriting features; in other words, $C = 1$.

For individual letters, however, case-specific evaluability is more relevant. For demonstration purposes, sufficient known handwriting material is assumed to be available for all letters except β and the uppercase letters U , Z , X , and Y , which is a realistic situation in practical examination.

Most lowercase letters occur at least five times in the questioned sample. Exceptions include v and w , with two occurrences each, and j , q , x , y , and β , which do not occur. Uppercase letters, when present, generally occur fewer than five times and therefore receive lower case-specific evaluability values.

In this illustrative example, a discrepancy in lowercase g ($Z = 0.25$) receives approximately eight times the weight of a discrepancy in ä or ü ($Z = 0.03$). This ratio is specific to the current illustrative values and should not be interpreted as a general forensic weighting rule.

5. Discussion

The present study proposes a formalized approach to the identification value of handwriting features and individual letters. This approach is consistent with the broader shift in forensic science toward more transparent, data-driven, and empirically testable evaluation of evidence, in which expert judgement is increasingly supported by quantitative models and reference data [20]. Its central premise is that handwriting findings should not be treated uniformly. A similarity or difference involving a common, unstable, or condition-dependent characteristic cannot have the same evidential importance as a finding involving a rare, consistent, and highly discriminative feature. The proposed model addresses this problem by combining population distinctiveness, empirical discriminative performance, and case-specific evaluability, thereby providing a structured basis for weighting individual findings within the overall handwriting examination.

The illustrative calculations demonstrate how the proposed framework differentiates handwriting findings based on their prevalence, relative discrimination, and case-specific representation. In the illustrative calculations, comparatively low case-independent identification values were obtained for normal line spacing, a horizontal baseline, left-margin form, and pressure strength. Normal line spacing and a horizontal baseline are common manifestations and therefore provide little population distinctiveness. Left-margin form was also weakly informative, partly because it cannot always be assessed reliably in contemporary handwriting material. Many samples are written on small sheets or non-standard pages that do not permit meaningful evaluation of page layout. Pressure strength showed particularly low relative discrimination. This is plausible because general pressure is influenced not only by the writer's habitual motor behavior but also by the pen, paper, supporting surface, writing position, and quality of reproduction.

The illustrative calculations yielded comparatively higher identification values for very small handwriting, the absence of interletter spacing, a strong leftward slant, and increased pressure on diacritical marks. These manifestations are comparatively uncommon and therefore receive higher objective discrimination values. At the same time, they represent relatively specific aspects of spatial or motor organization. Absence of interletter spacing, in which adjacent letter-body strokes touch without a genuine connecting stroke, reflects a distinctive organization of successive movements. A strong leftward slant is also unusual when consistently maintained. Increased pressure in diacritical strokes may be especially informative because it represents a localized and habitual movement emphasis rather than a general pressure level that is strongly affected by external conditions.

The structure of the proposed model also illustrates why rarity alone is insufficient to determine identification value. A rare feature may still have limited identification value if the corresponding handwriting sign shows low consistency within writers or limited discrimination between writers. Conversely, a moderately common feature may remain useful when it is stable and frequently differs in non-matching comparisons. The multiplication of objective and relative discrimination power, therefore, prevents an uncommon but unreliable characteristic from receiving an unjustifiably high weight.

For handwriting features, objective discrimination power can be estimated from the prevalence of a defined feature manifestation in the reference population. For individual letters, the situation is more complex because the alphabetic character itself is not writer-specific. Nevertheless, the expected linguistic frequency of a letter may be used as an additional measure of its objective availability or potential identification value. A letter that occurs frequently in the language is more likely to be represented repeatedly in a text, providing a better basis for assessing habitual execution and natural variation. Conversely, a rare letter may be absent or represented only once, even in a relatively long text.

Expected linguistic frequency should, however, be interpreted cautiously. It reflects how likely a letter is to occur in the language, not how distinctive its execution is among writers. A frequent letter is not necessarily weakly discriminative, and a rare letter is not automatically highly informative. Linguistic frequency is therefore most appropriately considered an availability-related or normalizing factor, whereas the actual discriminative performance of the letter should be derived from forensic comparison data.

The illustrative letter-level calculations also show how different graphical elements can receive different relative discrimination values within the proposed framework. These values should not be interpreted as a stable ranking of individual letters, because the current HSForen dataset is limited and the underlying classifications are based on forensic examination conclusions rather than independently established ground truth. The purpose of the present calculations is therefore to demonstrate the operation of the weighting procedure rather than to establish which letters are intrinsically more discriminative.

Case-specific evaluability is particularly important at the letter level. A letter may have high general discriminative potential but contribute little to a particular examination when it is absent or represented only once. The proposed capped occurrence model reduces its identification value when fewer than five evaluable examples are available. This reflects the need for repeated occurrences to assess whether a particular execution is habitual and to observe natural variation. The use of the geometric mean also ensures that abundant known material cannot fully compensate for insufficient questioned material.

A further important distinction is between the similarity grade of a feature or letter and its identification value. The two quantities describe different aspects of the examination. Similarity grade expresses the degree of correspondence between questioned and known writings for a particular analytical unit. Identification value, in contrast, describes how important that similarity or difference is for writer identification.

Thus, a feature may show a very high similarity grade but have a low identification value because it is common or weakly discriminative. Agreement in normal line spacing, for example, may be nearly complete but provide limited support because the feature is widespread. Conversely, a highly discriminative feature may show only moderate similarity, yet the observed discrepancy may be important. The same principle applies to individual letters. Similarity determines the direction and degree of the comparison result, while identification value determines how strongly that result contributes to the overall score.

This distinction is central to the weighting model. The identification value is transformed into a normalized weight, whereas the observed feature or letter similarity is entered separately into the calculation. A high weight does not automatically support common authorship; it increases the influence of whichever result is observed, whether agreement or difference. The model therefore separates the importance of a finding from its direction.

The proposed framework may also be applicable to handwritten signatures, although signatures require separate methodological treatment. Compared with ordinary handwriting, signatures are usually more highly automatized, may be strongly stylized, and often contain fewer directly comparable graphical elements. Their identification value may therefore depend on a different set of structural and dynamic characteristics and on signature-specific reference data. In addition, recent evidence indicates that neurological and psychiatric conditions can influence both static and dynamic signature features, although no disorder-specific signature pattern has yet been established [21]. A formalized signature model would therefore require dedicated operational definitions, population data, and independent validation.

The present study should be interpreted primarily as a methodological proposal rather than as a validation of the resulting numerical weights. The available database material is used to demonstrate how the proposed components can be defined, calculated, and combined within a formalized handwriting examination. The resulting values are provisional and dataset-dependent. Establishing stable forensic reference values will require larger and more representative datasets, independently established same-writer and different-writer ground truth, systematic assessment of coding reliability, analysis of dependencies

among handwriting characteristics, sensitivity analysis of the model parameters, and external validation of weighted versus unweighted comparison scores.

The study has several limitations. First, estimates of population frequency and relative discrimination power depend on the current composition of HSDetect and HSForen. Although the databases include several hundred samples, they contain heterogeneous material and do not necessarily represent all relevant writing populations equally. The estimates should therefore be regarded as preliminary and may change as the databases grow.

Second, the present model does not yet provide empirically validated decision thresholds for the final similarity score. The coefficients assigned to handwriting-feature similarity and letter congruence were derived from practical experience and require further validation. This need for validation is particularly important in view of empirical studies of accuracy and error in forensic handwriting examination. Large-scale testing has demonstrated substantial accuracy and reliability among trained forensic handwriting examiners, while also showing that erroneous and inconclusive conclusions occur and that performance varies across comparisons and examiners [22]. A comparative review of published studies similarly found considerably lower error rates among experts than among non-experts, but also substantial variation across studies and examination conditions [23]. Statistical analyses of forensic signature examination further emphasize the need for caution when translating examiner conclusions into quantitative statements of evidential strength [24].

Third, the present framework was developed and illustrated using alphabetic handwriting, and the operational definitions and reference values presented here should therefore be interpreted within that context. Extension to non-alphabetic writing systems would require substantial adaptation of the analytical units. In logographic or syllabic systems, for example, the relevant units may be characters, radicals, stroke sequences, spatial configurations, or other script-specific structures rather than letters in the alphabetic sense. The general principles of separating similarity from identification value and combining population, discriminative, and case-specific information could nevertheless remain applicable. Any such extension would require script-specific operational definitions, reference data, and independent validation.

A further limitation is that statistical dependencies among handwriting characteristics have not yet been systematically quantified. Correlated characteristics may partially represent the same underlying handwriting property and could therefore receive disproportionate combined weight in an additive model. This issue should be addressed in future validation through correlation and redundancy analyses and, where necessary, consolidation or adjustment of related analytical units.

The proposed framework nevertheless provides a more transparent basis for weighting handwriting evidence. It does not replace expert interpretation, since the examiner must still assess comparability, natural variation, technical limitations, disguise, imitation, and the coherence of the findings as a whole. Its advantage lies in making the basis of weighting explicit and empirically adjustable [8]. As HSDetect and HSForen expand, estimates of prevalence, discrimination, consistency, and letter-specific performance can be updated without changing the general structure of the model.

The general methodological principles of the framework may also be adaptable to clinical assessment of handwriting disorders. In that context, the objective would not be writer identification but structured quantification of handwriting abnormalities associated with neurological, motor, cognitive, or psychiatric conditions. Operationally defined features, standardized coding, and empirically derived weighting could potentially improve reproducibility and allow changes to be summarized in condition-specific indices. Such applications would, however, require dedicated feature sets, appropriate control data, and independent clinical validation.

Overall, the proposed framework formalizes the principle that handwriting findings need not contribute uniformly to forensic comparison. It provides a structured method for separating similarity from identification value and assigning different weights based on population distinctiveness, empirical

comparison behavior, and case-specific representation. The present numerical examples illustrate this principle but do not establish validated forensic weights.

6. Conclusions

1. The present study proposes a methodological framework for formalizing the identification value of handwriting features and individual graphical elements in forensic handwriting examination. Its central principle is that the importance of a finding should be considered separately from the degree of similarity or difference observed in the comparison.
2. For handwriting features, identification value is modeled through objective discrimination power, relative discrimination power, and case-specific presence and representation. For individual graphical elements, relative discrimination and case-specific availability provide the principal components of the current formulation.
3. The framework allows identification values to be transformed into explicit weights within a formalized handwriting comparison. Similarity grade determines the comparison result, whereas identification value determines how strongly that result contributes to the overall score.
4. The numerical values presented in this study are illustrative and preliminary. They demonstrate how the proposed framework can be implemented, but should not be interpreted as validated population parameters, definitive rankings of handwriting characteristics or letters, or established forensic decision weights.
5. Further validation requires larger and more representative datasets, independently established same-writer and different-writer ground truth, systematic inter-examiner reliability assessment, investigation of dependencies among handwriting characteristics, sensitivity analysis of model parameters, and comparison of weighted and unweighted examination performance. Subject to such validation, the proposed framework may provide a transparent and empirically adjustable basis for weighting handwriting findings while retaining the expert examiner's role in evaluating natural variation, comparability, technical limitations, disguise, imitation, and the overall coherence of the evidence.

Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript the author used Deeple and ChatGPT for language editing and improvement. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Ethics Statement

The study involved the analysis of anonymized handwriting data. All handwriting material was handled in accordance with the applicable consent and data-protection requirements. No additional ethical approval was required for the present retrospective analysis of anonymized data.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

This research received no external funding.

Declaration of Competing Interest

The author declares no competing interests.

References

- Osborn AS. *Questioned Documents*, 2nd ed.; Nelson-Hall: Chicago, IL, USA, 1929.
- Hilton O. *Scientific Examination of Questioned Documents, Revised*; Elsevier: New York, NY, USA, 1982.
- Michel L. *Gerichtliche Schriftvergleichung. Eine Einführung in Grundlagen, Methoden und Praxis*; Walter de Gruyter & Co.: Berlin, Germany, 1982.
- Huber RA, Headrick AM. *Handwriting Identification: Facts and Fundamentals*; CRC Press: Boca Raton, FL, USA, 1999.
- Ellen D, Day S, Davies C. *Scientific Examination of Documents. Methods and Techniques*, 4th ed.; CRC Press: Boca Raton, FL, USA, 2018.
- Harralson HH, Miller LS. *Huber and Headrick's Handwriting Identification. Facts and Fundamentals*; CRC Press: Boca Raton, FL, USA, 2018.
- Morris R. *Forensic Handwriting Identification. Fundamental Concepts and Principles*, 2nd ed.; Academic Press: London, UK, 2021.
- 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
- ENFSI. *Best Practice Manual for the Forensic Handwriting Examination*, 4th ed.; ENFSI: Wiesbaden, Germany, 2022.
- Chernov Y, Caspers C. A Structured Framework for Formalized and Quantitative Handwriting Examination. *Perspect. Leg. Forensic Sci.* **2025**, *2*, 10007. DOI:10.70322/plfs.2025.10007
- Jones K, Bird C. Complexity, Features, and Comparisons in Forensic Handwriting Examination. *WIREs Forensic Sci.* **2025**, *7*, e1537. DOI:10.1002/wfs2.1537
- Angel M. Handwriting Complexity. In *Forensic Document Examination in the 21st Century*; Kelly JS, Angel M, Eds.; CRC Press: Boca Raton, FL, USA, 2021; pp. 111–118.
- Marti U-V, Bunke H. The IAM-database: An English sentence database for offline handwriting recognition. *Int. J. Doc. Anal. Recognit.* **2002**, *5*, 39–46. DOI:10.1007/s100320200071
- Kleber F, Fiel S, Diem M, Sablatnig R. CVL-Database: An off-line database for writer retrieval, writer identification and word spotting. In *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR)*, Washington, DC, USA, 25–28 August 2013; pp. 560–564.
- Crawford A, Ray A, Carriquiry A. A database of handwriting samples for applications in forensic statistics. *Data Brief* **2020**, *28*, 105059. DOI:10.1016/j.dib.2019.105059
- 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
- Chernov Y, Caspers C. Formalized Computer-Aided Handwriting Psychology: Validation and Integration into Psychological Assessment. *Behav. Sci.* **2020**, *10*, 27. DOI:10.3390/bs10010027
- Vastring TW. Frequency of Occurrence Proportions in Handwriting. In *Forensic Document Examination in the 21st Century*; Kelly JS, Angel M, Eds.; CRC Press: Boca Raton, FL, USA, 2021.
- Johnson M, Vastring T, Boulanger M, Schuetzner E. Measuring the frequency occurrence of handwriting and handprinting characteristics. *J. Forensic Sci.* **2017**, *62*, 142–163. DOI:10.1111/1556-4029.13248
- Morrison GS. Advancing a paradigm shift in evaluation of forensic evidence: The rise of forensic data science. *Forensic Sci. Int. Synerg.* **2022**, *5*, 100270. DOI:10.1016/j.fsisyn.2022.100270
- Vitale S, Guidali G, Zago S, Bolognini N. Handwritten signatures in neurological and psychiatric disorders: A systematic review with clinical and forensic implications. *Neurol. Sci.* **2026**, *47*, 658. DOI:10.1007/s10072-026-09266-z
- 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
- Crot S, Marquis RA. Comparative Review of Error Rates in Forensic Handwriting Examination. *J. Forensic Sci.* **2024**, *69*, 2127–2138. DOI:10.1111/1556-4029.15589
- Jemia AB, Plant A, Foster S. Examining Errors in Forensic Signature Analysis: A Statistical Perspective. *Int. J. Public Policy Law Dev.* **2024**, *1*, 42–64. Available online: <https://ijpld.com/ijpld/article/view/14> (accessed on 17 September 2026).