

## CLASSIFYING HANDWRITING SAMPLES ACCORDING TO THEIR TYPE USING DISCRIMINANT ANALYSIS

Jagoda DZIDA 

*Laboratory of Criminalistics, Department of Law and Administration, Adam Mickiewicz University, Poznań, Poland*

### Abstract

Classifying handwriting samples according to their type (i.e. natural, disguised, traced, simulated or unintentionally unnatural) is an important task in handwriting analysis. It may facilitate the collection of writing standards and also help experts to assess the differences between questioned material and comparative samples or to choose the best writing features and the most relevant examination protocol for the case. Current research aimed to create a method for classifying the type of a handwriting sample using discriminant analysis. Five basic types (i.e. natural, disguised, traced, simulated and unintentionally unnatural) and some subtypes were included in this study. Participants (N = 139) wrote their full signatures, fictional signatures or a short sentence. Motor and dimensional features were assessed. The methods proved to be more than twice as accurate in classifying samples according to their type than a random choice probability (e.g. 44% as opposed to 17% for the 6-types classifier). This proof-of-a-concept study demonstrates that handwriting samples may be classified according to their type with satisfying accuracy based on their writing features and statistical tools of discriminant analysis. However, further studies are necessary to improve the accuracy of the method.

### Keywords

Analysis of handwriting; Handwriting samples; Classifying handwriting samples; Types of handwriting samples; Discriminant analysis.

*Received 27 December 2022; accepted 31 January 2023*

### 1. Introduction

Analysis of handwriting in criminalistics is often aimed at determining the authorship of the handwriting sample. Therefore, a handwriting expert needs to compare the questioned material with comparative samples. Comparative material should consist of the same type of handwriting samples as the questioned material. By the type we mean not only the type of lettering (e.g. cursive writing, handprinting, etc.) but also if the sample is a natural one or unnatural one<sup>1</sup>. In the research, the second meaning is used.

If the type of the material in question is known, the collection of samples would be much easier.

under unaltered conditions (in usual position, with usual materials, without guidance, stress or any sort of violence) and includes natural habits of the writer. The unnatural samples can be further divided into unintentionally unnatural samples (like samples written in an unusual position or using an unusual writing device) and intentionally unnatural samples. The second group of samples consists of disguised samples (samples written in a manner that is to disguise the identity of the writer, by making some alterations in his writing habits), traced samples (by following the guiding line of the original sample, e.g. in transmitted light) and simulated samples (being imitations of some other handwriting, but without guiding line). All of the following samples are described in detail later in the article.

<sup>1</sup> In general, we can divide all handwriting samples into natural and unnatural ones. The natural sample is the sample written

Moreover, knowing the type of the questioned material is important to experts, as it allows them to properly assess the observed differences and similarities between the questioned and the comparative material (e.g., if the material in question consists of traced writing, we expect some tremor to be visible in it). Having determined the type of handwriting sample or, in other words, having classified a handwriting sample according to its type, experts would know which handwriting features they should focus on, and which ones may be less important for the case.

A few classifications of types of handwriting samples have been described in the literature – for example, by Huber and Headrick (1999), who described natural handwriting, disguised handwriting and forged handwriting (traced or simulated); similarly, these types were described by Koppenhaver (2007), who paid more attention to differences between some specific types of samples (like, for example, disguised samples and simulated ones). Somewhat different terminology was proposed by Matuszewski (2011). He proposed the following classification of the handwriting samples:

- natural samples,
- unnatural samples:
  - intentionally unnatural samples:
    - disguised samples,
    - simulated samples:
      - ‘learn and simulate’ samples (simulation by memory),
      - ‘look and simulate’ samples (freehand simulation),
      - auto-simulated samples: ‘learn and simulate’/‘look and simulate’
    - traced samples:
      - samples traced in transmitted light (by a look-through method),
      - samples traced by indented guidelines,
      - auto-traced samples: in transmitted light/by indented guidelines,
  - unintentional unnatural samples,
  - mixed samples.

In total, that gives us 11 types of handwriting samples (excluding mixed samples, as they are created by mixing some other types of handwriting samples): natural, unintentionally unnatural, disguised, auto-traced (both auto-traced in transmitted light and by indented guidelines), traced (in transmitted light and by indented guidelines), auto-simulated (both auto-‘learn and simulate’ and auto-‘look and simulate’ ones) and simulated (‘learn and simulate’ and ‘look and simulate’).

As it was suggested in the literature, some types of handwriting samples do have some characteristic

features that can distinguish them from other types of samples. For example, in disguised samples there are not as many writing habits and natural variations of handwriting as in natural samples (Matuszewski, 2011); there may be an apparent lack of fluency in the writing line (Bird, Found, Ballantyne, Rogers, 2010) and no uniformity of letter design or dimensions (Matuszewski, 2011). Similarly, in simulated samples we can encounter tremors, sudden changes of slant and lack of fluency in the writing line (Huber, Headrick, 1999; Matuszewski, 2011), very little or even no writing habits present (Muehlberger, 1990) and blunt endings (Matuszewski, 2011). At the same time, in traced samples there may be many pen stops and some sort of guiding line may be visible (Matuszewski, 2011), although such a line could be mistaken for an indentation made while signing a document lying atop of other documents (Ellen, Day, Davies, 2018).

Unfortunately, none of the authors provided a complex study on the differences observed. Most of the researchers compared only one type of handwriting sample with another one – usually with a natural writing (Harris, 1952–1953; Alford, 1970; Durina, 2005; Wendt, 2000; Leung, Cheng, Fung, Poon, 1993). Such a comparison is not sufficient when we are to decide if a handwriting sample is a natural sample or a handwriting sample of another type. Without comparing the sample with proper comparative material (e.g., because we have no such specimen), we can only rely on the writing features present in the sample itself. Therefore, it would be helpful to be able to use these features to classify the type of sample.

The chance of guessing the type of a handwriting sample correctly (so-called random selection probability) when choosing from 11 possible types of samples is only 1/11 (about 9%) – as there is no satisfactory method of classifying a handwriting sample as to its type, it is only possible to guess. Having a 9% chance of guessing the type of handwriting sample correctly is not sufficient. Even taking into consideration not 11, but 6 types of samples (excluding four types of auto-forgeries and taking all types of tracing for one, general type of traced sample), there is still merely a 17% chance for making correct classification (1/6). Therefore, such a method should be created. Moreover, a method should allow determining the type of a handwriting sample without the need of comparing it with any other sample. Therefore, the study was conducted in this area.

The three main research hypotheses were articulated:

(H1) Some types of handwriting samples do have characteristic features.

(H2) These features should allow to distinguish the said types of handwriting samples from other types of samples.

(H3) It is possible to create a method of classifying handwriting samples according to their type using the characteristic features.

To conclude, the research aimed at finding the characteristic features of 11 types of handwriting samples and using them to create an effective method of classifying a handwriting sample according to its type. By 'effective' it is meant 'more accurate than random selection probability'. To do that, some writing features of the collected handwriting samples were counted and measured and then, using discriminant analysis, used to classify samples as to their type. During validation, the accuracy of discriminant functions was tested as to their accuracy on a new group of samples. As a result, two different methods, called 'classifiers', were created (depending on the number of graphic features used). The classifiers were also created in two versions each, depending on how many types of handwriting

samples were taken into consideration (11 or 6), which gives us 4 methods in total.

## 2. Material and methods

The main part of the study was conducted on handwriting samples: full signatures of participants, fictional full signatures (the same fictional full name for all participants) and a single sentence, collected from 100 participants (66 females and 34 males, students of the Faculty of Law and Administration, Adam Mickiewicz University in Poznań). The sentences were collected as 11 types of handwriting samples, while both participants' full signatures and fictional full signatures were collected as 7 types of samples. Each type of sample used in the research was described in Table 1.

The number of types of handwriting samples collected in each of the three groups of samples (single

Table 1  
*The description of types of handwriting samples used in the research*

Type of handwriting sample	Description
Natural	a sample written in normal writing conditions, without any modifications (in a normal writing position, with normal writing habits, etc.)
Unintentionally unnatural	a sample written in unnatural writing conditions: in a standing position
Disguised	a sample in which a person tries to change his/her writing habits, e.g., by changing slant, the width of body letters, letter construction, etc.
Traced in transmitted light	a sample in which a person tries to copy a model sample, without revealing his/her writing habits, using transmitted light to create a guiding line from a model sample; to do this, a person puts a model sample under a piece of paper and then, using a source of light (e.g., a sun behind a window) thanks to the partial transparency of paper is able to see through the paper the model sample and can simply copy it
Traced by indented guidelines	a sample in which a person tries to copy a model sample, without revealing his/her writing habits, using indented guidelines from a model sample; to do this, a person puts a model sample on a piece of paper and traces the outline of each letter with a pen to create an indented impression of it on the paper below; in the next step, he gently writes over that indented impression with a pen to fill it with writing substance (e.g. ink)
'Learn and simulate'	a sample that simulates a model sample, in which a person is not able to observe the model sample during the writing and can only use its memorized image
'Look and simulate'	a sample that simulates a model sample, in which a person can simultaneously observe the model sample during the writing
Auto-traced (by indented guidelines/in transmitted light)	a sample in which a person tries to copy a model sample, without revealing his/her writing habits; however, the model sample has been created by the person who is now trying to copy it
Auto-simulated ('look and simulate'/'learn and simulate')	a sample that simulates a model sample; however, the model sample has been created by the person who is now trying to simulate it

Table 2  
Types of handwriting samples collected within each group of samples

Types of handwriting samples	Participant's signatures	Full names given to the participants	Single sentence
Natural	+	+	+
Unintentionally unnatural	+	+	+
Disguised	+	+	+
Auto-traced in transmitted light	+	-	+
Auto-traced by indented guidelines	+	-	+
Auto-‘learn and simulate’	+	-	+
Auto-‘look and simulate’	+	-	+
Traced in transmitted light	-	+	+
Traced by indented guidelines	-	+	+
‘Learn and simulate’	-	+	+
‘Look and simulate’	-	+	+

sentence, fictional full name given to the participants, participant's full name) are marked in Table 2.

In the second part of the study, the validation of the methods, the additional samples of the corresponding 11 types of samples, collected from another 39 participants (24 females, 15 males), were used (which gives 139 participants in total). Written consent was obtained from all participants and their personal data were not used in public – the research was conducted on the collected samples only.

In the handwriting samples both motor (Table 3) and dimensional features (change of slant, width or size of the letters) were assessed. All motor features were represented by their number per character in a sample; the dimensional features were represented by, respectively, their standard deviation, their coefficient of variation and arithmetic average.

These features, or, more precisely, the differences observed in number or within the size of features, were examined to find any existing statistical tendencies

Table 3  
Motor features counted in handwriting samples

Motor feature	Description
Pen stop	a sudden interruption in a stroke without removing the writing instrument from the paper
Pen lift	a sudden interruption in a stroke caused by removing the writing instrument from the paper
Hairlines	fine lines caused by continuation of a stroke regardless letter construction
Flying finishes	terminal strokes, growing out of the word ending
Changes of line direction	sudden changes of the direction within the single stroke, despite the letter construction
Retouches	corrections of the letter, made to change the shape or construction of the letter and not to change one letter into another one
Tremors	two or more consecutive changes of direction

and to check if they can be used to classify handwriting samples according to their type.

During the research, the following computer programmes were used:

- Graphlog 2.20 (Cyborg Idea) – to measure the dimensional features (standard deviation, arithmetic average and coefficient of variation of the features);
- Statistica 13.1 (StatSoft) – to perform statistical calculations (using discriminant analysis to calculate

the observed tendencies and then turn them into classifiers, thanks to the discriminant functions);

- Excel (Microsoft) – to validate the found classifiers.

Due to the length of samples, the largest number of writing features was assessed in sentences (23 of them in total); while in participants’ full names and in the fictional full names given to the participants there were 10 features assessed (Table 4).

Consequently, two different methods were created: the first, based on the smallest number of writing features that was possible to assess in each group of handwriting samples (Classifier 1); and the second based on 23 features that were possible to assess in sentences only (Classifier 2).

Each classifier was created in two versions:

- version ‘a’ – to classify 11 types of handwriting samples (11-type version, vide Table 1 and Table 2);
- version ‘b’ – to classify 6 types of handwriting samples (natural, disguised, unintentionally unnatural, traced – either traced in transmitted light or traced by indented guidelines, ‘look and simulate’ and ‘learn and simulate’ samples; 6-type version).

In this article only a basic method is described in full (Classifier 1), as it can be used for every sample, despite its length and the number of characters. Classifier 2 is discussed briefly, giving only its accuracy, as its application procedure is identical to that of Classifier 1. The discriminant functions of both classifiers are described in Appendix 1.

In the first step, the handwriting samples were collected and their writing features were assessed. The features were used to calculate the discriminant functions of the discriminant analysis [multivariate comparative analysis, Statistica 13.1 (Statsoft)] to classify handwriting samples as to their type. The classifying result is determined by the choice of the highest value of all discriminant functions. Each classifier consists

of a set of discriminant functions (their number depended on the number of possible types of handwriting samples, and varied from 6 to 11). To count the value of each discriminant function, we put the number of features assessed in the sample into the formula.

E.g., in Classifier 1 (11-types), the discriminant function for the natural sample ( $C_N$ ) was:

$$C_N = -0.50639 \cdot x_1 + 1.37615 \cdot x_2 + 1.81204 \cdot x_3 - 4.30809 \cdot x_4 + 0.32638 \cdot x_5 + 0.31590 \cdot x_6 + 8.70609 \cdot x_7 + 0.16095 \cdot x_8 + 0.25836 \cdot x_9 + 0.10630 \cdot x_{10} - 3.54610,$$

where  $x_1, x_2, \dots, x_{10}$  stand for the values of the subsequent features, measured in the sample.

Trying to determine if a handwriting sample is a natural one, we count the value of all discriminant functions and check if the discriminant function for a natural sample is the highest of all functions (for natural samples, unintentionally unnatural samples, disguised samples etc.). If not – the sample in question is possibly of another type (that obtained the highest value).

### 3. Results

It was demonstrated that most types of handwriting samples do have some characteristic features (Table 5). The observed tendencies were used in classifiers. The accuracy of the basic method (Classifier 1) was about 2.5 times higher than random selection probability when classifying 11 types of handwriting samples and about 2.17 times higher than random selection probability when classifying 6 types of handwriting samples (Table 6). During validation, Classifier 1 developed accuracy, respectively: 2.67 and 2.5 times higher than random selection probability. Similar results were achieved by Classifier 2. All methods turned out to be

Table 4  
Graphic features assessed in full names and in sentences

Handwriting sample	Writing features
Full names (both participant’s and a fictional one)	pen stops, pen lifts, hairlines, flying finishes, changes of direction, retouches, tremors, change of body-height of the letters (standard deviation), change of width in oval forms (standard deviation), change of slant in double-length letters (standard deviation)
A sentence	pen stops, pen lifts, hairlines, flying finishes, changes of direction, retouches, tremors, change of body-height of the linear letters (standard deviation, coefficient of variation), change of height in supralinear letters (standard deviation, coefficient of variation), change of height in infralinear letters (standard deviation, coefficient of variation), change of width in oval forms (standard deviation, coefficient of variation), change of slant in double-length letters (standard deviation, coefficient of variation), the arithmetic average of body-height of the linear letters, the arithmetic average of height in supralinear letters, the arithmetic average of height in infralinear letters, the arithmetic average of width in oval forms, the arithmetic average of slant in double-length letters, number of used variations of the same letter

**Table 5**  
*The prevalence rate of some writing features in six types of handwriting samples*

Writing feature	Type of handwriting sample					
	Natural	Disguised	Unintentionally unnatural	'Look and simulate'	'Learn and simulate'	Traced
Pen stops, pen lifts, changes of direction	low	average	low	average	low	high
Hairlines	high	average	high	average	high	low
Tremors	low	low/average	low	average	low	high
Flying finishes	average	average	high	average	average	low
Retouches	low	average	low	average	low	high
Change of body-height of the linear letters	low	high	average/high	low/average	average	average/high
Number of used variations of the same letter	low/average	low	low/average	average	average	average

**Table 6**  
*The accuracy [%] of the classifiers based on 10 or 23 writing features*

Number of types of handwriting samples	Random selection probability	Classifier1 (10 features)		Classifier 2 (23 features)	
		Test	Validation	Test	Validation
11	9%	23.23%	24%	31.12%	25%
6	17%	37.36%	43%	45.3%	44%

more than twice as accurate as random selection probability. Moreover, during validation, they turned out to be at least equally as effective as in the first part of the research.

All methods were also compared in terms of their accuracy when classifying different kinds of samples (full signature of participants, fictional full signature, and a single sentence); for example, Classifier 1 achieved the best accuracy when used on single

sentences (47% when dealing with 6 types of handwriting samples, see Table 7). Similar results were achieved by the second classifier. However, it was least effective when used on fictional full signatures (37%). And again, the classifiers proved to be more than twice as accurate as the random selection probability.

**4. Discussion**

The main thesis of the research has been confirmed – most types of handwriting samples do have specific features that allow them to be distinguished from other types of handwriting samples. Moreover, these features can be used to create a method of classifying handwriting samples according to their type.

As for writing features, some of them seem to be highly connected with specific types of samples, e.g., the number of pen stops is significantly higher in traced samples. Although some of these tendencies have previously been briefly described in the literature, a full comparison has not been provided before. Moreover,

**Table 7**  
*The accuracy [%] of Classifier 1 when used on different kinds of samples*

Handwriting sample	Number of types of handwriting samples	
	11	6
Participant's signature	16	42
Signature based on a name given to the participant	32	37
Single sentence	23	47

some of the researched features were omitted in the literature. During research, all writing features were used in discriminant functions of classifiers, and proved to be useful when classifying handwriting samples as to their types. However, the described features are only a few of the writing features that can be assessed in a handwriting sample. Therefore, it would be vital to take into consideration more writing features; such a research could possibly contribute to improving the accuracy of classifiers.

Due to the current lack of any scientific method for classifying handwriting samples as to their type, the results of the current research are satisfactory. Each method of classifying handwriting samples as to its type proved to be more than twice as effective as random selection probability, regardless of the number of types of handwriting sample considered. Classifier 1 was similarly effective during the first part of research and validation. It also proved to be reasonably effective: despite being the simplest one (using 10 handwriting features as opposed to 23), it resulted in equally good accuracy when classifying both full signatures and sentences. It turns out that greater number of features does not always lead to better accuracy in classification. Moreover, the simpler the classifier is, the more useful it becomes in practice. This is particularly important when dealing with short handwriting samples, with very few writing features.

It must be clearly stated that these methods of classifying handwriting samples as to their type are only the first attempt at solving the problem of creating such methods – the proposed classifiers should be improved to obtain greater accuracy.

Some issues should be taken into consideration in the future research on classifying handwriting samples as to their type. Firstly, using additional writing features could improve the classifiers – for example, the use of dimensional features should be possible when classifying longer handwriting samples. Secondly, the method of collecting handwriting samples should be simplified – instead of collecting every type of handwriting sample from every participant, perhaps it would be better to focus on each type separately and gather as many specimens as possible; it would also be less time-consuming for the participants. Last, but not least, the participants should be a more diverse group, as well as for their age, vocation and, perhaps, their geographic origin.

## 5. Conclusions

In conclusion, the proposed methods of classifying handwriting samples according to their type seem to have scientific potential, as they proved to be more than twice as accurate as random selection probability. The simplest of the methods, Classifier 1, seems to be the most useful one, as it does not require too many features to be present in a sample (it can be used both on shorter and longer handwriting samples) and yet it yields results that are as good as those provided by more complicated methods.

However, the final effectiveness of all methods is still too low for them to be used in practice. That is why more complex and thorough research needs to be provided in this area – perhaps including new writing features and improving the collection of handwriting samples will lead to higher accuracy of the methods.

## Notes from the author

The study is based on the larger research, done by the author as a part of her Doctor's dissertation ('Classifying handwriting as to its type', Poznań, 2021, unpublished).

## References

1. Alford, E. F. (1970). Disguised handwriting. A statistical survey of how handwriting is most frequently disguised. *Journal of Forensic Sciences*, 15, 476–488.
2. Bird, C., Found, B., Ballantyne, K., Rogers, D. (2010). Forensic handwriting examiners' opinions on the process of production of disguised and simulated signatures. *Forensic Science International*, 195, 103–107.
3. Durina, M. E. (2005). Disguised signatures: random or repetitious? *Journal of the American Society of Questioned Document Examiners*, 8, 9–16.
4. Ellen, D., Day, S., Davies, C. (2018). *The scientific examination of documents. Methods and techniques* (4th ed., p. 248). Boca Raton, Florida: CRC Press.
5. Harris, J. J. (1953). Disguised handwriting. *Journal of Criminal Law, Criminology & Police Science*, 43, 5, 685–689.
6. Huber, R. A., Headrick, A. M. (1999). *Handwriting identification: Facts and fundamentals* (p. 434). Boca Raton, Florida: CRC Press.
7. Koppenhaver, K. M. (2007). *Forensic document examination: Principles and practice* (p. 307). Totowa, New York: Humana Press Inc.
8. Leung, S. C., Cheng, Y. S., Fung, H. T., Poon, N. L. (1993). Forgery I – simulation. *Journal of Forensic Sciences*, 38, 402–412.

9. Matuszewski, S. (2011). Types of handwriting samples. *Problems of Forensic Sciences*, 87, 181–192.
10. Muehlberger, J. (1990). Identifying simulations: practical considerations. *Journal of Forensic Sciences*, 35, 368–374.
11. Wendt, G. W. (2003). Statistical observations of disguised signatures. *Journal of the American Society of Questioned Document Examiners*, 3, 19–27.

**ORCID**Jagoda Dzida  0000-0001-9057-5294

---

**Corresponding author**

Dr. Jagoda Dzida  
Laboratory of Criminalistics, Department of Law and  
Administration  
Adam Mickiewicz University in Poznań  
ul. Św. Marcin 90  
PL 61-809 Poznań  
e-mail: jagoda.dzida@amu.edu.pl

---



## APPENDIX – CLASSIFIERS

### 1. Classifier 1

The first method of classifying handwriting samples as to their type uses 10 graphic features:

$x_1$  – number of pen lifts in a sample

$x_2$  – number of pen stops in a sample

$x_3$  – number of changes of direction (despite the proper letter construction) within a sample

$x_4$  – number of tremors in a sample

$x_5$  – number of flying finishes (related to the number of words in a sample)

$x_6$  – number of retouches in a sample

$x_7$  – number of hairlines in a sample

$x_8$  – standard deviation of the change of body-height of the letters

$x_9$  – standard deviation of the change of slant in double-length letters

$x_{10}$  – standard deviation of the change of width in oval forms

#### Classifier 1a, used to classify 11 types of handwriting samples:

$C_N$  – natural sample

$C_{UN}$  – unintentionally unnatural sample

$C_D$  – disguised sample

$C_{aTl}$  – auto-traced in transmitted light

$C_{aTg}$  – auto-traced by indented guidelines

$C_{Tl}$  – traced in transmitted light

$C_{Tg}$  – traced by indented guidelines

$C_{aSle}$  – auto-simulated ‘learn and simulate’

$C_{aSlo}$  – auto-simulated ‘look and simulate’

$C_{Sle}$  – simulated ‘learn and simulate’

$C_{Slo}$  – simulated ‘look and simulate’

Discriminant functions for each type of handwriting sample:

$$C_N = -0.50639 \cdot x_1 + 1.37615 \cdot x_2 + 1.81204 \cdot x_3 - 4.30809 \cdot x_4 + 0.32638 \cdot x_5 + 0.31590 \cdot x_6 + 8.70609 \cdot x_7 + 0.16095 \cdot x_8 + 0.25836 \cdot x_9 + 0.10630 \cdot x_{10} - 3.54610$$

$$C_{UN} = 0.33823 \cdot x_1 + 0.61408 \cdot x_2 + 1.92933 \cdot x_3 - 4.01528 \cdot x_4 + 0.97182 \cdot x_5 + 0.59117 \cdot x_6 + 20.62315 \cdot x_7 + 0.08868 \cdot x_8 + 0.25104 \cdot x_9 + 0.12394 \cdot x_{10} - 4.27656$$

$$C_D = 4.69673 \cdot x_1 + 6.12979 \cdot x_2 + 2.64383 \cdot x_3 - 8.35297 \cdot x_4 + 0.34299 \cdot x_5 + 7.89985 \cdot x_6 - 0.18655 \cdot x_7 + 0.10160 \cdot x_8 + 0.28412 \cdot x_9 + 0.20776 \cdot x_{10} - 4.48035$$

$$C_{aTl} = 6.84151 \cdot x_1 + 5.73008 \cdot x_2 + 3.95199 \cdot x_3 - 5.70635 \cdot x_4 + 0.00650 \cdot x_5 - 1.36779 \cdot x_6 - 1.49988 \cdot x_7 + 0.10323 \cdot x_8 + 0.28523 \cdot x_9 + 0.12981 \cdot x_{10} - 5.08275$$

$$C_{aTg} = 9.64383 \cdot x_1 + 6.26925 \cdot x_2 + 3.69185 \cdot x_3 + 1.68855 \cdot x_4 - 0.08061 \cdot x_5 - 3.84100 \cdot x_6 - 1.27796 \cdot x_7 + 0.11135 \cdot x_8 + 0.30518 \cdot x_9 + 0.13147 \cdot x_{10} - 5.82064$$

$$C_{T1} = 10.89122 \cdot x_1 + 4.56654 \cdot x_2 + 3.39189 \cdot x_3 + 0.47408 \cdot x_4 - 0.11157 \cdot x_5 + 1.04413 \cdot x_6 - 1.64181 \cdot x_7 + 0.11834 \cdot x_8 + 0.29832 \cdot x_9 + 0.15370 \cdot x_{10} - 5.55724$$

$$C_{Tg} = 10.32443 \cdot x_1 + 5.61368 \cdot x_2 + 3.36626 \cdot x_3 + 6.55360 \cdot x_4 - 0.12554 \cdot x_5 - 3.25517 \cdot x_6 - 1.60897 \cdot x_7 + 0.11999 \cdot x_8 + 0.30273 \cdot x_9 + 0.15622 \cdot x_{10} - 6.01443$$

$$C_{aSle} = -0.53311 \cdot x_1 + 0.66983 \cdot x_2 + 2.51803 \cdot x_3 - 5.73353 \cdot x_4 + 0.50360 \cdot x_5 + 2.31452 \cdot x_6 + 11.45235 \cdot x_7 + 0.09370 \cdot x_8 + 0.30082 \cdot x_9 + 0.12224 \cdot x_{10} - 4.04797$$

$$C_{aSlo} = -0.41953 \cdot x_1 + 4.17300 \cdot x_2 + 2.71915 \cdot x_3 - 6.14231 \cdot x_4 + 0.17231 \cdot x_5 + 4.62019 \cdot x_6 + 0.13384 \cdot x_7 + 0.09108 \cdot x_8 + 0.26211 \cdot x_9 + 0.11450 \cdot x_{10} - 3.74801$$

$$C_{Sle} = 2.93319 \cdot x_1 + 2.97093 \cdot x_2 + 1.98706 \cdot x_3 - 5.33022 \cdot x_4 + 0.15899 \cdot x_5 + 6.22575 \cdot x_6 + 13.25517 \cdot x_7 + 0.11279 \cdot x_8 + 0.25604 \cdot x_9 + 0.13536 \cdot x_{10} - 3.93176$$

$$C_{Slo} = 6.56276 \cdot x_1 + 5.64105 \cdot x_2 + 2.52555 \cdot x_3 - 5.59608 \cdot x_4 - 0.02405 \cdot x_5 + 7.18898 \cdot x_6 - 0.86159 \cdot x_7 + 0.11694 \cdot x_8 + 0.29833 \cdot x_9 + 0.13728 \cdot x_{10} - 4.57483$$

To check what is the type of the handwriting sample, we need to calculate values for each discriminant function. The highest of the received values indicates the type of our handwriting sample. For example, if the highest value is calculated for  $C_N$  function, our sample is probably a natural sample.

The second version of the first method enables user to classify 6 types of handwriting samples (Classifier 1b):

$C_N$  – natural sample

$C_{UN}$  – unintentionally unnatural sample

$C_D$  – disguised sample

$C_T$  – traced (either in transmitted light or by indented guidelines)

$C_{Sle}$  – simulated ‘learn and simulate’

$C_{Slo}$  – simulated ‘look and simulate’

Discriminant functions:

$$C_N = -0.55143 \cdot x_1 + 1.37787 \cdot x_2 + 1.81261 \cdot x_3 - 4.33349 \cdot x_4 + 0.32608 \cdot x_5 + 0.34534 \cdot x_6 + 8.75200 \cdot x_7 + 0.16123 \cdot x_8 + 0.25808 \cdot x_9 + 0.10650 \cdot x_{10} - 2.93978$$

$$C_{UN} = 0.32093 \cdot x_1 + 0.62067 \cdot x_2 + 1.92829 \cdot x_3 - 4.03260 \cdot x_4 + 0.97182 \cdot x_5 + 0.63650 \cdot x_6 + 20.69676 \cdot x_7 + 0.08890 \cdot x_8 + 0.25076 \cdot x_9 + 0.12429 \cdot x_{10} - 3.67196$$

$$C_D = 4.37263 \cdot x_1 + 6.06900 \cdot x_2 + 2.66555 \cdot x_3 - 8.50697 \cdot x_4 + 0.35492 \cdot x_5 + 7.81786 \cdot x_6 - 0.16035 \cdot x_7 + 0.10055 \cdot x_8 + 0.28338 \cdot x_9 + 0.20713 \cdot x_{10} - 3.85787$$

$$C_T = 8.96484 \cdot x_1 + 5.39154 \cdot x_2 + 3.63019 \cdot x_3 + 0.33130 \cdot x_4 - 0.06087 \cdot x_5 - 1.80477 \cdot x_6 - 1.48835 \cdot x_7 + 0.11166 \cdot x_8 + 0.29620 \cdot x_9 + 0.14193 \cdot x_{10} - 4.90495$$

$$C_{Sle} = 1.22184 \cdot x_1 + 1.90516 \cdot x_2 + 2.23641 \cdot x_3 - 5.58138 \cdot x_4 + 0.31928 \cdot x_5 + 4.44308 \cdot x_6 + 12.49008 \cdot x_7 + 0.10387 \cdot x_8 + 0.27600 \cdot x_9 + 0.12929 \cdot x_{10} - 3.33126$$

$$C_{Slo} = 2.95244 \cdot x_1 + 4.89069 \cdot x_2 + 2.63769 \cdot x_3 - 6.02388 \cdot x_4 + 0.07900 \cdot x_5 + 5.91184 \cdot x_6 - 0.36225 \cdot x_7 + 0.10375 \cdot x_8 + 0.28039 \cdot x_9 + 0.12574 \cdot x_{10} - 3.49422$$

## 2. Classifier 2

### Classifier 2

The third method of classifying handwriting samples as to their type uses 23 graphic features:

- $x_1$  – number of pen lifts in a sample
- $x_2$  – number of pen stops in a sample
- $x_3$  – number of changes of direction (despite the proper letter construction) within a sample
- $x_4$  – number of tremors in a sample
- $x_5$  – number of flying finishes (related to the number of words in a sample)
- $x_6$  – number of retouches in a sample
- $x_7$  – number of hairlines in a sample
- $x_8$  – coefficient of variation of the change of body-height of the linear letters
- $x_9$  – coefficient of variation of the change of height in supralinear letters
- $x_{10}$  – coefficient of variation of the change of height in infralinear letters
- $x_{11}$  – coefficient of variation of the change of slant in double-length letters
- $x_{12}$  – coefficient of variation of the change of width in oval forms
- $x_{13}$  – the average number of used variations of the same letter
- $x_{14}$  – standard deviation of the change of body-height of the linear letters
- $x_{15}$  – arithmetic average of the change of body-height of the linear letters
- $x_{16}$  – standard deviation of the change of height in supralinear letters
- $x_{17}$  – arithmetic average of the change of height in supralinear letters
- $x_{18}$  – standard deviation of the change of height in infralinear letters
- $x_{19}$  – arithmetic average of the change of height in infralinear letters
- $x_{20}$  – standard deviation of the change of slant in double-length letters
- $x_{21}$  – arithmetic average of the change of slant in double-length letters
- $x_{22}$  – standard deviation of the change of width in oval forms
- $x_{23}$  – arithmetic average of the change of width in oval forms

### Classifier 2a, used to classify 11 types of handwriting samples:

- $C_N$  – natural sample
- $C_{UN}$  – unintentionally unnatural sample
- $C_D$  – disguised sample
- $C_{aTl}$  – auto-traced in transmitted light
- $C_{aTg}$  – auto-traced by indented guidelines
- $C_{Tl}$  – traced in transmitted light
- $C_{Tg}$  – traced by indented guidelines
- $C_{aSle}$  – auto-simulated ‘learn and simulate’
- $C_{aSlo}$  – auto-simulated ‘look and simulate’
- $C_{Sle}$  – simulated ‘learn and simulate’
- $C_{Slo}$  – simulated ‘look and simulate’

Discriminant functions for each type of handwriting sample:

$$C_N = -11.003 \cdot x_1 - 16.499 \cdot x_2 + 5.630 \cdot x_3 - 30.080 \cdot x_4 - 12.899 \cdot x_5 + 18.347 \cdot x_6 + 3.185 \cdot x_7 + 175.392 \cdot x_8 + 715.779 \cdot x_9 + 10.929 \cdot x_{10} - 194.941 \cdot x_{11} + 42.063 \cdot x_{12} + 10.441 \cdot x_{13} - 40.466 \cdot x_{14} + 3.462 \cdot x_{15} - 101.536 \cdot x_{16} + 21.473 \cdot x_{17} + 0.829 \cdot x_{18} + 0.020 \cdot x_{19} - 1.260 \cdot x_{20} - 0.192 \cdot x_{21} - 1.158 \cdot x_{22} - 0.002 \cdot x_{23} - 114.457$$

$$C_{UN} = -5.829 \cdot x_1 - 19.059 \cdot x_2 + 6.244 \cdot x_3 - 32.766 \cdot x_4 - 5.309 \cdot x_5 + 18.337 \cdot x_6 + 21.148 \cdot x_7 + 183.760 \cdot x_8 + 714.860 \cdot x_9 + 9.967 \cdot x_{10} - 210.815 \cdot x_{11} + 42.157 \cdot x_{12} + 11.689 \cdot x_{13} - 40.945 \cdot x_{14} + 3.460 \cdot x_{15} - 100.432 \cdot x_{16} + 21.623 \cdot x_{17} + 1.283 \cdot x_{18} + 0.031 \cdot x_{19} - 1.354 \cdot x_{20} - 0.195 \cdot x_{21} - 2.245 \cdot x_{22} - 0.027 \cdot x_{23} - 122.545$$

$$C_D = -3.995 \cdot x_1 - 2.025 \cdot x_2 + 7.069 \cdot x_3 - 35.831 \cdot x_4 - 13.344 \cdot x_5 + 40.481 \cdot x_6 - 12.478 \cdot x_7 + 178.860 \cdot x_8 + 724.727 \cdot x_9 + 7.863 \cdot x_{10} - 206.561 \cdot x_{11} + 44.201 \cdot x_{12} + 11.394 \cdot x_{13} - 42.214 \cdot x_{14} + 3.809 \cdot x_{15} - 101.309 \cdot x_{16} + 21.598 \cdot x_{17} + 1.523 \cdot x_{18} + 0.057 \cdot x_{19} - 1.214 \cdot x_{20} - 0.192 \cdot x_{21} - 2.19 \cdot x_{22} - 0.027 \cdot x_{23} - 122.307$$

$$C_{aTI} = 6.545 \cdot x_1 - 4.286 \cdot x_2 + 7.313 \cdot x_3 - 26.959 \cdot x_4 - 20.313 \cdot x_5 + 23.914 \cdot x_6 - 10.344 \cdot x_7 + 181.064 \cdot x_8 + 705.180 \cdot x_9 + 12.824 \cdot x_{10} - 190.240 \cdot x_{11} + 44.205 \cdot x_{12} + 12.147 \cdot x_{13} - 41.252 \cdot x_{14} + 3.579 \cdot x_{15} - 98.996 \cdot x_{16} + 21.217 \cdot x_{17} + 0.998 \cdot x_{18} + 0.032 \cdot x_{19} - 1.078 \cdot x_{20} - 0.192 \cdot x_{21} - 2.703 \cdot x_{22} - 0.016 \cdot x_{23} - 121.179$$

$$C_{aTg} = 11.777 \cdot x_1 - 6.521 \cdot x_2 + 7.546 \cdot x_3 - 16.482 \cdot x_4 - 21.094 \cdot x_5 + 9.704 \cdot x_6 - 10.861 \cdot x_7 + 170.773 \cdot x_8 + 712.496 \cdot x_9 + 17.077 \cdot x_{10} - 169.223 \cdot x_{11} + 6.962 \cdot x_{12} + 12.449 \cdot x_{13} - 38.499 \cdot x_{14} + 3.455 \cdot x_{15} - 100.502 \cdot x_{16} + 21.353 \cdot x_{17} + 0.798 \cdot x_{18} + 0.026 \cdot x_{19} - 0.758 \cdot x_{20} - 0.189 \cdot x_{21} - 3.650 \cdot x_{22} - 0.014 \cdot x_{23} - 123.152$$

$$C_{TI} = 15.187 \cdot x_1 - 4.459 \cdot x_2 + 7.599 \cdot x_3 - 16.385 \cdot x_4 - 19.566 \cdot x_5 + 17.462 \cdot x_6 - 12.797 \cdot x_7 + 180.009 \cdot x_8 + 699.138 \cdot x_9 + 17.488 \cdot x_{10} - 184.862 \cdot x_{11} + 7.763 \cdot x_{12} + 12.997 \cdot x_{13} - 39.913 \cdot x_{14} + 3.576 \cdot x_{15} - 98.793 \cdot x_{16} + 20.926 \cdot x_{17} + 0.406 \cdot x_{18} - 0.027 \cdot x_{19} - 1.024 \cdot x_{20} - 0.191 \cdot x_{21} - 0.391 \cdot x_{22} - 0.018 \cdot x_{23} - 122.111$$

$$C_{Tg} = 14.238 \cdot x_1 - 3.990 \cdot x_2 + 7.582 \cdot x_3 - 5.212 \cdot x_4 - 20.025 \cdot x_5 + 6.226 \cdot x_6 - 12.719 \cdot x_7 + 170.575 \cdot x_8 + 697.403 \cdot x_9 + 18.152 \cdot x_{10} - 188.292 \cdot x_{11} + 43.048 \cdot x_{12} + 12.861 \cdot x_{13} - 38.226 \cdot x_{14} + 3.512 \cdot x_{15} - 98.422 \cdot x_{16} + 20.786 \cdot x_{17} + 0.304 \cdot x_{18} - 0.027 \cdot x_{19} - 1.016 \cdot x_{20} - 0.190 \cdot x_{21} - 1.143 \cdot x_{22} - 0.017 \cdot x_{23} - 121.205$$

$$C_{aSle} = -9.536 \cdot x_1 - 17.126 \cdot x_2 + 6.222 \cdot x_3 - 32.741 \cdot x_4 - 13.719 \cdot x_5 + 22.562 \cdot x_6 + 2.299 \cdot x_7 + 182.756 \cdot x_8 + 718.247 \cdot x_9 + 9.831 \cdot x_{10} - 228.314 \cdot x_{11} + 43.615 \cdot x_{12} + 11.692 \cdot x_{13} - 39.421 \cdot x_{14} + 3.516 \cdot x_{15} - 101.535 \cdot x_{16} + 21.527 \cdot x_{17} + 1.218 \cdot x_{18} + 0.031 \cdot x_{19} - 1.504 \cdot x_{20} - 0.210 \cdot x_{21} - 1.804 \cdot x_{22} - 0.010 \cdot x_{23} - 123.332$$

$$C_{aSlo} = -6.503 \cdot x_1 - 8.805 \cdot x_2 + 6.471 \cdot x_3 - 34.876 \cdot x_4 - 16.986 \cdot x_5 + 24.381 \cdot x_6 - 10.060 \cdot x_7 + 184.712 \cdot x_8 + 718.703 \cdot x_9 + 7.999 \cdot x_{10} - 194.732 \cdot x_{11} + 37.906 \cdot x_{12} + 10.787 \cdot x_{13} - 43.324 \cdot x_{14} + 3.560 \cdot x_{15} - 101.529 \cdot x_{16} + 21.561 \cdot x_{17} + 1.322 \cdot x_{18} + 0.046 \cdot x_{19} - 1.156 \cdot x_{20} - 0.192 \cdot x_{21} - 0.143 \cdot x_{22} - 0.027 \cdot x_{23} - 117.115$$

$$C_{Sle} = -1.852 \cdot x_1 - 12.597 \cdot x_2 + 5.622 \cdot x_3 - 34.186 \cdot x_4 - 18.017 \cdot x_5 + 25.173 \cdot x_6 + 8.350 \cdot x_7 + 175.235 \cdot x_8 + 718.936 \cdot x_9 + 9.775 \cdot x_{10} - 208.558 \cdot x_{11} + 44.687 \cdot x_{12} + 12.001 \cdot x_{13} - 37.497 \cdot x_{14} + 3.511 \cdot x_{15} - 101.739 \cdot x_{16} + 21.398 \cdot x_{17} + 1.111 \cdot x_{18} - 0.004 \cdot x_{19} - 1.393 \cdot x_{20} - 0.192 \cdot x_{21} - 1.370 \cdot x_{22} + 0.011 \cdot x_{23} - 121.178$$

$$C_{Slo} = 3.954 \cdot x_1 - 6.225 \cdot x_2 + 6.879 \cdot x_3 - 32.049 \cdot x_4 - 20.042 \cdot x_5 + 37.586 \cdot x_6 - 11.739 \cdot x_7 + 188.644 \cdot x_8 + 704.925 \cdot x_9 + 10.202 \cdot x_{10} - 195.955 \cdot x_{11} + 9.691 \cdot x_{12} + 12.610 \cdot x_{13} - 41.671 \cdot x_{14} + 3.634 \cdot x_{15} - 100.214 \cdot x_{16} + 21.150 \cdot x_{17} + 1.484 \cdot x_{18} + 0.048 \cdot x_{19} - 1.160 \cdot x_{20} - 0.191 \cdot x_{21} + 0.051 \cdot x_{22} - 0.018 \cdot x_{23} - 122.550$$

**Classifier 2b, used to classify 6 types of handwriting samples:**

$C_N$  – natural sample

$C_{UN}$  – unintentionally unnatural sample

$C_D$  – disguised sample

$C_{TI}$  – traced (either in transmitted light or by indented guidelines)

$C_{Sle}$  – simulated ‘learn and simulate’

$C_{Slo}$  – simulated ‘look and simulate’

Discriminant functions for each type of handwriting sample:

$$C_N = -11.613 \cdot x_1 - 16.241 \cdot x_2 + 5.528 \cdot x_3 - 29.467 \cdot x_4 - 12.748 \cdot x_5 + 15.947 \cdot x_6 + 3.908 \cdot x_7 + 174.771 \cdot x_8 + 719.958 \cdot x_9 + 10.788 \cdot x_{10} - 195.645 \cdot x_{11} + 41.432 \cdot x_{12} + 10.267 \cdot x_{13} - 40.708 \cdot x_{14} + 3.473 \cdot x_{15} - 102.046 \cdot x_{16} + 21.566 \cdot x_{17} + 0.723 \cdot x_{18} + 0.008 \cdot x_{19} - 1.283 \cdot x_{20} - 0.191 \cdot x_{21} - 0.800 \cdot x_{22} - 0.001 \cdot x_{23} - 113.751$$

$$C_{UN} = -6.682 \cdot x_1 - 18.863 \cdot x_2 + 6.130 \cdot x_3 - 32.202 \cdot x_4 - 5.058 \cdot x_5 + 15.530 \cdot x_6 + 22.036 \cdot x_7 + 182.989 \cdot x_8 + 719.415 \cdot x_9 + 9.759 \cdot x_{10} - 211.498 \cdot x_{11} + 41.518 \cdot x_{12} + 11.471 \cdot x_{13} - 41.218 \cdot x_{14} + 3.469 \cdot x_{15} - 100.973 \cdot x_{16} + 21.728 \cdot x_{17} + 1.175 \cdot x_{18} + 0.019 \cdot x_{19} - 1.377 \cdot x_{20} - 0.194 \cdot x_{21} - 1.902 \cdot x_{22} - 0.027 \cdot x_{23} - 121.747$$

$$C_D = -5.409 \cdot x_1 - 1.921 \cdot x_2 + 6.953 \cdot x_3 - 35.551 \cdot x_4 - 12.980 \cdot x_5 + 37.416 \cdot x_6 - 11.763 \cdot x_7 + 178.079 \cdot x_8 + 729.944 \cdot x_9 + 7.518 \cdot x_{10} - 207.414 \cdot x_{11} + 43.594 \cdot x_{12} + 11.099 \cdot x_{13} - 42.595 \cdot x_{14} + 3.816 \cdot x_{15} - 101.912 \cdot x_{16} + 21.725 \cdot x_{17} + 1.423 \cdot x_{18} + 0.047 \cdot x_{19} - 1.236 \cdot x_{20} - 0.192 \cdot x_{21} - 1.913 \cdot x_{22} - 0.027 \cdot x_{23} - 121.416$$

$$C_T = 9.401 \cdot x_1 - 4.907 \cdot x_2 + 7.365 \cdot x_3 - 17.446 \cdot x_4 - 19.732 \cdot x_5 + 11.676 \cdot x_6 - 10.803 \cdot x_7 + 175.428 \cdot x_8 + 709.756 \cdot x_9 + 15.582 \cdot x_{10} - 184.498 \cdot x_{11} + 42.411 \cdot x_{12} + 12.180 \cdot x_{13} - 40.158 \cdot x_{14} + 3.538 \cdot x_{15} - 99.850 \cdot x_{16} + 21.243 \cdot x_{17} + 0.564 \cdot x_{18} - 0.005 \cdot x_{19} - 0.999 \cdot x_{20} - 0.191 \cdot x_{21} - 1.811 \cdot x_{22} - 0.018 \cdot x_{23} - 120.644$$

$$C_{Sle} = -7.009 \cdot x_1 - 14.773 \cdot x_2 + 5.801 \cdot x_3 - 33.228 \cdot x_4 - 15.568 \cdot x_5 + 20.942 \cdot x_6 + 6.144 \cdot x_7 + 178.266 \cdot x_8 + 723.654 \cdot x_9 + 9.467 \cdot x_{10} - 219.312 \cdot x_{11} + 43.574 \cdot x_{12} + 11.571 \cdot x_{13} - 38.808 \cdot x_{14} + 3.520 \cdot x_{15} - 102.228 \cdot x_{16} + 21.585 \cdot x_{17} + 1.067 \cdot x_{18} + 0.003 \cdot x_{19} - 1.473 \cdot x_{20} - 0.200 \cdot x_{21} - 1.300 \cdot x_{22} + 0.000 \cdot x_{23} - 121.313$$

$$C_{Slo} = -2.942 \cdot x_1 - 7.487 \cdot x_2 + 6.550 \cdot x_3 - 33.432 \cdot x_4 - 18.143 \cdot x_5 + 27.853 \cdot x_6 - 10.129 \cdot x_7 + 185.954 \cdot x_8 + 717.278 \cdot x_9 + 8.665 \cdot x_{10} - 196.157 \cdot x_{11} + 8.225 \cdot x_{12} + 11.369 \cdot x_{13} - 42.930 \cdot x_{14} + 3.602 \cdot x_{15} - 101.499 \cdot x_{16} + 21.494 \cdot x_{17} + 1.311 \cdot x_{18} + 0.038 \cdot x_{19} - 1.180 \cdot x_{20} - 0.191 \cdot x_{21} + 0.208 \cdot x_{22} - 0.023 \cdot x_{23} - 118.659$$

## KLASYFIKOWANIE PRÓBY PISMA RĘCZNEGO WEDŁUG TYPU Z WYKORZYSTANIEM ANALIZY DYSKRYMINACYJNEJ

### 1. Wprowadzenie

Badania pisma ręcznego są zazwyczaj nastawione na określenie wykonawstwa danej próby. Biegły musi w tym celu porównać materiał porównawczy z materiałem dowodowym. Materiał porównawczy powinien składać się z prób pisma tego samego typu co materiał dowodowy (a więc – przykładowo – musi obejmować naturalną próbę pisma, jeśli materiał dowodowy jest piśmem naturalnym danej osoby; jeśli materiał kwestionowany składa się z prób maskowanych, to materiał porównawczy również musi takie próby zawierać).

Jeśli znany jest typ próby stanowiącej materiał dowodowy, to pobranie adekwatnego materiału porównawczego staje się dużo prostsze i szybsze. Ponadto znajomość typu próby stanowiącej materiał kwestionowany jest również przydatna dla samego biegłego pismoznawcy, gdyż pozwala mu ona na poprawną ocenę zaobserwowanych przez siebie różnic i podobieństw między badanymi próbami pisma; na przykład jeśli materiał kwestionowany jest próbą kopiowaną, to spodziewamy się w niej obecności tremoru. Biegły, sklasyfikowawszy daną próbę pisma według jej typu, mógłby skupić się w szczególności na konkretnych cechach graficznych, które w danym przypadku (z uwagi na typ próby) będą najważniejsze.

W doktrynie znanych jest kilka możliwych klasyfikacji typów prób: Huber i Headrick (1999) dokonali opisu prób naturalnych, maskowanych i sfalszowanych (naśladowanych i kopiowanych); podobnie typy opisała również Koppenhaver (2007), która więcej uwagi poświęciła różnicom między danymi typami prób (na przykład maskowanymi i naśladowanymi).

Nieco odmienna terminologia została zaproponowana przez Matuszewskiego (2011). Dokonał on następującego podziału typów prób:

- próby naturalne;
- próby nienaturalne:
  - próby zamierzenie nienaturalne:
    - próby maskowane,
    - próby naśladowane:
      - próby naśladowane swobodnie,
      - próby naśladowane ściśle (niewolniczo),
      - próby autonaśladowane: swobodnie/ściśle (niewolniczo),
    - próby kopiowane:
      - próby kopiowane w prześwicie,
      - próby kopiowane przez przeciskanie,
      - próby autokopiowane: w prześwicie/przez przeciskanie,
  - próby niezamierzenie nienaturalne,
  - próby mieszane.

Łącznie daje nam to 11 typów prób pisma ręcznego (z wyłączeniem prób mieszanych, powstałych poprzez połączenie kilku spośród wskazanych wyżej typów): próby naturalne, niezamierzenie nienaturalne, maskowane, autokopiowane (autokopiowane w prześwicie i autokopiowane przez przeciskanie), kopiowane (w prześwicie i przez przeciskanie), autonaśladowane (autonaśladowane ściśle i autonaśladowane swobodnie) oraz naśladowane (swobodnie i ściśle).

Jak zasugerowano w literaturze, niektóre typy prób mają pewne cechy charakterystyczne, które umożliwiają ich odróżnienie od prób innych typów. Na przykład w próbach maskowanych należy się spodziewać niewielu cech nawykowych (nawyków pisarskich) i o wiele mniej zmian naturalnych niż w próbach naturalnych (Matuszewski, 2011); może tam również być widoczny brak płynności linii pisma (Bird, Found, Ballantyne, Rogers, 2010) i niejednorodność sposobu kreślenia znaków lub znaczące różnice w ich wymiarach (Matuszewski, 2011). Podobnie – w próbach naśladowanych możemy napotkać tremor, załamania i brak płynności linii pisma (Huber, Headrick, 1999; Matuszewski, 2011). Z kolei w próbach kopiowanych mogą być widoczne zatrzymania narzędzia pisarskiego, a także obecność linii wiodącej (Matuszewski, 2011), chociaż owa linia może być nieraz mylona z odciskiem linii tekstu powstałym w czasie podpisywania dokumentu leżącego na badanej próbce pisma (Ellen, Day, Davies, 2018).

Niestety, żaden z autorów nie zajął się opracowaniem kompleksowego porównania różnic występujących między wszystkimi 11 typami prób, najczęściej poprzestając na porównaniu próby naturalnej z jednym, innym typem próby (Harris, 1952–1953; Alford, 1970; Durina, 2005; Wendt, 2000; Leung, Cheng, Fung, Poon, 1993). Takie porównanie nie jest wystarczające, jeśli chcemy stwierdzić, czy dana próba pisma jest próbą naturalną czy też próbą innego typu. W przypadku gdy nie mamy możliwości porównać tej próby pisma z odpowiednim materiałem porównawczym (bo nim z pewnych względów nie dysponujemy), możemy polegać jedynie na cechach graficznych obecnych w danej próbce pisma. Wówczas byłoby przydatne opracowanie metody ustalania, na podstawie owych cech graficznych, typu danej próby pisma.

Prawdopodobieństwo tego, że, zgadując, prawidłowo określi się typ próby pisma (czyli tzw. klasyfikowania losowego), w przypadku gdy jest 11 możliwych typów próby, wynosić będzie 1/11 (około 9%). Jako że nie jest

znana żadna metoda klasyfikowania próby pisma według jej typu, to można jedynie zgadywać. Posiadanie 9% prawdopodobieństwa na poprawne określenie typu próby jest jednak ze względów praktycznych niewystarczające. Nawet jeśli zamiast 11 typów prób uwzględnili się jedynie 6 typów (wyłączając z rozważań 4 typy „auto” oraz bez rozróżniania od siebie dwóch rodzajów kopiowania), to trafność klasyfikowania losowego wzrosłaby do zaledwie 17% (1/6).

Z uwagi na powyższe opracowanie takiej metody jest konieczne. Ponadto metoda ta powinna umożliwić klasyfikowanie danej próby pisma według typu bez konieczności każdorazowego porównywania tej próby z materiałem porównawczym, a więc z wykorzystaniem do tego jedynie cech graficznych widocznych w tej próbce. Ze względu na potrzebę opracowania takiej metody przeprowadzono własne badania w tym zakresie.

Sformułowano w tym celu trzy główne hipotezy badawcze:

(H1) Niektóre typy prób pisma ręcznego posiadają pewne cechy charakterystyczne.

(H2) Cechy te umożliwiają odróżnienie od siebie owych typów prób.

(H3) Na podstawie cech charakterystycznych dla pewnych typów prób można opracować metodę klasyfikowania prób pisma ręcznego według ich typu.

Podsumowując: celem badawczym było znalezienie cech charakterystycznych dla 11 typów prób pisma ręcznego oraz ich wykorzystanie do opracowania skutecznej metody klasyfikowania próby pisma według jej typu. Przez „skuteczną” rozumiano tu „bardziej trafną niż klasyfikowanie losowe”. Aby tego dokonać, pewne cechy graficzne pobranych prób pisma zostały zliczone lub zmierzone, a następnie wykorzystane do klasyfikowania prób pisma według typu przy użyciu analizy dyskryminacyjnej i jej funkcji. Podczas walidacji opracowanej metody sprawdzono trafność opracowanych wcześniej funkcji dyskryminacyjnych na nowych próbach pisma. W wyniku badań opracowano dwie różne (z uwagi na liczbę uwzględnionych w nich cech graficznych, na podstawie których dokonywana jest klasyfikacja) metody klasyfikowania, nazywane dalej „klasyfikatorami”. Same klasyfikatory również zostały opracowane w dwóch wariantach w zależności od liczby możliwych typów branych pod uwagę w trakcie klasyfikowania (11 lub 6). W sumie daje to 4 różne metody klasyfikowania prób pisma według ich typu.

## 2. Materiał i metody

Główna część badań została przeprowadzona na próbach pisma: wszystkich 100 probantów (66 kobiet i 34 mężczyzn, studentów Wydziału Prawa i Administracji Uniwersytetu im. Adama Mickiewicza w Poznaniu)

poproszono o napisanie własnego imienia i nazwiska oraz fikcyjnego imienia i nazwiska – „Jędrzej Mickiewicz”. Każdy z uczestników badania miał także do napisania pojedyncze zdanie. Zdania zostały pobrane jako próby 11 typów, a imiona i nazwiska (zarówno własne probantów, jak i fikcyjne) jako próby 7 typów. Każdy z owych typów prób został opisany w tabeli 1, podczas gdy dokładne wyszczególnienie pobranych typów prób w przypadku każdego rodzaju próby (własnego imienia i nazwiska, fikcyjnego imienia i nazwiska oraz zdania) zostało uwzględnione w tabeli 2.

W drugiej części badania, walidacji opracowanych metod, wykorzystano próby pisma pochodzące od dodatkowych 39 osób (24 kobiet i 15 mężczyzn), co daje łącznie 139 probantów.

Pisemna zgoda na udział w badaniu została udzielona autorce przez wszystkich probantów; ich dane osobowe nie zostały w żaden sposób publicznie wykorzystane, gdyż badania prowadzono wyłącznie na próbach pisma i nie udostępniano ich nikomu.

We wszystkich pobranych próbach pisma zostały ocenione zarówno cechy motoryczne (Tabela 3), jak i mierzalne (zmiany nachylenia znaków, zmiany szerokości lub wielkości znaków). W przypadku cech motorycznych uwzględniano ich liczbę przypadającą na znak w danej próbce, podczas gdy cechy mierzalne były reprezentowane przez – odpowiednio – ich odchylenie standardowe w badanej próbce, współczynnik zmienności lub średnią arytmetyczną.

Owe cechy graficzne, a dokładniej: zaobserwowane różnice w ich liczbie czy rozmiarach, zostały przebadane celem znalezienia ich statystycznej zależności od typu próby i sprawdzenia, czy nadają się one do wykorzystania dla klasyfikowania prób pisma według typu.

Badań tych dokonano z wykorzystaniem programów komputerowych:

- Graphlog 2.20 (Cyborg Idea) – aby dokonać pomiarów cech mierzalnych (odchylenia standardowego, współczynnika zmienności i średniej arytmetycznej owych cech);
- Statistica 13.1 (Statsoft) – do obliczeń statystycznych (z wykorzystaniem analizy dyskryminacyjnej w celu obliczenia zaobserwowanych zależności między cechami graficznymi a typami prób oraz aby następnie zastosować je do klasyfikatorów dzięki zastosowaniu funkcji dyskryminacyjnych);
- Excel (Microsoft) – celem dokonania walidacji opracowanych metod.

Z uwagi na długość próby pisma największa liczba cech graficznych została przeanalizowana w zdaniach (23 cechy), podczas gdy w przypadku badania imion i nazwisk (zarówno probantów, jak i fikcyjnego) było to 10 cech (Tabela 4).

Ze względu na liczbę cech możliwych do analizy zostały opracowane dwie różne metody: pierwsza, bazująca

na najmniejszej liczbie cech, gdyż były one możliwe do zmierzenia we wszystkich rodzajach prób pisma (Klasyfikator 1) oraz druga, bazująca na 23 cechach graficznych, które były możliwe do zmierzenia lub policzenia w jednozdaniowej próbie pisma (Klasyfikator 2).

Obydwa klasyfikatory zostały też opracowane w dwóch wersjach:

- wersja „a” – do klasyfikowania 11 typów prób (naturalne, niezamierzenie nienaturalne – kreślone w pozycji stojącej, maskowane, autokopiuwane – w prześwicie i przez przeciskanie, kopiowane – w prześwicie i przez przeciskanie, autonaśladowane – ściśle i swobodnie, naśladowane – ściśle i swobodnie);
- wersja „b” – do klasyfikowania 6 typów prób (naturalne, maskowane, niezamierzenie nienaturalne, kopiowane – w prześwicie i przez przeciskanie, naśladowane – ściśle i swobodnie).

W niniejszym artykule opisano w sposób bardziej szczegółowy jedynie jedną, podstawową wersję klasyfikatora – Klasyfikator 1 (wersja „b”). Wszystkie funkcje dyskryminacyjne obydwu klasyfikatorów zostały opisane w załączniku 1.

W pierwszym kroku pobrano próbki pisma i zbadano ich cechy graficzne. Cechy te zostały następnie wykorzystane do obliczenia funkcji dyskryminacyjnych analizy dyskryminacyjnej (wielowymiarowa analiza porównawcza) w programie Statistica 13.1 (Statsoft), aby dokonać sklasyfikowania zebranych próbek według ich typu. Proces klasyfikowania był zdeterminowany wyborem najwyższej wartości funkcji dyskryminacyjnych. Każdy klasyfikator składa się z zestawu funkcji dyskryminacyjnych (liczba tych funkcji zależy od liczby możliwych do sklasyfikowania typów i w toku badań wynosiła odpowiednio 6 lub 11). Aby obliczyć wartość danej funkcji, należy podstawić liczbę obliczonych przez siebie cech graficznych danej próbki do wzoru.

Przykładowo: w jednym z wariantów Klasyfikatora 1 funkcja dyskryminacyjna dla próby naturalnej ( $K_N$ ) ma postać:

$$K_N = -0,50639 \cdot x_1 + 1,37615 \cdot x_2 + 1,81204 \cdot x_3 - 4,30809 \cdot x_4 + 0,32638 \cdot x_5 + 0,31590 \cdot x_6 + 8,70609 \cdot x_7 + 0,16095 \cdot x_8 + 0,25836 \cdot x_9 + 0,10630 \cdot x_{10} - 3,54610,$$

gdzie  $x_1, x_2, \dots, x_{10}$  oznaczają wartości kolejnych cech graficznych badanej próbki pisma.

Aby sprawdzić, czy dana próba pisma jest próbą naturalną, należy obliczyć wartości wszystkich funkcji klasyfikacyjnych (dla próby naturalnej, niezamierzenie nienaturalnej, maskowanej itp.) i sprawdzić, czy wartość dla klasyfikatora  $K_N$  jest najwyższa. Jeśli nie – badana próba pisma jest próbą innego typu, czyli tego, dla którego wartość funkcji była najwyższa.

### 3. Wyniki

Większość zbadanych typów prób posiada pewne cechy charakterystyczne (Tabela 5). Zaobserwowane zależności między występowaniem owych cech graficznych a typem próby zostały wykorzystane w klasyfikatorach. Trafność klasyfikowania podstawowej metody, nazwanej Klasyfikatorem 1, była około 2,5 razy wyższa niż klasyfikowanie losowe przy klasyfikowaniu 11 typów prób oraz około 2,17 razy wyższa przy klasyfikowaniu 6 typów prób (Tabela 6). W trakcie walidacji Klasyfikator 1 uzyskał trafność odpowiednio 2,67 i 2,5 razy wyższą od klasyfikowania losowego. Podobne wyniki uzyskał Klasyfikator 2. Każda z opracowanych 4 metod okazała się ponad dwukrotnie bardziej skuteczna niż klasyfikowanie losowe. Ponadto w trakcie walidacji metody te były przynajmniej tak samo skuteczne jak w pierwszym etapie badań.

Opracowane metody zostały również przetestowane (porównano ich skuteczność) dla różnych rodzajów badanych prób pisma (imienia i nazwiska probanta, fikcyjnego imienia i nazwiska oraz zdania); przykładowo – Klasyfikator 1 uzyskał najwyższą trafność, gdy został użyty do klasyfikowania zdań – około 47% (dla klasyfikowania 6 typów prób – Tabela 7). Podobne wyniki uzyskał Klasyfikator 2, jednakże w przypadku klasyfikowania zdań był on mniej skuteczny (około 37%). W tym wypadku jednak znowu każdy z klasyfikatorów okazał się ponad dwukrotnie bardziej skuteczny niż klasyfikowanie losowe.

### 4. Dyskusja wyników

Wyniki badań potwierdziły tezę o posiadaniu przez większość typów prób pewnych cech charakterystycznych, umożliwiających ich odróżnienie od pozostałych typów prób. Ponadto cechy te mogą również zostać wykorzystane do opracowania metody klasyfikowania prób pisma według ich typu.

Jeśli chodzi o cechy graficzne, to niektóre z nich wydają się silnie związane z danym typem próby, np. liczba zatrzymań jest zdecydowanie wyższa dla prób kopiowanych niż pozostałych typów prób. Niektóre z tych zależności zostały już wcześniej wspomniane w literaturze, jednak pełne ich zestawienie nie zostało dotąd opublikowane; niektóre z badanych cech nie zostały nigdy wcześniej uwzględnione przez badaczy. W toku badań wszystkie z badanych cech zostały uwzględnione w klasyfikatorach i okazały się pomocne dla klasyfikowania. Należy jednak mieć na uwadze fakt, że są to jedynie niektóre cechy graficzne możliwe do zbadania w próbkach. Z tego powodu należy w przyszłości przeprowadzić podobne badania również z uwzględnieniem innych cech



graficznych, gdyż mogłoby to poprawić skuteczność opracowanych klasyfikatorów.

Z uwagi na dotychczasowy brak jakichkolwiek obiektywnych metod klasyfikowania prób pisma według ich typu wyniki uzyskane w badaniach są satysfakcjonujące. Każda z metod okazała się przynajmniej dwa razy bardziej skuteczna niż klasyfikowanie losowe niezależnie od liczby branych pod uwagę cech graficznych. Klasyfikator 1 w głównej części badania oraz w trakcie walidacji uzyskał podobne wyniki i okazał się dość efektywny: pomimo uwzględnienia najmniejszej liczby cech graficznych (10 – w przeciwieństwie do 23 cech wykorzystywanych przez drugi klasyfikator) dokonywał klasyfikowania z trafnością podobnie dużą w przypadku imion i nazwisk, jak i całych zdań. Dodatkowo – im prostszy klasyfikator, tym jest bardziej użyteczny w praktyce – jest to szczególnie istotne w przypadku klasyfikowania krótkich prób pisma zawierających niewielką liczbę cech graficznych.

Należy podkreślić, że opracowane metody klasyfikowania prób pisma według ich typu stanowią jedynie wstęp do dalszych badań w tym kierunku. Po pierwsze – wykorzystanie w klasyfikatorach dodatkowych cech graficznych zapewne poprawiłoby ich skuteczność – na przykład w przypadku dłuższych prób pisma należałoby też rozważyć wykorzystanie cech mierzalnych. Po drugie – należy również poprawić metodę pobierania prób pisma do badań – zamiast pobierania wszystkich typów prób od każdego probanta być może należałoby się skupić na pobieraniu każdego typu od jak największej liczby probantów, aby zebrać jak najwięcej próbek pisma, a dodatkowo – uczynić całą procedurę sporządzania próbek przez probanta mniej czasochłonną. Wreszcie – sami probanci powinni stanowić bardziej zróżnicowaną grupę pod względem wieku, wykonywanego zawodu i, być może, pochodzenia geograficznego.

## 5. Wnioski

Podsumowując: wydaje się, że opracowane metody klasyfikowania prób pisma według ich typu posiadają potencjał naukowy, gdyż mimo wszystko okazały się ponad dwa razy bardziej skuteczne w klasyfikowaniu prób niż klasyfikowanie losowe. Najprostszą z metod, Klasyfikator 1, wydaje się najbardziej użyteczna, gdyż nie wymaga, by w badanej próbce było obecnych wiele cech graficznych (zatem może być wykorzystana zarówno dla długich, jak i krótkich prób pisma), a mimo to uzyskała wyniki równie dobre jak metody bardziej złożone.

Należy jednak podkreślić, że ostateczna efektywność wszystkich opracowanych metod nie pozwala na ich praktyczne zastosowanie – ich skuteczność wymaga poprawy. Z tego powodu dalsze badania w tej dziedzinie są

niezbędne – być może z wykorzystaniem dodatkowych cech graficznych oraz innego sposobu pobierania próbek.

### Notka od autorki

Niniejszy artykuł powstał na podstawie szerszego opracowania badawczego stanowiącego część rozprawy doktorskiej autorki („Klasyfikowanie pisma ręcznego według jego typu”, Poznań, 2021, niepublikowana).