A study into evaluating the location of fingermarks on letters given activity level propositions

de Ronde, Anouk; van Aken, Marja; de Poot, Christianne J.; de Puit, Marcel

DOI
10.1016/j.forsciint.2020.110443

Publication date
2020

Document Version
Final published version

Published in
Forensic Science International

Forensic Science International 315 (2020) 110443

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

A study into evaluating the location of fingermarks on letters given activity level propositions

Anouk de Ronde[a,b,c,*], Marja van Aken[b], Christianne J. de Poot[a,c], Marcel de Puit[b,d]

[a] Amsterdam University of Applied Sciences, Forensic Sciences, P.O. Box 1025, 1000 BA Amsterdam, the Netherlands
[b] Netherlands Forensic Institute, Digital Technology and Biometrics, P.O. Box 24044, 2490 AA The Hague, the Netherlands
[c] VU University Amsterdam, Criminology Department, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands
[d] Delft University of Technology, Faculty of Applied Sciences, Chemical Engineering, Van der Maasweg 9, 2629 HZ, Delft, the Netherlands

ARTICLE INFO

Article history:
Received 17 December 2019
Received in revised form 9 July 2020
Accepted 30 July 2020
Available online 2 August 2020

Keywords:
Fingerprints
Activity level
Document examination
Fingermark location

ABSTRACT

A previous paper published in this journal proposed a model for evaluating the location of fingermarks on two-dimensional items (de Ronde, van Aken, de Puit and de Poot (2019)). In this paper, we apply the proposed model to a dataset consisting of letters to test whether the activity of writing a letter can be distinguished from the alternative activity of reading a letter based on the location of the fingermarks on the letters. An experiment was conducted in which participants were asked to read a letter and write a letter as separate activities on A4- and A5-sized papers. The fingermarks on the letters were visualized, and the resulting images were transformed into grid representations. A binary classification model was used to classify the letters into the activities of reading and writing based on the location of the fingermarks in the grid representations. Furthermore, the limitations of the model were studied by testing the influence of the length of the letter, the right- or left-handedness of the donor and the size of the paper with an additional activity of folding the paper. The results show that the model can predict the activities of reading or writing a letter based on the fingermark locations on A4-sized letters of right-handed donors with 98 % accuracy. Additionally, the length of the written letter and the handedness of the donor did not influence the performance of the classification model. Changing the size of the letters and adding an activity of folding the paper after writing on it decreased the model's accuracy. Expanding the training set with part of this new set had a positive influence on the model's accuracy. The results demonstrate that the model proposed by de Ronde, van Aken, de Puit and de Poot (2019) can indeed be applied to other two-dimensional items on which the disputed activities would be expected to lead to different fingermark locations. Moreover, we show that the location of fingermarks on letters provides valuable information about the activity that is carried out.

1. Introduction

Focus on the activity that was carried out during the deposition of evidence has recently become an important aspect in the field of forensic science [1,2]. Establishing a link between the donor and the crime scene by determining the source of the trace is often not sufficient to determine what happened at the crime scene. Frequently, the question in court is about the activity that led to the deposition of the traces, which requires the use of activity level propositions instead of source level propositions [3]. For fingermark evidence, the evaluation of activity level propositions is a rather unexplored territory. However, recent research has shown that evaluating fingermarks given activity level propositions may add valuable information when one is reconstructing a crime [4].

An important variable for the evaluation of fingermarks at activity level is the location of the fingermarks on the object of interest. de Ronde, van Aken, de Puit and de Poot [5] presented a model for evaluating fingermark locations on pillowcases in relationship to the activity level questions of whether the pillowcase was used for smothering or was simply changed. The paper proposed that this model could be applied to all two-dimensional items for which it is expected that different activities result in different fingermark locations. An interesting application for this model is the evaluation of the location of fingermarks on handwritten letters since it might be expected that different activities—such as writing and reading—leave fingermarks on different locations and that the location of fingermarks on a letter can be used to determine what activity has taken place.

* Corresponding author at: Amsterdam University of Applied Sciences, Forensic Sciences, P.O. Box 1025, 1000 BA, Amsterdam, the Netherlands.
E-mail addresses: a.de.ronde2@hva.nl, a.de.ronde@nfi.nl (A. de Ronde).

http://dx.doi.org/10.1016/j.forsciint.2020.110443

Although examinations of handwritten documents seem less relevant as a forensic discipline in the digital world, a study into the demand for document examination showed that this may not be the case [6]. Besides cases of fraud or counterfeiting, handwritten document examination is still considered very important in counter-terrorism because terrorists appear to prefer to use handwritten texts to avoid digital traces. Handwritten document examination is also still considered relevant when the authenticity of suicide notes is questioned. An example of this is the case *R v. Stephen Port* [7], in which Port was convicted of four murders. In one of these murders, Port left a suicide note next to the victim in an attempt to divert suspicion. Another application of handwritten document examination is in cases involving illegal drugs. Evidence collected in these cases regularly includes handwritten notes describing the manufacturing steps for the synthesis of drugs[1] . For all these cases, it might be relevant to determine who wrote the notes or letters discovered at the crime scene. In cases regarding handwritten documents, a plausible alternative explanation for the presence of fingermarks on letters may be the activity of reading the letter instead of writing the letter.

The current approach for evaluating these types of questions about handwritten documents is to perform a handwriting examination [8]. We propose a complementary innovative approach: the evaluation of the location of the fingermarks on the letter.

This study investigates whether the model proposed by de Ronde, van Aken, de Puit and de Poot [5] to analyze the location of fingermarks could also be used to distinguish the activity of writing a letter from the alternative activity of reading a letter. For this purpose, we designed an experiment in which participants carried out two tasks: reading a preprinted letter and writing a letter. The fingermarks were visualized using conventional visualization techniques for fingermarks on paper. Afterwards, the binary classification model proposed by de Ronde, van Aken, de Puit and de Poot [5] was used to categorize the letters into the classes of writing and reading. In this study we have focussed only on the fingermarks visualised and not any palm marks that have potentially been left during writing, normally referred to as writers palm. This model is based on the distance between grid representations of the letters and classifies each grid into one of two classes that represent an activity by using quadratic discriminant analysis. The model was first trained using a training set consisting of written and read letters. The trained model was then used to predict the class of an unseen test set.

The previous study of this model on pillowcases had a few limitations. First, the objects in the training set were created by exactly the same protocol as the objects that were tested. Furthermore, for pillowcases, it was not deemed relevant to study the difference between left- and right-handed donors since the activities of smothering and changing were carried out using both hands. However, for written letters, the handedness of the donor may be an important factor. In this study, the limitations of the model were investigated by testing the influence of the length of the letter, the left- or right-handedness of the donor and the size of the paper with an additional activity of folding the paper on the model's performance.

2. Materials and methods experiment

2.1. Experimental design

The study is divided into two experiments. In the first experiment, we studied the possibility of differentiating between the two activities of writing and reading based on the fingermark locations present on A4-sized letters for right-handed donors. For this experiment, we used a dataset of 84 right-handed donors who wrote a letter of regular length on A4-sized paper and divided this set into a training set (70 %) and a test set (30 %) by random selection. The training set was used to train the classification model, and the unseen test set was used to study the performance of the model. We also tested the classification performance of the model when only the front side of the letter was used to determine the influence of the back side of the letter on the classification performance.

To study the limitations of the model for different variations of the letters, we conducted a second experiment in which the classification performance of the trained model based on A4-sized letters of regular length for right-handed donors was tested on three extra test sets:

a) a test set consisting of 13 right-handed donors who wrote a full-page letter;
b) a test set consisting of 12 left-handed donors, of whom two wrote a full-page letter; and
c) a test set consisting of 15 donors who used A5-sized paper and folded their letters after writing them.

2.2. Experimental protocol for A4-sized letters

A total of 110 students of the Amsterdam University of Applied Sciences read a letter on A4-sized paper and wrote a letter on A4-sized paper. The participants were first presented with a letter printed on one side of the paper that was placed on a table. The participants were asked to pick up the letter and read it. This letter was printed by a printer that was loaded by a person wearing gloves with clean, brand-new paper. Next, the participant was given a new, blank sheet of clean paper on which the participant was asked to write. Since it was observed that the letters written by the participants were mostly the length of half an A4-sized paper, we asked 15 participants to write a letter that was the length of a full A4-sized paper.

To visualize the fingermarks, the letters were treated with indanedione followed by ninhydrin. The results of one donor were excluded from the dataset due to heavy staining on a letter as a result of incorrect application of the visualization method. After each treatment, the letters were documented using a scanner and edited using Photoshop CS by cropping the images and adjusting the brightness for optimal contrast between the fingermarks and the background. The custom-made software tool Lexie translated the pictures into grid representations using a segmentation process, as described by de Ronde, van Aken, de Puit and de Poot [5] (supplementary material). A grid representation of 15×20 cells was used, which was found to be the optimal grid size.

2.3. Experimental protocol for A5-sized papers

To study the influence of the size of the paper on the performance of the model, an existing dataset consisting of grids representing A5-sized paper was used[2] . For this experiment, 15 participants were asked to perform three tasks: to read a letter printed on A5-sized paper, to write a threatening letter on A5 sized-paper and to write a love letter on A5-sized paper. The experimental protocol used for reading the letter was the same as

[1] Case example: Rb Gelderland 20 December 2018, ECLI:NL:RBGEL:2018:5606. Available via www.rechtspraak.nl, a database of randomly selected Dutch verdicts.

[2] For the A5-sized data, we have only used the grid data that were generated from this study.

that described in Section 2.2, whereas in the protocol for the writing scenario, an extra step of folding the paper was carried out by all participants after they finished writing. For the visualization of the fingermarks, the paper was treated with indanedione followed by ninhydrin and an additional treatment with physical developer. These letters were photographed instead of scanned, and the photographs were manually transformed into a grid representation of 15×20 cells.

2.4. Materials

For the A4-sized papers, clean regular white paper of the brand Canon Black Label Zero was used. For the A5-sized papers, clean, ruled paper of the brand Staples was used. For the development of the fingermarks, 1,2-indanedione, ninhydrin and physical developer were used. Indanedione solution was prepared by mixing 8 mL stock solution of $ZnCl_2$ with 100 mL of 1,2-indanedione stock solution (100 mL), which results in an IND-Zn solution (7,4% v/v). The stock solution of $ZnCl_2$ is prepared by adding 0.8 g $ZnCl_2$ to 10 mL EtOH, to which 1 mL ethyl acetate and 190 mL HFE 7100 was added. The stock solution of 1,2-indanedione is prepared by mixing 1.0 g 1,2-indanedione with 60 mL ethyl acetate, to which 10 mL acetic acid and 900 mL HFE 7100 are added and stirred for 20 min. The letters were immersed in the solution and air dried for 2 min. Ninhydrin solution was prepared by mixing 5 g of ninhydrin with 45 mL of ethanol, 2 mL of ethyl acetate and 5 mL acetic acid, to which 1 L of HFE7100 was added. The letters were immersed in the solution and air dried for 2 min. The A5-sized documents were additionally treated with the physical developer technique as described by Wilson, Cantu, Antonopoulos and Surrency [9]. All solutions were prepared freshly before use, from pre-weighed reagents except the silver nitrate. The application of the developer solution occurred on a slow shaking device in order to circumvent silver deposition on the bottom of the container. All the glassware was salinized before use, to prevent silver deposition on the slightly acidic surface of the glass. $ZnCl_2$ (>99 %), EtOH (absolute, > 99 %), Ethyl acetate (>98 %) were obtained from Sigma Aldrich (Zwijndrecht, NL). HFE 7100 was obtained from 3 M (Delft, NL). 1,2-indanedione (99 %) was obtained from BVDA (Haarlem, NL). Silver nitrate, maleic acid, iron nitrate monohydrate, ammonium iron sulfate hexahydrate and citric acid monohydrate were obtained from Merck & Co (Darmstadt, Germany). n-Dodecylamine acetate was obtained from ICN/Hicol (Aliso Viejo, CA) and Synperonic N from BDH/VWR (Amsterdam, the Netherlands).

3. Analysis

All analyses were conducted using the software R, a freely available software for statistical computing, version 0.99.896 [10].

3.1. Construction of the datasets

For the data pre-processing, we used the design shown in Fig. 1 for both the datasets of A4-sized papers and A5-sized papers. Each picture was transformed into a grid representation of 15×20 cells. In the grid representations, the presence of a fingermark in a cell is denoted by a 1 and the absence of a fingermark in a cell is denoted by a 0, resulting in a binary grid that represents the picture. Because the front side and the back side of each letter are considered dependent, we decided to concatenate the grids into a 30×20 grid representing one letter, of which the left side represents the front side of the letter and the right side represents the back side of the letter. The final datasets consisted of one concatenated grid for each scenario per donor.

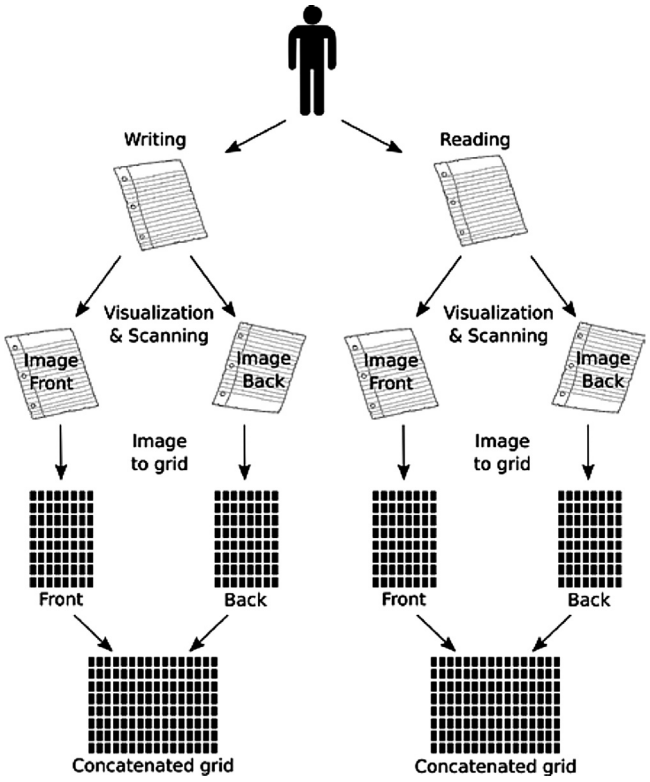

Fig. 1. Data construction of the grids representing the letters.

3.2. Visual analysis

In order to visualize the location of the fingermarks on the paper for the two scenarios reading and writing, we make use of heat maps. A heat map is a graphical representation, in which the distribution of fingermarks for all grids of one scenario is visually shown by the use of colors. From a heat map, the observed fingermark locations that are characteristic for each scenario can directly be observed.

3.3. Classification task

The purpose of the classification model we used is to assign the objects (letters) to a class (writing or reading) based on the location of the fingermarks on the letter. This is done by training the model with the use of a training set, for which for every letter is known to which class the letter belongs. The trained algorithm is then used to predict the class of letters in an unseen test set. The accuracy of the model is determined by comparing the model predictions of the test set to the known classes of the letters in the test set. Fig. 2 shows the structure of the datasets. In the first phase of testing whether we can differentiate between the two activities of writing and reading based on the location of the fingermarks, we used the training set consisting of 59 right-handed donors (denoted in blue in Fig. 2) to train the classification model. An unseen test set consisting of 25 right-handed donors (also denoted in blue in Fig. 2) was used to study the performance of the model. The limitations of the model were studied by testing test sets consisting of different variations of the letters to see the performance of the model trained on right-handed A4-sized letters of regular length on variations of this data, denoted by test sets A, B and C in Fig. 2.

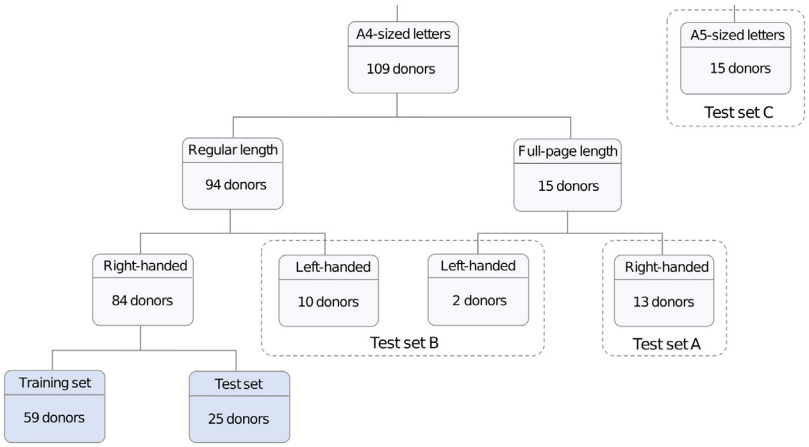

Fig. 2. Structure of the dataset.

3.4. Classification model

For the analysis, we used the classification model de Ronde, van Aken, de Puit and de Poot [5] proposed. This classification model is based on a similarity and distance measure between grids. For grids that belong to the same class is expected that there is a higher similarity between them than for grids that belong to a different class. The similarity between grids is represented by the similarity index (*SI*) of Sokal and Michener [11]:

$$SI = \frac{a+d}{n} \quad (1)$$

In which a represents the number of cells for which both grids contain a fingermark, d represents the number of cells for which both grids contain no fingermark and n represents the total number of cells. The *SI* is used to determine the Euclidean distance (d) between two grids, which can be expressed as:

$$d = \sqrt{1 - SI} \quad (2)$$

This distance measure is used to determine the distance of each grid to each of the grids in the training set consisting of writing letters and its distance to each of the girds in the training set consisting of reading letters. As a result, each grid can be represented as a feature vector $\binom{x_1}{x_2}$ where x_1 represents its mean distance to the training set of writing letters and x_2 represents its mean distance to the training set of reading letters. The classification is based on the expectation that a grid representing a writing letter has a lower distance to the training set consisting of writing letters compared to its distance to the training set consisting of reading letters, and vice versa. The feature vectors of all letters form a so-called feature space, which can be partitioned in classes with the use of a classification rule, for which we used Quadratic Discriminant Analysis (QDA). For a further explanation of QDA, we refer the reader to James, Witten, Hastie and Tibshirani [12].

3.5. Programming in R

For the implementation of the analysis in R, the following packages were used:

- *Raster* for all grid computations [13];
- *Ade4* to compute distance measures [14];
- *MASS* to perform QDA [15]; and
- *MVN* to test assumptions for QDA [16].
- *ggplot2* to produce the figures [17].

4. Results

4.1. Right-handed donors on A4-sized paper

Figs. 3 and 4 show the heat maps for the 59 right-handed donors in the training set for the scenarios of reading and writing, respectively. The heat maps show the concatenated grids of the front sides and the back sides of the letters. Fig. 3 shows that for the read letters, the fingermarks are mostly distributed around the left and right edges, on both sides of the paper. The heat map for the written letters in Fig. 4 shows that on the front side of the paper, the fingermarks are mostly distributed in an area on the middle top of the paper and along the left edge. The fingermarks on the middle top of the paper are caused by the placement of the right palm on the paper while writing. The fingermarks around the left edges on the front side of the paper are caused by holding the paper with the left hand. There were almost no fingermark observations on the back side of the paper.

4.2. The classification model

For each letter in the trainings set, its mean distances to the training set of written letters and to the training set of read letters are calculated. Fig. 5 shows the resulting feature space, in which the distance to the training set of written letters is plotted on the x-axis and the distance to the training set of read letters on the y-axis.

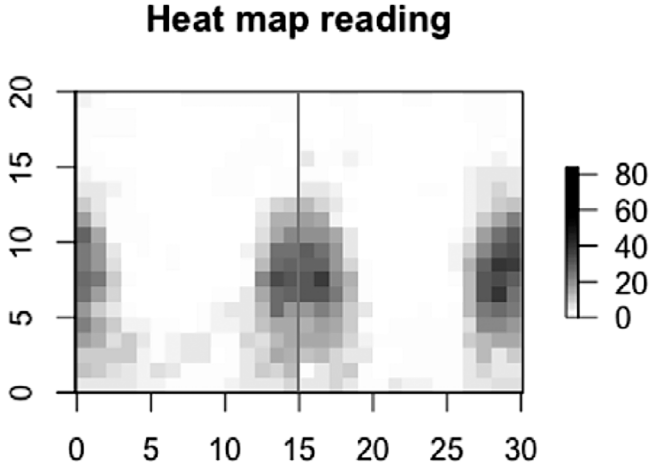

Fig. 3. Heat map for the training set of the reading scenario.

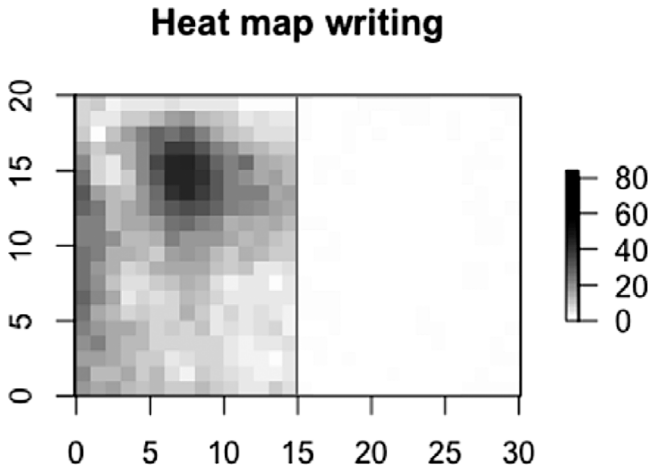

Fig. 4. Heat map for the training set of the writing scenario.

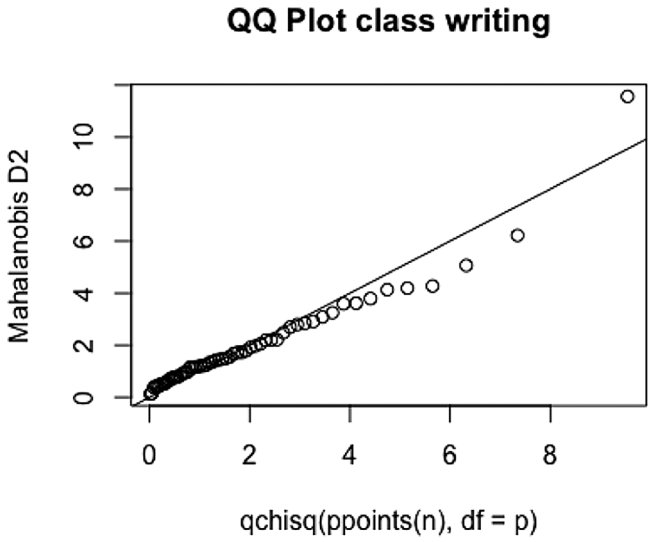

Fig. 6. QQ plot for the class of writing.

The red dots represent the read letters, and the blue triangles represent the written letters. Fig. 5 shows that the two classes of reading and writing form two reasonably separate regions, raising the expectation that a classification based on a QDA classifier as used in [5] may be appropriate for this dataset.

For the use of the QDA classifier, the assumption is that both classes follow a multivariate normal distribution. This hypothesis is tested with the use of the Mardia test and by studying QQ plots. The Mardia test is used to assess multivariate normality for the separate classes writing and reading based on the Mardia's multivariate skewness and kurtosis coefficients. For a further explanation of the Mardia test, we refer the reader to Kres [18]. The Mardia test result showed that the data were not multivariate normally distributed within the classes of writing and reading. Because multivariate outliers may be the reason for violation of the multivariate Gaussian assumption, we studied the QQ plot of each class, a widely used graphical approach to visually evaluate multivariate normality [16]. Using a QQ plot makes it possible to directly observe outliers that may cause a violation of the multivariate normality assumption. From the QQ plot shown in Fig. 6 for the class of writing, we observed that one outlier distorted the normality assumption. Aside from this outlier, the Mardia test shows that the data are indeed distributed following the multivariate Gaussian assumption. The QQ plot for the class of reading, shown in Fig. 7, shows three possible outliers. Aside from the most extreme outlier in the upper right corner, the Mardia test shows that the data are also distributed following the multivariate Gaussian assumption.

4.3. Evaluation of the model

Table 1 shows the confusion matrix for the QDA classification of the test set consisting of 25 right-handed donors writing a letter of regular length and reading a letter. The model classified 49 of the 50 letters correctly, representing an accuracy of 98.0 %. One read letter was misclassified as being a written letter. Fig. 8 shows a visual representation of the concatenated grid of the front side and the back side of this letter, indicating that the fingermarks on this letter are around the edges, as we would expect from the heat map for read letters, but additional fingermarks are found in the middle of the front of the paper, indicated by a black circle. We expect that these fingermarks in the middle of the paper caused the model to classify it as a written letter.

Since QDA classification is based on the posterior probabilities, the use of a QDA classifier allows for the calculation of a likelihood ratio for each object present in the test set using the formula $\frac{Pr(X=\boldsymbol{x} \mid Y=writing)}{Pr(X=\boldsymbol{x} \mid Y=reading)}$, in which $\boldsymbol{x}$ represents a feature vector of the corresponding letter. Fig. 9 shows the $\log_{10}$ likelihood ratio distributions for both classes of both the training set and the test set. The distributions for the classes of writing and reading are

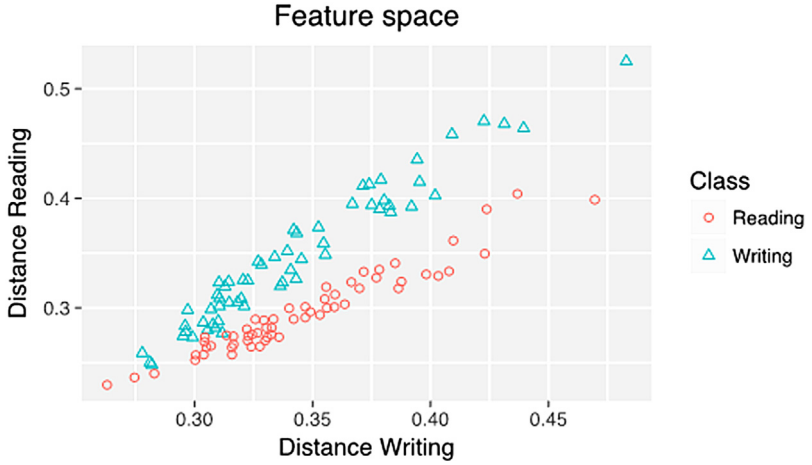

Fig. 5. Feature space for the training set consisting of right-handed donors.

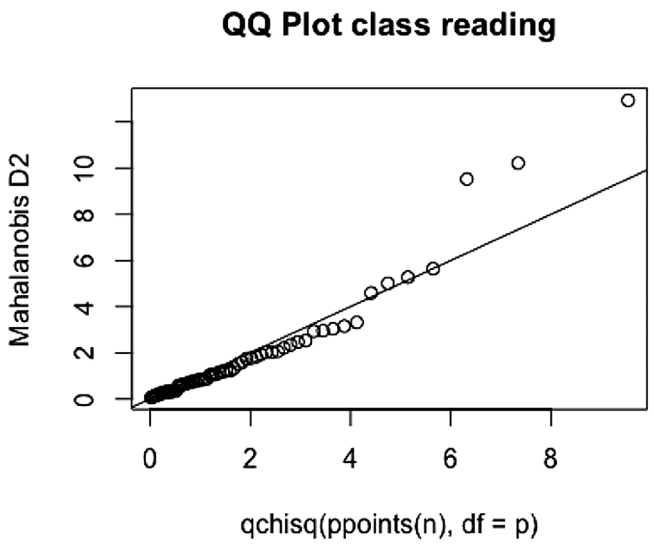

Fig. 7. QQ plot for the class of reading.

Table 1
Confusion matrix for the test set consisting of right-handed donors on A4-sized paper.

Test set	Reading	Writing
Reading predicted	24	0
Writing predicted	1	25

Fig. 8. Visual representation of the grid for the misclassified read letter.

quite well separated, although some letters obtain a relatively low likelihood ratio in favor of the wrong class. One of these is the letter shown in Fig. 8, and the other three letters were present in the training set on which the model is trained. From the distributions, we observe that the likelihood ratios reach extreme values. This will be further explained in the discussion.

4.4. Only the front side of the letter

Because the heat map for the writing scenario in Fig. 4 shows that there were almost no fingermark observations on the back side of the written letters, the question of whether the model only uses the empty back side of the letter as an indication for the class of writing or reading might arise. This would make the applicability of the model questionable if the activities slightly change such that the back side of the letter also contains fingermarks in the writing scenario. To account for this, we tested the performance of the model when only using the front side of the letters. The confusion matrix shown in Table 2 demonstrates that when using only the front side of the letters, the model classified 48 of the 50 letters correctly, an accuracy of 96 %. One additional read letter was misclassified as being a written letter. These results show that the model is able to classify the letters based on only the front side of the letter; however, the accuracy increases slightly when taking the dependency between the front and the back sides of the letters into account by concatenating both sides.

4.5. Full-page letters (test set A)

For the analysis of the full-page letters, a test set of 13 full-page letters was predicted by the classification model trained on the training set consisting of right-handed donors who wrote letters of regular length. Fig. 10 shows the heat map for the full-page read letters, and Fig. 11 shows the heat map for the written letters. The heat map for the read letters shows the same characteristics as the heat map for the training set shown in Fig. 3. The heat map for the written letters shows a somewhat different distribution of the fingermarks than the heat map for the training set shown in Fig. 4. The area on the middle top of the paper observed for the regular length letters is more spread over the front side of the letter. However, the heat maps show somewhat the same characteristics as the heat maps used for the training set, which leads to the expectation that this test set will be quite well predicted by the model.

Table 3 shows the confusion matrix for the test set. The results show that the activity of reading and the activity of writing were predicted correctly in all cases, although the heat map for the written letters looked slightly different. This is because the written letters are still quite different from the read letters. Whereas for the writing scenario, fingermarks are mostly observed in the middle of the paper and almost no fingermarks are observed on the back side of the paper, the fingermarks for the scenario of reading are still mostly placed along the edges of the paper on both sides of the paper. These results show that writing a full-page letter instead of a shorter letter on A4-sized paper does not influence the performance of the classification model.

4.6. Left-handed donors (test set B)

For the analysis of the letters of the left-handed donors, we used a test set consisting of 12 read and written letters, of which two donors wrote full-page letters. This test set was also predicted by the classification model trained on the training set consisting of right-handed donors who wrote letters of regular length. Since the results in Section 4.5 show that the length of the letter does not influence the performance of the model, these two full-page letters were also included in the left-handed test set. Figs. 12 and 13 show the heat maps for the left-handed donors for the classes of reading and writing, respectively. Fig. 12 shows that for the read letters, left-handed donors have a similar pattern as right-handed donors. Fig. 13 shows that for the written letters, the fingermarks of left-handed donors are distributed over the whole page, while for right-handed donors, the fingermarks were mostly distributed in an area on the middle top of the letter and along the left edge. Since the heat maps for the left-handed donors show somewhat the same characteristics as the heat maps for the full-page letters and the full-page letters were all correctly predicted, we expect that the model will also be able to predict the correct class of most of the left-handed donors.

Table 4 shows the confusion matrix for the test set consisting of left-handed donors. The results show that all read letters and

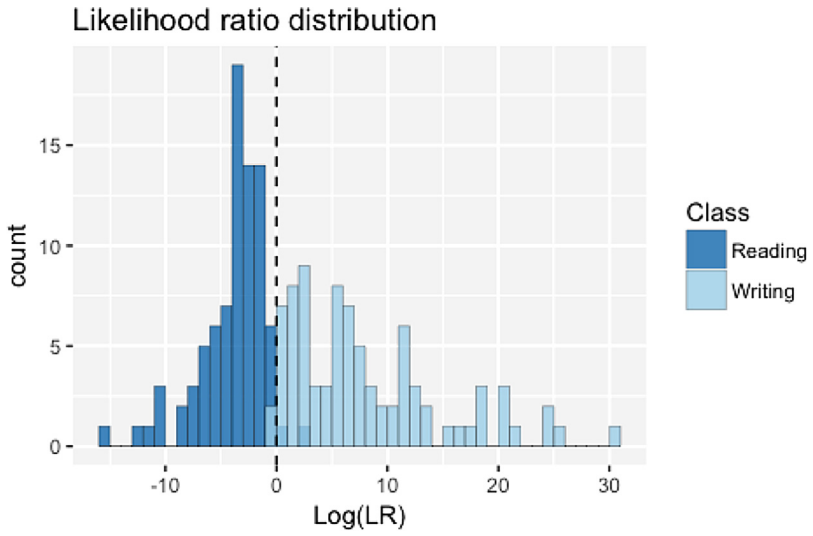

Fig. 9. Likelihood ratio distribution for the complete dataset.

Table 2
Confusion matrix for the test set consisting of right-handed donors on A4-sized paper using only the front side of the paper.

Test set	Reading	Writing
Reading predicted	23	0
Writing predicted	2	25

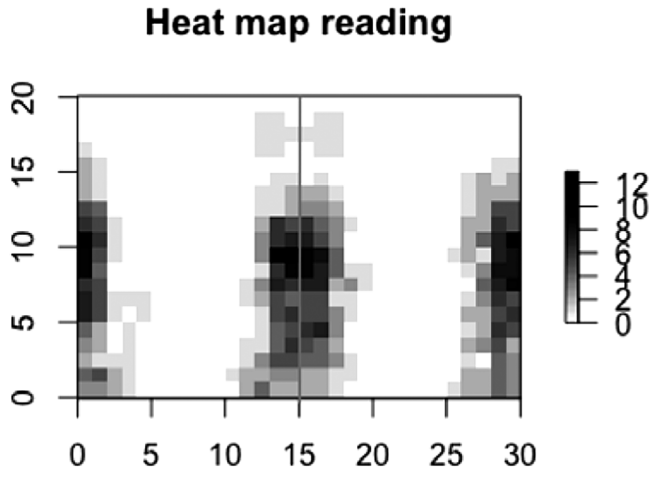

Fig. 10. Heat map of read letters for test set A consisting of full-page letters.

written letters were predicted correctly. Apparently, training the model with a dataset consisting of right-handed letters does not affect the classification of the left-handed letters, although the fingermark patterns differ for the writing scenario.

4.7. A5-sized letters (test set C)

For the analysis of the size of the letters, a test set consisting of 15 read letters and 30 written letters was also predicted by the classification model trained on the training set consisting of right-handed donors who wrote letters of regular length. Figs. 14 and 15 show the heat maps for these A5-sized letters for the scenario of reading and the scenario of writing, respectively. Fig. 14 shows for the A5-sized read letters, the fingermarks are mostly distributed along the edges on both sides of the paper, as we also observed for the A4-sized read letters. Additionally, some donors placed their hands around the bottom of the paper, which was also observed for the A4-sized read letters in Fig. 3. The heat map for the A5-sized written letters in Fig. 15 shows that the distribution of the fingermarks is clearly different from the distribution we observed for the A4-sized written letters in Fig. 4, for which we observed that on the front side of the paper, the fingermarks are mostly distributed on the middle top of the letter and along the left edge. For the A5-sized written letters, we observe that this area has shifted to the middle bottom of the paper and is concentrated on the entire width of the paper, and almost no fingermarks are found in the middle top area of the letter. An explanation for this may be that the palm is placed lower on the paper since the paper is smaller. Furthermore, the fingermarks around the edges caused by holding the paper with the other hand may interfere with the palm placement because the paper is narrower, so the areas almost overlap. The fingermarks on the back side of the written letters can be explained by the additional activity of folding the paper before it was put back on the table. This also differs from the heat map observed for the A4-sized written letters, since almost no fingermarks were found on the back side of the paper.

For the classification, we tested a test set consisting of all 15 read letters and all 30 written letters (love letters and threatening letters). The confusion matrix in Table 5 shows that the model had an accuracy 645%. All 15 read letters were predicted correctly, but the model had difficulty classifying the written letters. One explanation for the model's poor classification accuracy for the A5-sized letters might be the influence of the additional post-activity of folding the paper after writing on it. Since we expect that folding the paper mostly affects fingermarks to be present on the backside of the letter, the classification was repeated with only using the front sides of the letters. Table 6 shows the classification results. Although the model accuracy increased to 75.6 %, the model still wrongly predicted 11 of the writing letters. A possible explanation for this will be further explained in the discussion.

One way to achieve higher accuracy for A5-sized letters may be to expand the training set consisting of A4-sized letters by adding A5-sized letters to train the model for A5-sized letters as well. For this analysis, 70 % of the first 15 donors who read and wrote a love letter on A5-sized paper were added to the training set (11 donors). The remaining 30 % of the donors represent the test set (4 donors), together with the extra 15 threatening letters written by the donors. For this, we assumed that there is no difference in

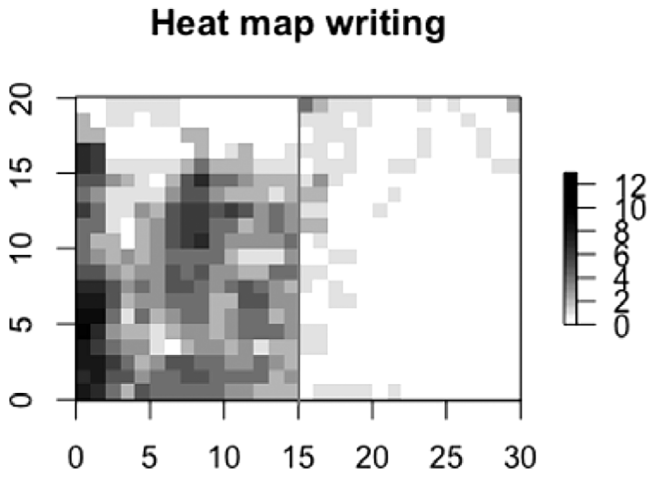

Fig. 11. Heat map of written letters for test set A consisting of full-page letters.

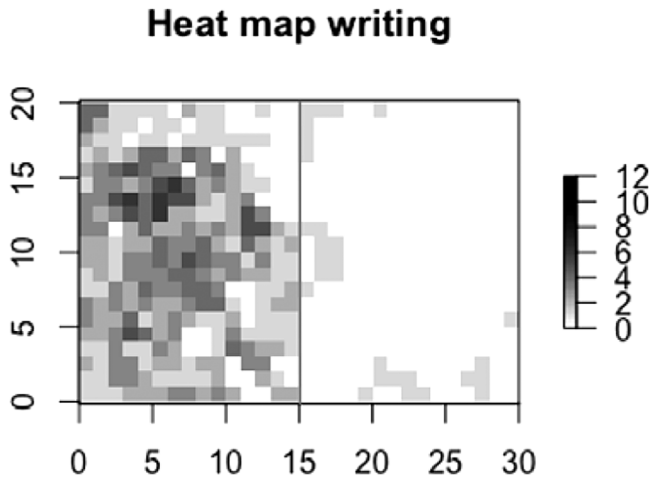

Fig. 13. Heat map of written letters for test set B consisting of letters by left-handed donors.

Table 3
Confusion matrix for test set A consisting of full-page letters.

Test set	Reading	Writing
Reading predicted	13	0
Writing predicted	0	13

Table 4
Confusion matrix for test set B consisting of letters by left-handed donors.

Test set	Reading	Writing
Reading predicted	12	0
Writing predicted	0	12

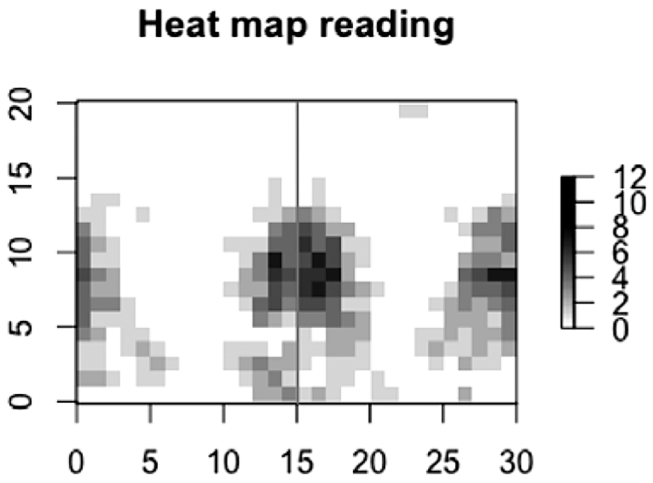

Fig. 12. Heat map of read letters for test set B consisting of letters by left-handed donors.

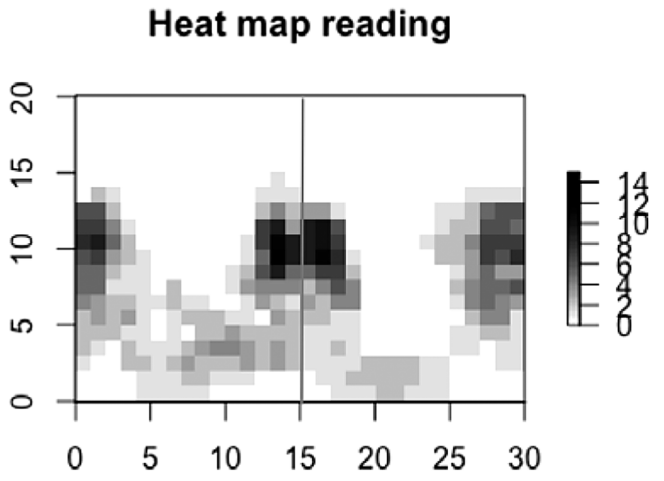

Fig. 14. Heat map of read letters for test set C consisting of A5-sized letters.

fingermark deposition between the type of message (love or threatening) that is written. The new training set was used to train the model, and afterward, the performance of the model was tested on the unseen test set. Table 7 shows the confusion matrix, which indicates that five written letters are wrongly classified as read letters, resulting in an accuracy of 78.3 %, which is significantly increased compared to the accuracy of 64.4 % obtained for a training set consisting of only A4-sized letters.

5. Discussion and conclusion

This research studied whether the model for the activity level analysis of the location of fingermarks proposed by de Ronde, van Aken, de Puit and de Poot [5] could also be used on letters to distinguish the activity of writing from the alternative activity of reading. The results have shown that the model could very well be applied to fingermarks on letters of right-handed donors to differentiate between the two activities, with a classification accuracy of 98.0 %. Furthermore, we showed that the length of the written letter and the handedness of the donor did not influence the performance of the classification model. For letters on a smaller sized paper (A5) and with an additional activity of folding the paper after writing on it, the model accuracy decreased to 64.4 %. If the training set consisting of A4-sized letters used to train the model is expanded with A5-sized letters, the model accuracy increases to 78.3 %. These results show that the location of fingermarks on letters provides valuable information about the activity that was carried out.

Despite the fact that the heat map for the written letters of the left-handed donors showed significant differences from the heat map of written letters of the right-handed donors, all letters written by left-handed donors were correctly predicted by the

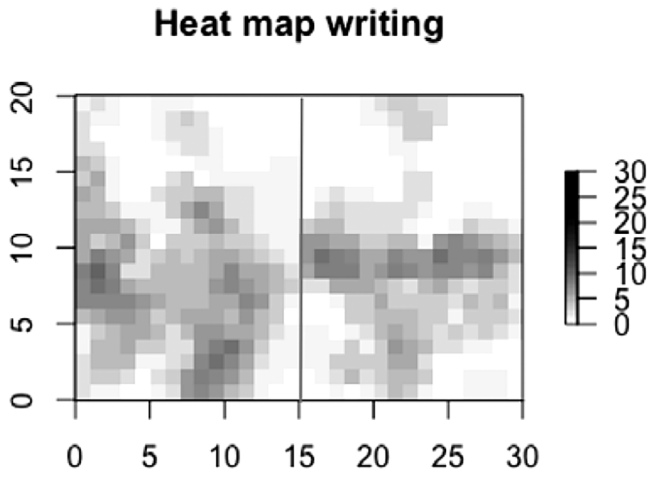

Fig. 15. Heat map of written letters for test set C consisting of A5-sized letters.

Table 5
Confusion matrix for classifying test set C consisting of A5-sized letters based on a training set of A4-sized letters.

Test set A5 size	Reading	Writing
Reading predicted	15	16
Writing predicted	0	14

Table 6
Confusion matrix for classifying test set C consisting of A5-sized letters based on a training set of A4-sized letters, when using the front side of the letters.

Test set A5 size	Reading	Writing
Reading predicted	15	11
Writing predicted	0	19

Table 7
Confusion matrix for classifying a test set consisting of A5-sized letters based on a training set of A4- and A5-sized letters.

Test set A5 size	Reading	Writing
Reading predicted	4	5
Writing predicted	0	14

model trained on right-handed donors. The difference in fingermark patterns for the scenario of writing between left- and right-handed donors may be caused by the variation in hand placement that was observed for left-handed people when they were writing letters. The reason that the classification was not affected by these different fingermark patterns is because the grids that represent the written letters of the left-handed donors still have a distinctive pattern from that of the read letters. However, care should be taken when testing left-handed donors on a right-handed–trained model. Since we tested only a small sample of left-handed donors (12 donors), the possibility exists that not all variations of left-handed writing are incorporated in our dataset, and variations that are not represented may be classified incorrectly. To correct for this, a larger sample of left-handed donors should be tested.

The model trained on A4-sized letters wrongly predicted more than half of the written A5-sized letters. There can be two explanations: the difference in activity that is carried out and the difference in the size of the paper. An additional activity of folding the paper was carried out by the participants in the experiment with A5-sized paper, causing the appearance of fingermarks on the back side of the paper in the writing scenario. Since the results for testing only the front side of the letters for the A5-sized papers have shown that still 36.7 % of the written letters are wrongly predicted by the model, this extra activity of folding the paper does not explain the poor classification results on itself and we expect that the difference in the size of the paper between the training set (A4) and the test set (A5) is an important factor to consider. Since the model is constructed such that the training set and the test set have to contain grids of similar dimensions, the number of cells is the same for both sizes of letters (15×20), but the size of the cells differs between the grids for the A4-sized letters ($1.5\text{cm} \times 1.5\text{cm}$) and the grids of the A5-sized letters ($1\text{cm} \times 1\text{cm}$). However, the sizes of the fingermarks do not change when using a smaller paper, so one fingermark may fill more cells in the grid representing A5-sized paper than it does in the grid representing A4-sized paper. This means that if the size of the objects present in the training set significantly changes from the size of the object being tested, the training set will probably not be representative of the test set. One solution may be to expand the dataset with new data, as we have shown for the A4- and A5-sized letters. Another may be to not work with squared cells but to choose larger areas on the letters that are representative for the activities of reading and writing and to standardize different sizes of paper to this representation. This may be a topic for further research. For now, we propose expanding the training set so that the dimensions of the object to be tested are also represented.

The likelihood ratio values that were provided as output from our model are in a higher and lower order then expected, given the size of our dataset. Since the assumptions for the use of QDA we have made are based on a limited dataset, we have no proof of the applicability of QDA beyond our dataset, which means that the likelihood ratios provided by the system may be sensitive to extrapolation errors [19]. A solution for this is to calibrate the likelihood ratio system that results from the model. There are several methods for performing this calibration [20]. Further research is needed to determine which calibration method is most suitable for our dataset to obtain likelihood ratio values that can be directly applied to casework.

In this research, the source level information of the fingermarks is not taken into account. This means that the model is not only based on identifiable fingermarks present on the letters, but also on additional stains such as smears that were visualized. We decided to not work only with identifiable fingermarks since smears and stains are also a direct result of the activity. For example, a smear created by the placement of the palm on the paper during writing may not result in a fingermark suitable for identification. However, this smear provides information about the placement of the hand during the activity. A drawback to this is that care should be taken when using this model on visualized fingermarks: if the fingermark visualization method is not correctly applied, causing the appearance of drops or spots on the object of interest, these drops and spots will also be interpreted as marks.

In this experiment, we exclusively tested the activities of writing and reading a letter for the training set used, without testing any pre- or post-activities such as grabbing the paper or folding the paper. As a consequence of this, if this dataset is applied in casework, it is of great importance to clearly state the activity hypotheses tested, to know exactly what activities are at stake. As we have shown, any additional pre- or post-activities may slightly influence the performance of the model; adding an extra step of folding the paper may influence the performance of the model if the model is trained based on a training set that does not involve this extra folding step. Thus, when applying this dataset to casework, it should be considered whether the training set should be expanded with appropriate examples of additional

activities if any extra activities were carried out in the particular case.

Another factor to take into account when applying the generated data to casework is that in this study, we clearly separated the activities of writing and reading. In real casework, this may not always be expected and these activities could have occurred successively. However, by studying these activities separately, we have shown that both activities cause a distinctive fingermark pattern on the letter. The heat maps show particular areas on the letter that are representative of writing traces or reading traces, making it possible to select the traces on a letter that are specific for the activity of writing or for the activity of reading. In this way, the investigation can focus on the marks that provide an indication of a certain activity, and if no identifiable fingermarks are found, a targeted sampling for DNA is possible.

The focus of this study was to distinguish the activity of writing a letter from the alternative activity of reading a letter, based on the variable location of the fingermarks. As discussed by de Ronde, Kokshoorn, de Poot and de Puit [4], there are several other variables that may be of interest when evaluating fingermarks given activity level propositions. The data from the conducted experiment shows that the presence of a large area on the front side of the letter, caused by placing the palm on the paper while writing a letter, is probably very distinctive between the disputed activities writing and reading, raising the suspicion that the presence of a palm print on the front side of the paper may provide valuable information on the activity that was carried out with the paper. The variable area of friction ridge skin that left the fingermark may be an interesting variable for further research into fingermarks given activity level propositions on letters.

With this research, we have confirmed that the model proposed [5] could very well be applied to any two-dimensional item for which it is expected that different activities lead to different fingermark locations. Instead of using paint to directly visualize the fingermarks as was done in the previous study on pillowcases [5], conventional techniques to visualize fingermarks on paper were used, resulting in traces that represent fingermark traces that would be obtained in real casework. We now have access to a database consisting of written and read letters on A4-sized paper and A5-sized paper that represents the separate activities of reading and writing. Now that we have shown that the model is very well able to distinguish between the activities reading and writing, the next step for implementation to casework will be to perform further research into more realistic scenarios such as pre- and post-activities or carrying out reading and writing successively by performing a pseudo-operational trial on letters that were not collected under lab conditions to see how the model performs on more realistic casework materials.

CRediT authorship contribution statement

Anouk de Ronde: Conceptualization, Methodology, Software, Validation, Data curation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Marja van Aken:** Methodology, Software, Validation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Christianne J. de Poot:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Marcel de Puit:** Conceptualization, Methodology, Writing - review & editing, Supervision.

References

[1] S. Willis, L. McKenna, S. McDermott, G. O'Donell, A. Barrett, B. Rasmusson, A. Nordgaard, C. Berger, M. Sjerps, J. Lucena-Molina, ENFSI Guideline for Evaluative Reporting in Forensic Science, European Network of Forensic Science Institutes, 2015.
[2] A. Biedermann, C. Champod, G. Jackson, P. Gill, D. Taylor, J. Butler, N. Morling, T. Hicks, J. Vuille, F. Taroni, Evaluation of Forensic DNA Traces When Propositions of Interest Relate to Activities: Analysis and Discussion of Recurrent Concerns, Front. Genet. 7 (2016) 215.
[3] R. Cook, I.W. Evett, G. Jackson, P. Jones, J. Lambert, A hierarchy of propositions: deciding which level to address in casework, Sci. Justice 38 (1998) 231–239.
[4] A. de Ronde, B. Kokshoorn, C.J. de Poot, M. de Puit, The evaluation of fingermarks given activity level propositions, Forensic Sci. Int. 302 (2019).
[5] A. de Ronde, M. van Aken, M. de Puit, C. de Poot, A study into fingermarks at activity level on pillowcases, Forensic Sci. Int. 295 (2019) 113–120.
[6] S. Flight, C. Wiebes, Handschriftonderzoek in het kader van terrorismebestrijding, WODC, 2016.
[7] Rv. Stephen Port, Sentencing Remarks of Mr Justice Openshaw, Judiciary of England and Wales, Central Criminal Court, 2016.
[8] R.N. Morris, Forensic Handwriting Identification: Fundamental concepts and principles, Academic press, 2000.
[9] J.D. Wilson, A.A. Cantu, G. Antonopoulos, M.J. Surrency, Examination of the steps leading up to the physical developer process for developing fingerprints, J. Forensic Sci. 52 (2) (2007) 320–329.
[10] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016.
[11] R.R. Sokal, C.D. Michener, A Statistical Method for Evaluating Systematic Relationships, Univ. Kansas Sci. Bull. 38 (22) (1958).
[12] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer, New York, 2013.
[13] R.J. Hijmans, Raster: Geographic Data Analysis and Modeling, (2016) .
[14] S. Dray, A.B. Dufour, The ade4 package: implementing the duality diagram for ecologists, J. Stat. Softw. 22 (4) (2007) 1–20.
[15] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, 4th ed., Springer, New York, 2002.
[16] S. Korkmaz, D. Goksuluk, G. Zararsiz, MVN: An R Package for Assessing Multivariate Normality, R J. 6 (2) (2014) 151–162.
[17] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer, 2016.
[18] H. Kres, The Mardia-Test for Multivariate Normality, Skewness, and Kurtosis Tables by, in: K.V. Mardia (Ed.), Statistical Tables for Multivariate Analysis, Springer, New York, NY, 1983.
[19] P. Vergeer, A. van Es, A. de Jongh, I. Alberink, R. Stoel, Numerical likelihood ratios outputted by LR systems are often based on extrapolation: When to stop extrapolating? Sci. Justice 56 (6) (2016) 482–491.
[20] G.S. Morrison, N. Poh, Avoiding overstating the strength of forensic evidence: Shrunk likelihood ratios/Bayes factors, Sci. Justice 58 (2018) 200–218.