Forensic Science International 360 (2024) 112069

Contents lists available at ScienceDirect

Forensic Science International

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

Use of automated quality assessment algorithms in fingermark detection research – Application to IND/Zn vs DFO

Bérénice Bonnaz *, Christophe Champod, Andy Bécue *

[1] University of Lausanne, Lausanne-Dorigny 1015, Switzerland
[2] École des Sciences Criminelles (School of Criminal Justice), Faculty of Law, Criminal Justice and Public Administration, Lausanne-Dorigny 1015, Switzerland

ARTICLE INFO

Keywords:
Fingerprint
Clarity
Quality assessment
Algorithms
Latent Quality Metric (LQM)
Improved Latent Fingerprint Quality Metric (ILFQM)

ABSTRACT

When developing detection techniques for fingermarks, the detected fingermarks must be evaluated for their quality to assess the effectiveness of the new method. It is a common practice to compare the performance of the new (optimized) technique with the traditional or well-established ones. In current practice, this evaluation step is carried out by a group of human assessors. A new approach is applied in this paper and consists of using algorithms to perform this task. To implement this approach, the comparison between IND/Zn and DFO has been chosen because it has already been the subject of many articles published in recent years and a consensus exists on the superiority of IND/Zn over DFO. The quality of 3'600 fingermarks developed using both detection techniques was assessed automatically using two algorithms: LQM (Latent Quality Metric) and ILFQM (Improved Latent Fingerprint Quality Metric). The distribution of quality scores was studied for both detection techniques. The results showed that fingermarks detected with IND/Zn received higher scores on average than fingermarks detected with DFO, which is in line with the consensus in the literature based on human assessment. The results of this research are promising and shows that automated fingermark quality assessment is an efficient and viable way to comparatively assess fingermark detection techniques.

1. Introduction

Assessing fingermark detection techniques is a multi-step process which usually includes: (1) the collection of a set of fingermarks, (2) their processing using one or several detection techniques, (3) their recording using the most adequate optical methods, and (4) the assessment of their quality so that conclusions can be drawn regarding the performance of the investigated technique(s). Each step must be carefully designed to ensure the validity of the methodology and of the conclusions reached. Guidelines do exist and help researchers setting some of the major parameters (e.g., number of donors, number of substrates, secretion types, deposition strategy, depletion series, split marks) [1–3].

With regards to the quality assessment of the detected marks, the current practice consists in asking a group of examiners to review all the marks (or pairs of marks) and assign a score to each of them using a provided scale [4]. This way of doing has proven its reliability, despite some anticipated variations between examiners [5]. The ease of implementation of such an approach explains why it is widely used in the field. Its simplicity is however hindered by two major limitations: (1) the choice of the quality scale among dozens of existing ones and (2) the workload for the examiners when they are requested to grade several hundreds to thousands of marks. With regards to the first limitation, an extensive review recently showed the existence of dozens of quality scales, with a couple of well-accepted ones that are sometimes adapted by researchers to fit their study [4]. The authors in [4] also identified a lack of agreement regarding the parameters that examiners must consider when grading the fingermarks quality (e.g., levels of ridge details, ridge visibility/continuity, background development, contrast, clarity of image – to cite a few). As a result, researchers may struggle to choose a scale fitting their study. It also leads to a lack of readability of the current good practices related to the quality assessment of a set of fingermarks. With regards to the data management, Hockey *et al.* [6] emphasized that ordinal scales are often applied in an incorrect manner by researchers, leading to misguiding conclusions. With regards to the second limitation, the workload is a direct consequence of the recommended good practices in fingermark detection research. Referring to the IFRG guidelines [3], proof-of-concept studies (Phase 1) can rely on hundreds of marks, whereas optimization and validation studies (Phases 2–4) could quickly lead to the consideration of thousands of marks. For

* Corresponding authors.
E-mail addresses: berenice.bonnaz@unil.ch (B. Bonnaz), andy.becue@unil.ch (A. Bécue).

https://doi.org/10.1016/j.forsciint.2024.112069
Received 9 January 2024; Received in revised form 2 May 2024; Accepted 23 May 2024
Available online 27 May 2024

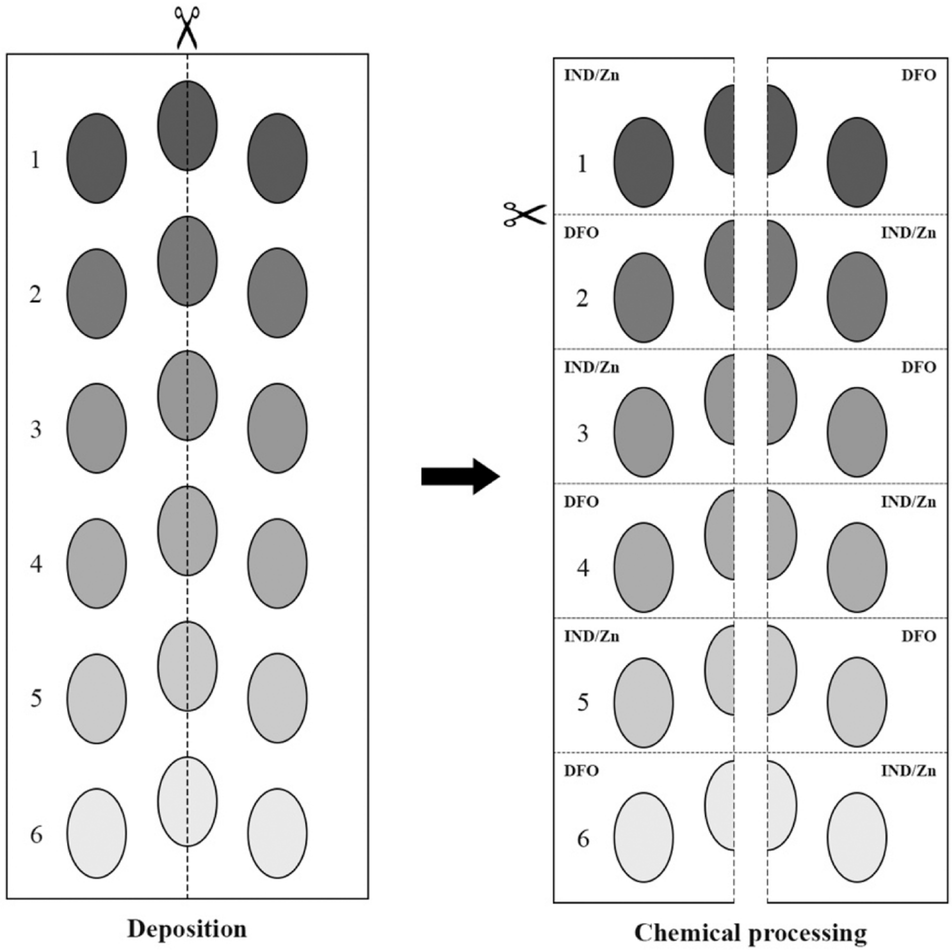

Fig. 1. Illustration of the depletion series distribution on a given substrate and of the association strategy of the fingermarks to IND/Zn or DFO.

Table 1
Recipes used to prepare the IND/Zn and DFO solutions.

IND/Zn [31]	DFO [Internal recipe]
Working solution	
0.125 g 1,2-indanedione (BVDA)	0.10 g 1,8-diazafluoren-9-one (BVDA)
50 ml ethyl acetate (Sigma-Aldrich, 99.5%)	25 ml dichloromethane (Sigma-Aldrich, ≥ 99.9%)
50 ml methanol (Sigma-Aldrich, ≥ 99.8%)	25 ml methanol (Sigma-Aldrich, ≥ 99.8%)
5 ml acetic acid (Sigma-Aldrich, ≥ 99%)	10 ml acetic acid (Sigma-Aldrich, 99%)
400 ml petroleum ether (Sigma-Aldrich)	440 ml petroleum ether (Sigma-Aldrich)
10 ml $ZnCl_2$ solution	
$ZnCl_2$ solution	
0.2 g $ZnCl_2$ (Sigma-Aldrich)	
100 ml ethanol (Sigma-Aldrich, ~ 96%)	

example, Luscombe *et al.* [7] used 7'500 split marks to validate an amino acid reagent. Chadwick *et al.* [8] took advantage of 14'000 fingermarks that were graded by examiners to better understand the mechanisms related to fingermark deposition. The current practice, based on human examiners is time-consuming, prone to variation among vetters, and tedious due to the repetitive nature of the process [5,9,10].

A new approach has been recently proposed to address these limitations: using the quality-assessment metrics integrated or not into automated comparison systems (e.g., AFIS) to grade the collected marks [11]. In their proof-of-concept study, Bonnaz *et al.* [11] compared the metrics obtained from several algorithms (i.e., LO (a proprietary quality measure for fingermarks part of a commercial AFIS system), LFIQ1 and LFIQ2 [12,13], LQM [14,15], ESLR [16], NFIQ and MINDTCT [17]) with the grading scores provided by a group of examiners. The LQM algorithm, designed to mimic human vision [14], stood out by offering a significant correlation with the human scoring. These results demonstrated that quality measure algorithms could constitute a valid alternative to human assessment, by being quicker, independent of the assessors, and requiring minimal human efforts (i.e., coding and data management).

The present study aims at investigating further the use of quality metrics in the fingermark detection research to assess large sets of marks. The selected scenario is the one that gathered a lot of research attention in the last years: the comparison between 1,8-diazafluorenone (DFO) and 1,2-indanedione/zinc (IND/Zn). Both amino acid reagents, the first one was introduced in 1990 [18,19] while the second one was optimized in the early 2000s [20–22]. Both reagents develop pinkish fingermarks that must be observed in photoluminescence. Since the proposition of an optimized formulation of IND/Zn in 2007, numerous publications aimed at assessing the relative performance of both reagents in controlled and (pseudo-)operational studies [7,23–30]. Overall, the conclusions clearly favor IND/Zn over DFO, in terms of numbers of detected marks, mark quality and brightness/luminescence. Pseudo-operational trials showed that IND/Zn detected more marks than DFO: +34% to +43% [23], +20% [24], +70% [7]. Bicknell and Ramotowski [28] showed that IND/Zn detected additional marks (i.e., not detected by DFO) on 72.4% of the processed samples. Luscombe *et al.* [7] also concluded that IND/Zn outperformed DFO on all the

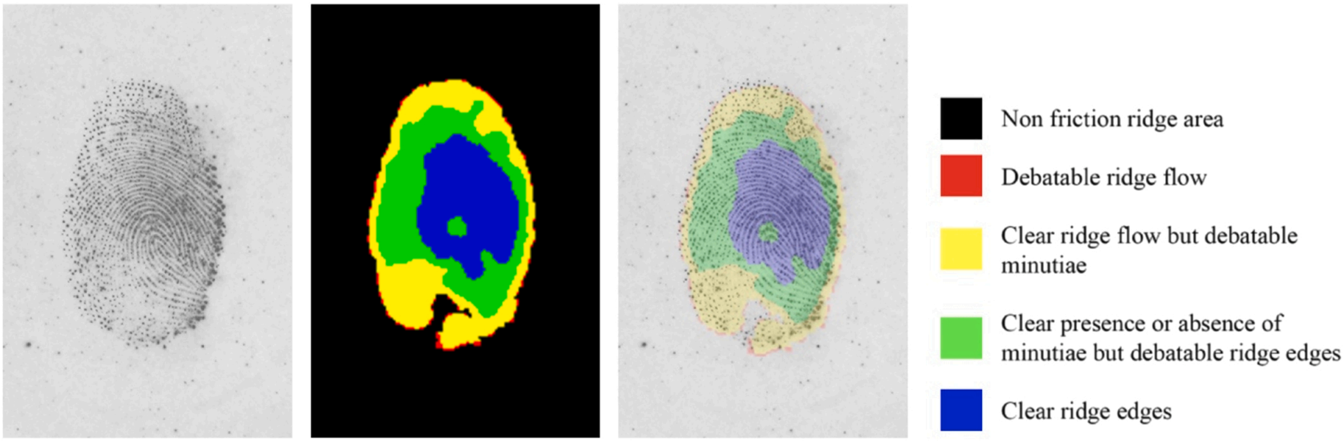

Fig. 2. Illustration of a clarity map (middle) generated by LQM when assessing a fingermark image (left). In this depiction, the levels of clarity range from red (lowest) to blue (highest) [15]. To better image the matching between the clarity areas and the fingermark, both representations have been superimposed (right).

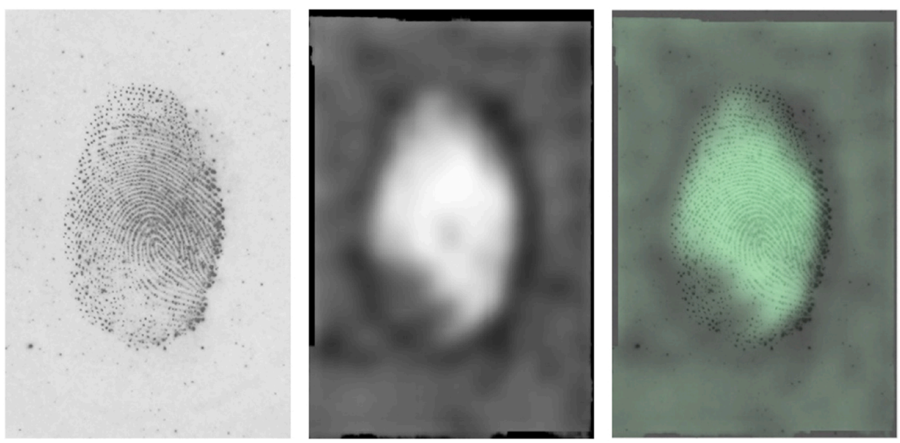

Fig. 3. Illustration of a clarity map (middle) generated by ILFQM when assessing a fingermark image (left). The maps are in greyscale, with the lighter areas corresponding to areas of higher clarity. On the right, the clarity map has been colored in green and superimposed to the fingermark to image the matching between the clarity areas and the ridge details.

considered substrates. Despite variations in the respective experimental designs (i.e., reagent formulations and types of substrates), all these studies led to a common conclusion: the superiority of IND/Zn over DFO. Such a consensus can be considered as the "ground truth" that any study carried out to compare DFO and IND/Zn should reach.

In this study, we aim at comparing DFO and IND/Zn using quality metric algorithms by considering a dataset composed of 1'800 full fingermarks and 1'800 half fingermarks. Our working hypothesis is that if such algorithms have any merit in this area, they should be able to show the superiority of IND/Zn over DFO as established by the consensus in the literature. The quality assessment step was performed by using two algorithms: LQM (Latent Quality Metrics) and ILFQM (Improved Latent Fingerprint Quality Metrics). The first algorithm was developed by Noblis/FBI and has shown good correlation with human scoring [11]. The second is an algorithm recently developed by IDEMIA taking advantage of deep learning techniques. It provides several metrics designed to mimic human manual quality assessment and to predict the accuracy obtained with an operational AFIS system.

2. Materials and methods

2.1. Fingermark collection

After a preliminary experiment involving the detection of natural fingermarks by IND/Zn (results not shown), ten people were selected from a group of potential donors to obtain a diversity in the quality of the fingermarks deposited. The resulting group was composed of three poor, four average and three good donors. The respective donor quality being overall assessed by the response each donor marks gave to IND/Zn.

Five porous substrates were used: white recycled office paper (Canon, 80 g/m^2), white recycled envelopes (Cora, 80 g/m^2), kraft envelopes (Cora, 90 g/m^2), pink blotting paper (Suffren, 125 g/m^2), and white office paper (Clairefontaine, 200 g/m^2).

Donors were asked to deposit natural fingermarks [2]. To do this, they first had to wash their hands with soap and water, dry them, and were then invited to return to their normal occupations for a period of at least 30 min. Just before the deposition, the donors were asked to rub their hands together to distribute the secretions evenly over the surface of their fingers. Depletion series were created on each substrate by simultaneously depositing the three central fingers (i.e., index, middle and ring) six times in a row, as shown in Fig. 1. During the deposition, the donors were instructed to apply a constant and similar pressure for a few seconds, the pressure being not controlled.

Three ageing times were considered in this study, namely one day, one week and one month. During these periods, the substrates bearing fingermarks were kept in a drawer, protected from light and under laboratory conditions (20 – 24 °C).

To summarize, each donor left 54 fingermarks on each substrate, resulting in a total of 2'700 fingermarks when considering the five

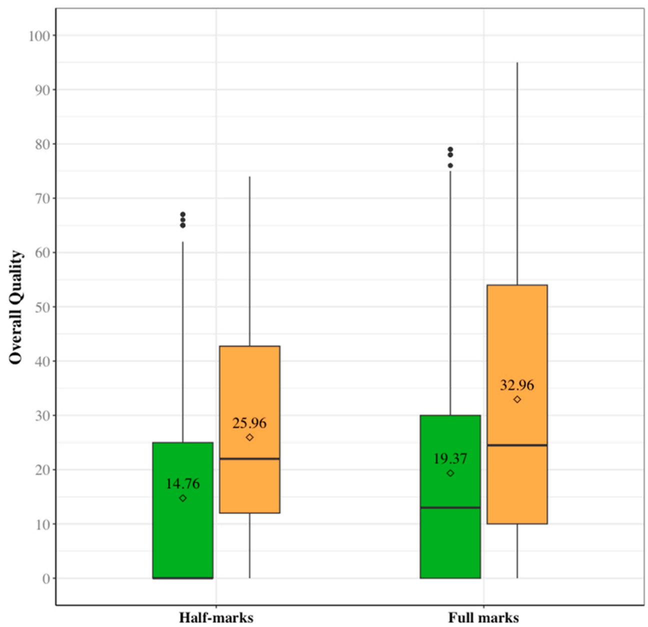

Fig. 4. Boxplot distribution of the *OQ* scores associated to the half and full fingermarks detected with DFO (green) and IND/Zn (orange), considering all donors and substrates. The median values are represented by the black line in the boxes and the average values are shown.

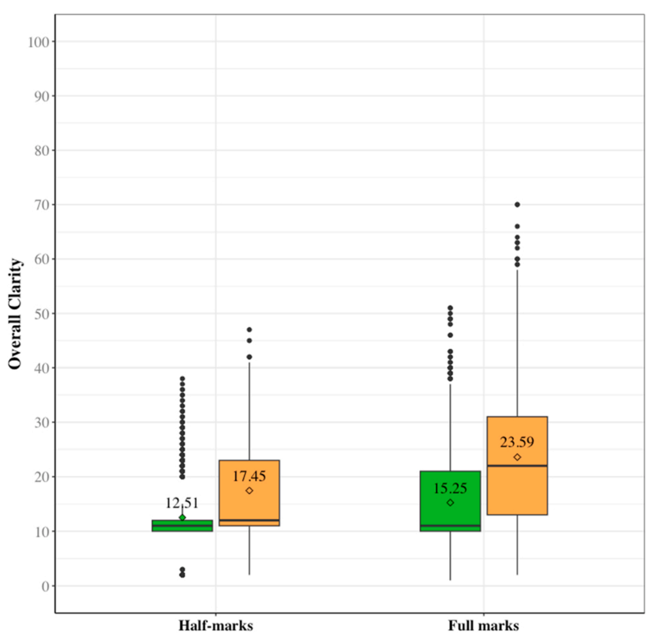

Fig. 5. Boxplot distribution of the *OC* scores associated to the half and full fingermarks detected with DFO (green) and IND/Zn (orange), considering all substrates and donors. The medians are represented by the black line in the boxes and the average values are shown.

substrates and the ten donors.

2.2. Fingermark processing

The fingermark sets were managed by following the strategy illustrated in Fig. 1. First, the substrates were cut in half heightwise, allowing the central marks to be divided into two half-marks. After this step, the set of fingermarks was composed of 3'600 distinct fingermarks (i.e., 1'800 full marks + 1'800 half-marks). To avoid any bias induced by an inhomogeneous pressure during deposition or by the finger type, the substrates were then cut widthwise so that fingermarks could be associated to IND/Zn and DFO by alternating the left and right sides of the substrates [2]. Overall, 1'800 fingermarks (i.e., 900 full marks + 900 half-marks) were processed with IND/Zn and 1'800 fingermarks with DFO. For the remainder of this paper, the term « corresponding » (half-) marks is used to designate the fingermarks which were deposited simultaneously by the three fingers at a given moment.

The IND/Zn and DFO solutions were prepared by following recipes summarized in Table 1. For both reagents, the substrates were quickly immersed in the working solution and left to air-dry. The fingermarks were then processed through a heat press for IND/Zn (i.e., 165 °C for 10 s) and in an oven for DFO (i.e., 100 °C for 20 min).

Table 2
Number of corresponding marks (and percentages) for which the *OQ* and *OC* scores were in favor of one of the three trends regarding IND/Zn and DFO.

	Half – marks		*Full marks*	
	OQ	*OC*	*OQ*	*OC*
IND/Zn > DFO	604 (67.1%)	660 (73.3%)	622 (69.1%)	710 (78.9%)
IND/Zn = DFO	145 (16.1%)	109 (12.1%)	94 (10.4%)	49 (5.4%)
IND/Zn < DFO	151 (16.8%)	131 (14.6%)	184 (20.5%)	141 (15.7%)
Total	900 (100%)	900 (100%)	900 (100%)	900 (100%)

2.3. Fingermark recording and image processing

The illumination and filtration conditions were chosen to correspond optimally to the maximum of the excitation and emission spectra of the two reagents studied [32,33].

The fingermarks detected with IND/Zn were observed in photoluminescence using a 532 nm TracER Compact laser (Coherent, USA) and an orange longpass filter (OG570) set on the camera (Canon EOS 6D equipped with a Canon Compact-Macro EF 50 mm + Life Size Converter EF). At the exception of the white recycled envelope, the fingermarks detected with DFO were observed using a 577 nm TracER Compact laser (Coherent, USA) and a purple band blocking filter (to block 575–579 nm) set on the camera (Canon EOS 6D equipped with a Canon Compact-Macro EF 50 mm + Life Size Converter EF). For the white recycled envelope, the background luminescence was too intense at 577 nm and impeded the observation of the fingermarks. Consequently, the DFO-processed fingermarks on this substrate were photographed under the same conditions as for IND/Zn, which turned out to be the most suitable conditions in this case.

The images were processed using Adobe Photoshop 2023. They have been converted to black and white (layer « *Black & White* » – grayscale 8 bits), and the contrast was reversed to obtain dark ridges on a lighter background (layer « *Invert* »). The images were resized to reach a 1:1 scale. Finally, the pictures were saved at 500 ppi, in JPG format with minimal compression. No further image enhancement was applied.

2.4. Fingermark quality assessment

The recorded fingermarks were evaluated by both algorithms: LQM and ILFQM. Both offer a range of metrics informing quality. The most relevant metrics are shortly described below:

- LQM [15,34]
  - o *Overall quality* (*OQ*) indicates the predicted probability of a successful AFIS search. The *OQ* scores range from 0 to 100.
  - o *Overall clarity* (*OC*) is a measure of the level of confidence in the presence or absence of friction ridge detail in the mark. The *OC* scores range from 0 to 100. Each clarity level is associated with a color ranging from black, red, yellow, green and blue. The meaning of these colors is detailed in Fig. 2.

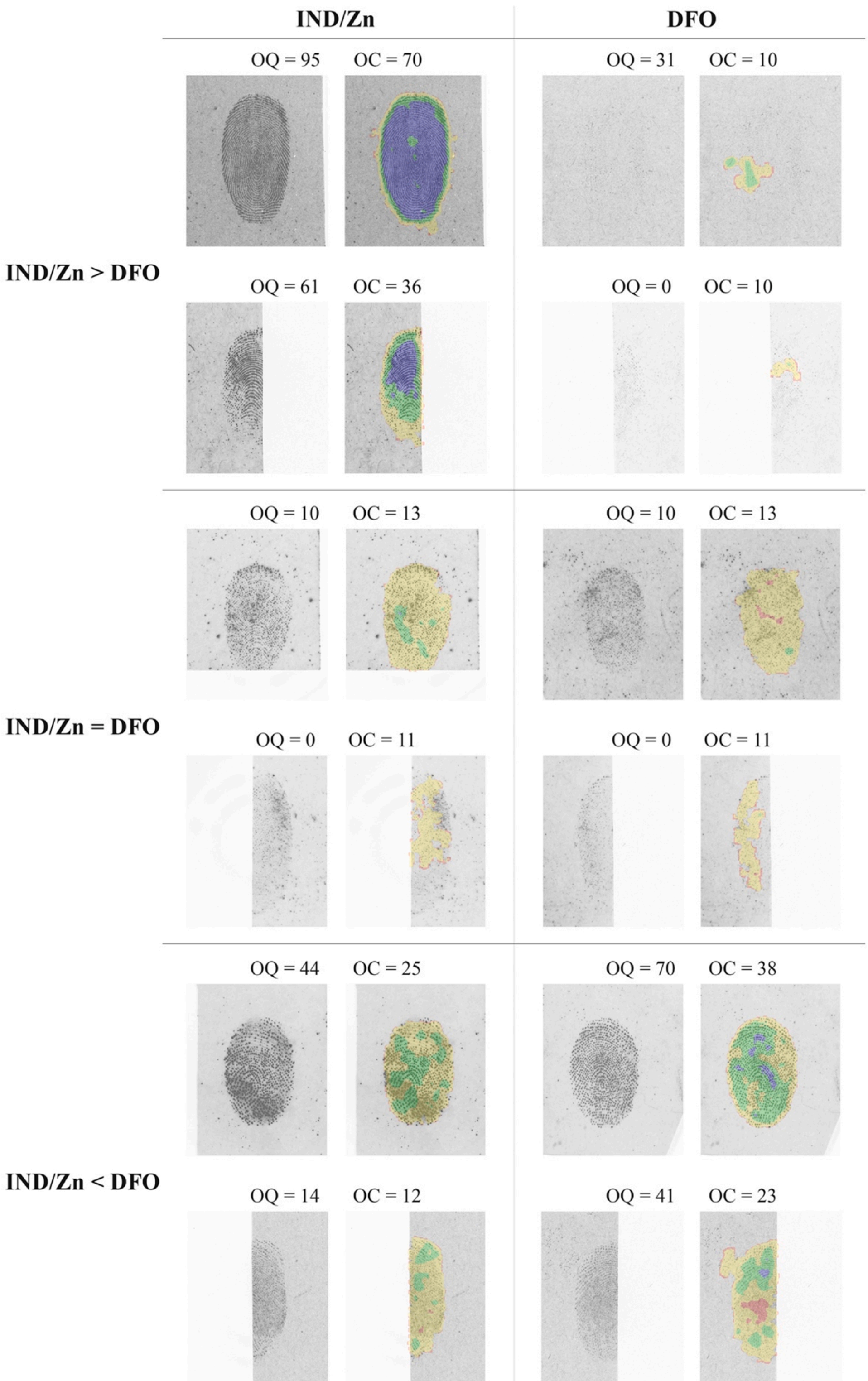

Fig. 6. Selection of corresponding (half-)marks processed with IND/Zn and DFO illustrating the three relative performance trends. The clarity maps generated by LQM metrics are shown, as well as the *OQ/OC* values.

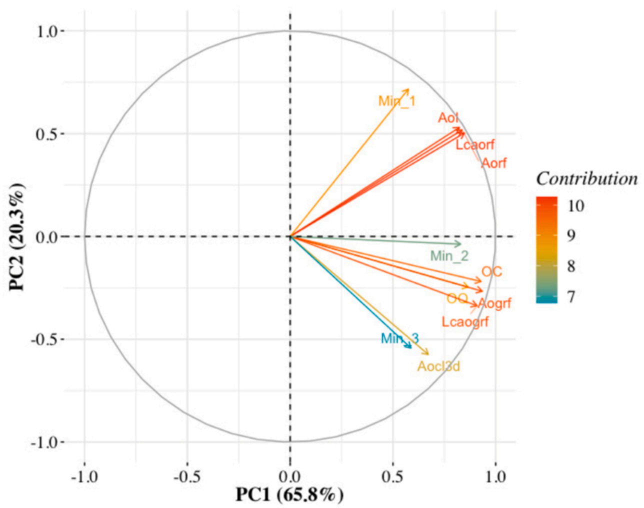

Fig. 7. Contribution of the LQM metrics according to the two principal components (PC1 and PC2) for the separation of the half-mark data.

- o *Area of impression* (*AoI*) provides a numerical value corresponding to the size of the ridge flow area (in mm^2) which presents a red clarity level or better.
- o *Area of ridge flow* (*Aorf*) provides a numerical value corresponding to the size of the ridge flow area (in mm^2) which presents a yellow clarity level or better.
- o *Area of good ridge flow* (*Aogrf*) provides a numerical value corresponding to the size of the ridge flow area (in mm^2) which presents green and blue clarity levels.
- o *Area of clear level 3 detail* (*Aocl3d*) provides a numerical value corresponding to the size of the ridge flow area (in mm^2) which presents only blue clarity level.
- o *Largest contiguous area of ridge flow* (*Lcaorf*) provides a numerical value corresponding to the size of the largest contiguous area of ridge flow (in mm^2) which presents yellow to blue clarity levels.
- o *Largest contiguous area of good ridge flow* (*Lcaogrf*) provides a numerical value corresponding to the size of the largest contiguous area of ridge flow (in mm^2) which presents green and blue clarity levels.
- o *Automated minutiae 1, 2* and *3* (*Min_1*, *Min_2*, *Min_3*) are the number of minutiae in yellow, green, and blue clarity areas respectively.

- ILFQM [proprietary of IDEMIA]
  - o *score* provides a discontinuous numerical value for a general quality regarding the automatic encoding and matching expectation on an AFIS.
  - o *expert_score* provides a value which reflects the quality assessment of a fingermark if made fingerprint examiners. The deep-learning algorithm was informed by labels (a range of positive (21) and negative (18) quality features) assigned manually by experts to 1000 fingermarks.

In addition to these metrics, both algorithms generate a clarity map for each image (Figs. 2 and 3). Such maps provide a visual indication of the area of clarity considered by the algorithms when assessing the fingermarks.

2.5. Statistical analysis

First, the distributions of scores for DFO and IND/Zn were plotted for each metric. That allowed for a general description of the data obtained. Second, still in a univariate approach, the average values were calculated and a Wilcoxon statistical test was applied to test for significant differences. Third, when considering corresponding pair of (half-)marks, the number of cases for which IND/Zn was shown to be superior, equal and inferior to DFO was counted.

Finally, the scores obtained from LQM and ILFQM were considered in a multivariate way using principal component analysis (PCA). Pearson correlation coefficient between the different metrics was calculated before applying a MANOVA and a PerMANOVA to test for overall multivariate differences between IND/Zn and DFO.

All statistical analysis and graphical representations were carried out using the RStudio© (version 2023.03.0+386) software with a range of libraries, such as *tidyverse* [35] and *ggplot2* (part of the *tidyverse* suit) for data wrangling and plots, *FactoMineR* [36] for PCA, *stats* [37] for MANOVA and *vegan* [38] for PerMANOVA.

3. Results

Given the differences between half and full marks (i.e., ridge flow area and ridge flow interruption), the scores for these two groups have

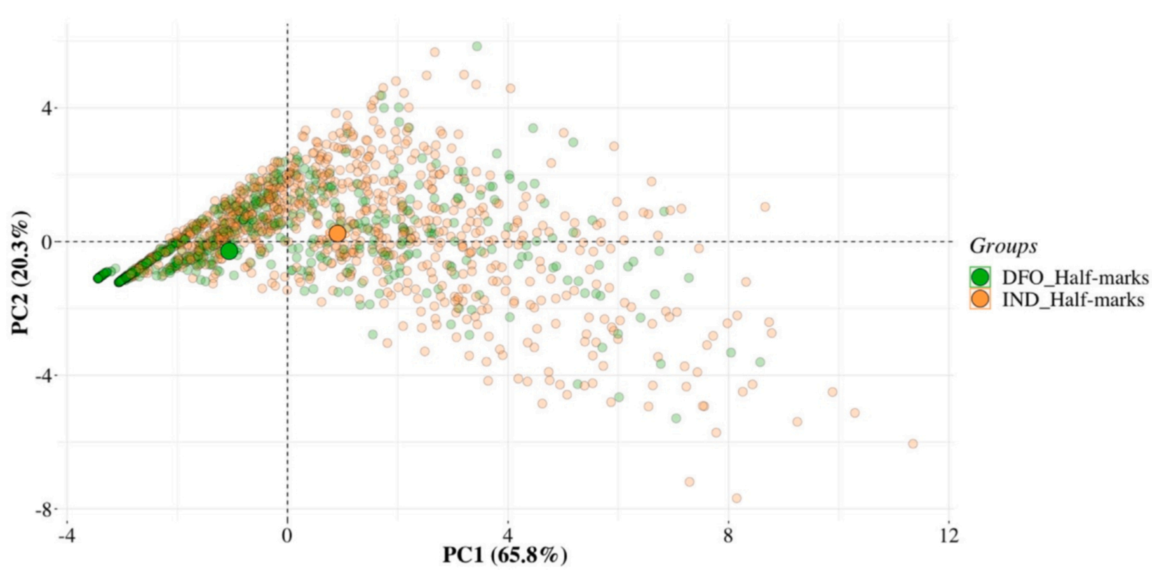

Fig. 8. PCA score plot of LQM data for the half-marks according to the two principal components (PC1 and PC2). The green points are associated with fingermarks detected with DFO and the orange ones with those detected with IND/Zn. The average for each group is represented by the two largest points.

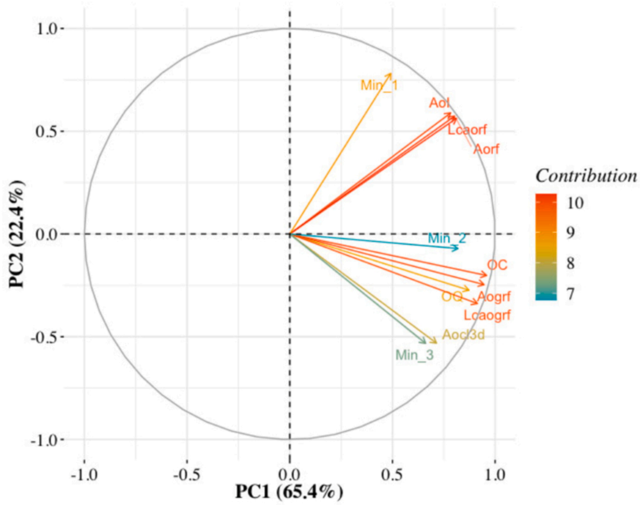

Fig. 9. Contribution of the LQM metrics according to the two principal components (PC1 and PC2) for the separation of the full mark data.

been considered separately. By doing so, any variation in the algorithm behavior when processing both types of marks may be observed.

Out of the 3'600 fingermarks (i.e., 1'800 for IND/Zn and 1'800 for DFO), 262 fingermarks (14.6%) processed with DFO (i.e., 124 full marks and 138 half-marks) and 16 marks (0.9%) processed with IND/Zn (i.e., 6 full marks and 10 half-marks) were not detected. They could not be photographed and submitted to the algorithms. Therefore, they did not receive any score from both algorithms.

3.1. Results for LQM

Overall, the scores obtained for the LQM metrics are on average higher for IND/Zn compared to DFO. The trend is comparable between full and half fingermarks. For the sake of clarity and to avoid redundancy, only the results for the two main metrics, namely *OQ* and *OC*, are described below. The results for the other metrics are available as Supplementary data.

The score distributions are presented in Figs. 4 and 5. On each plot, there are some overlaps between the boxplots. There is therefore no perfect separation between the different groups of data. However, some differences can be observed. The dispersion of the metrics for full and half-marks detected with DFO extends over smaller spreads than for those for IND/Zn marks. The average values are indeed lower. For example, in the case of full fingermarks for the *OQ* metric, the average obtained for DFO is 19.37, whereas it is 32.96 for IND/Zn. In addition, the half-mark scores for both metrics are generally lower than for full marks. For the *OC* metric for example, the average for half-marks detected with DFO is 12.51 and for full marks 15.25.

When comparing the corresponding (half-)marks, more than two thirds of the fingermarks received *OQ* scores (67.1% for half-marks and 69.1% for full marks) and *OC* scores (73.3% for half-marks and 78.9% for full marks) in favor of IND/Zn (Table 2). The fingermarks that were not detected were taken into consideration in this comparison: if one of the marks was not detected, but the corresponding one well, then the result of the comparison was in favor of the technique which detected the mark. In the case where both marks were not detected, the performance was classified as equivalent for both techniques.

Images of full and half fingermarks illustrating the different relative performance between IND/Zn and DFO are shown in Fig. 6. The *OQ* and *OC* scores are indicated for each fingermark. The differences between the scores for corresponding marks can also be seen in the clarity maps. For fingermarks that have been assessed by LQM as being of better quality than their corresponding mark (Fig. 6 – top and bottom), the clarity maps generated show larger green and blue clarity areas.

A Wilcoxon test was carried out to compare the average *OQ* and *OC* scores obtained for the two techniques (detailed results not shown). The obtained *p*-values (i.e., $< 2.2\times10^{-16}$) being lower than the threshold (i.e., 0.05), the differences observed between IND/Zn and DFO are hence to be considered significant on a variable-by-variable basis.

A PCA was applied to the LQM metrics. The results obtained for half and full marks are shown in Figs. 7, 8 and 9, 10 respectively. Figs. 7 and 9 highlight the contribution of each of the metrics to the data separation. When the angle between two vectors is small, then the variables are correlated (e.g., *AoI*, *Aorf* and *Lcaorf*). When the angle is close to 90°, the variables are unlikely to be correlated (e.g., *Min _1* and *Aocl3d*).

Figs. 8 and 10 show that there are overlaps between the fingermarks detected with IND/Zn and DFO. However, a separation between these data exists and is visible thanks to the average points plotted. A MANOVA (Wilks' lambda) and a PerMANOVA (Mahalanobis distance), which is more suitable because the variables do not strictly follow a normal

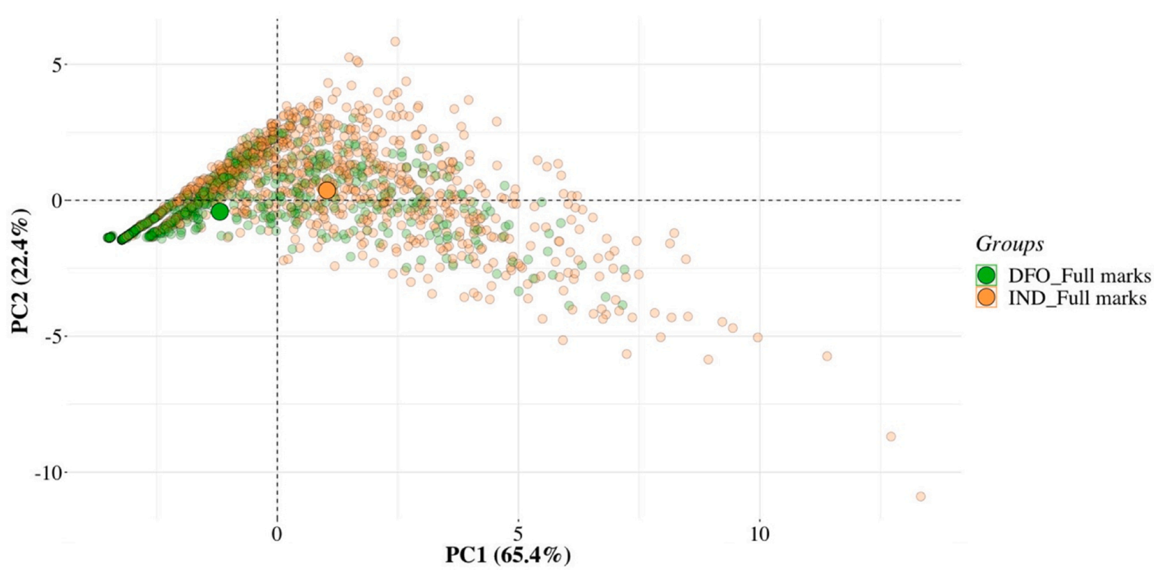

Fig. 10. PCA score plot of LQM data for the full marks according to the two principal components (PC1 and PC2). The green points are associated with fingermarks detected with DFO and the orange ones with those detected with IND/Zn. The average for each group is represented by the two largest points.

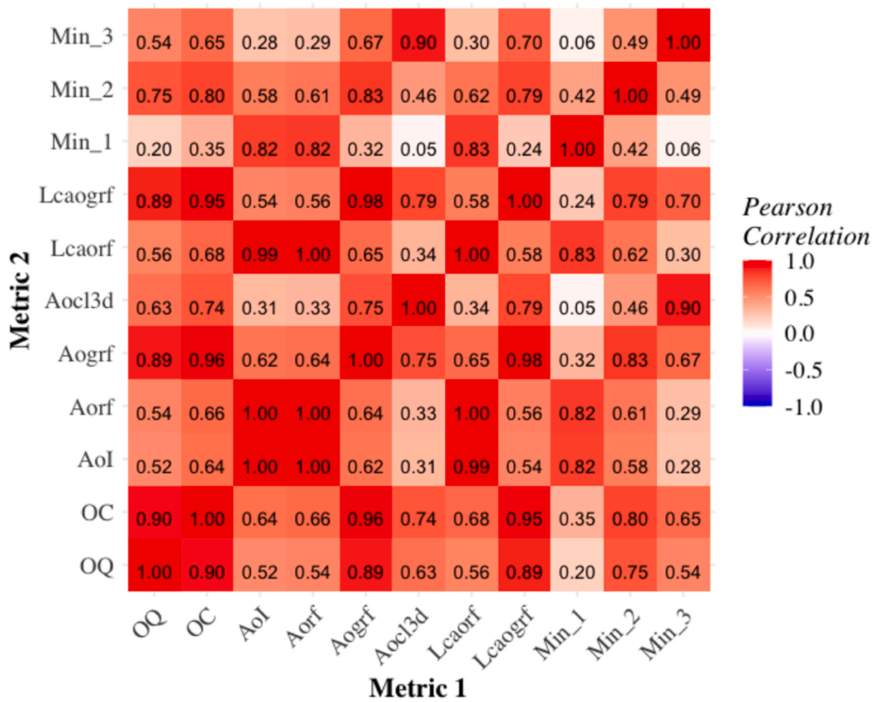

Fig. 11. Pearson correlation matrix between the metrics of LQM.

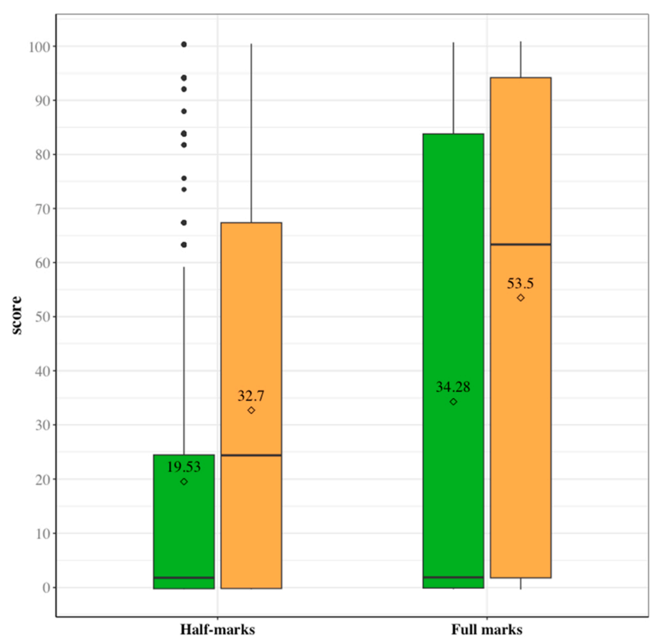

Fig. 12. Boxplot distribution of the *score* values associated to the half and full fingermarks detected with DFO (green) and IND/Zn (orange), considering all donors and substrates. The median values are represented by the black line in the boxes and the average values are shown.

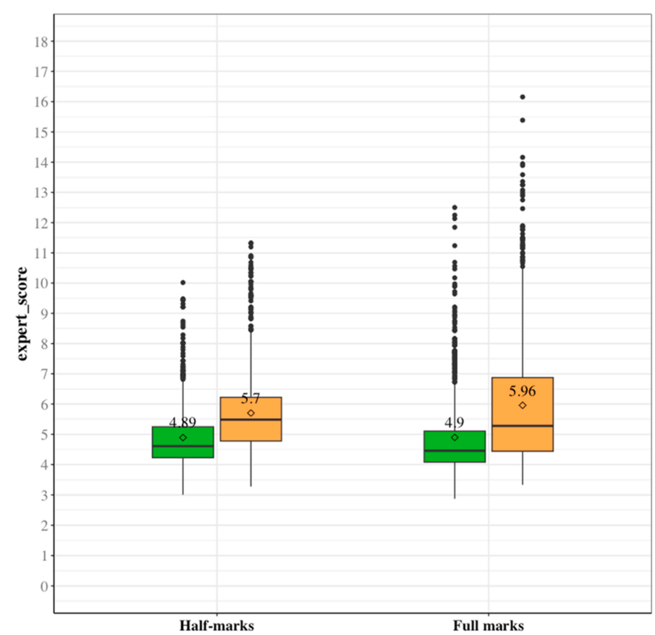

Fig. 13. Boxplot distribution of the *expert_score* values associated to the half and full fingermarks detected with DFO (green) and IND/Zn (orange), considering all donors and substrates. The median values are represented by the black line in the boxes and the average values are shown.

distribution, were applied to determine whether the data separation is significant. For their implementation, the variables studied must not be overly correlated with each other. The Pearson correlation was calculated between the different metrics (Fig. 11). It was decided that only metrics with a correlation strictly less than 0.80 would be retained for statistical testing: *OQ*, *Min_1*, *Min_2* et *Min_3*. For both full and half-marks, the *p*-values obtained for the MANOVA (i.e., $< 2.2\times10^{-16}$) and the PerMANOVA (i.e., 9.99×10^{-5}) are below the significant threshold set at 0.05. Based on these results, it can therefore be concluded that the difference observed between the scores obtained for the fingermarks detected with IND/Zn and DFO is significant.

3.2. Results for ILFQM

Overall, the values obtained for *score* and *expert_score* metrics are on average higher for IND/Zn compared to DFO. The trend is the same for

Table 3
Number of corresponding marks (and percentages) for which the *score* and *expert_score* values were in favor of one of the three trends regarding IND/Zn and DFO.

	Half – marks		*Full marks*	
	score	*expert_score*	*score*	*expert_score*
IND/Zn > DFO	562 (62.5%)	743 (82.6%)	620 (68.9%)	717 (79.7%)
IND/Zn = DFO	10 (1.1%)	8 (0.8%)	4 (0.4%)	3 (0.3%)
IND/Zn < DFO	328 (36.4%)	149 (16.6%)	276 (30.7%)	180 (20%)
Total	900 (100%)	900 (100%)	900 (100%)	900 (100%)

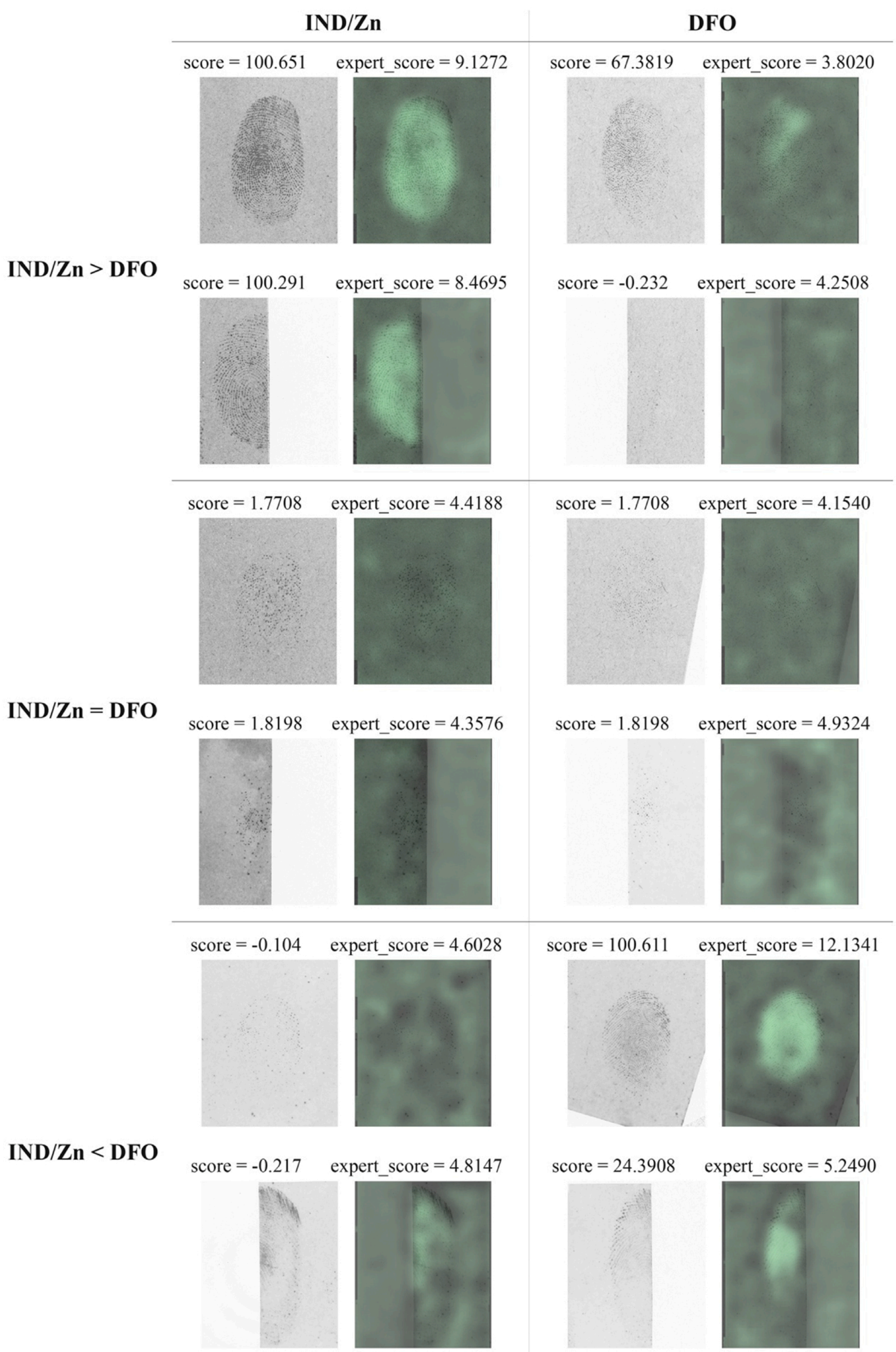

Fig. 14. Selection of corresponding (half-)marks processed with IND/Zn and DFO and illustrating the three relative performance trends. The clarity maps generated by ILFQM metrics are shown, as well as the *score/expert_score* values.

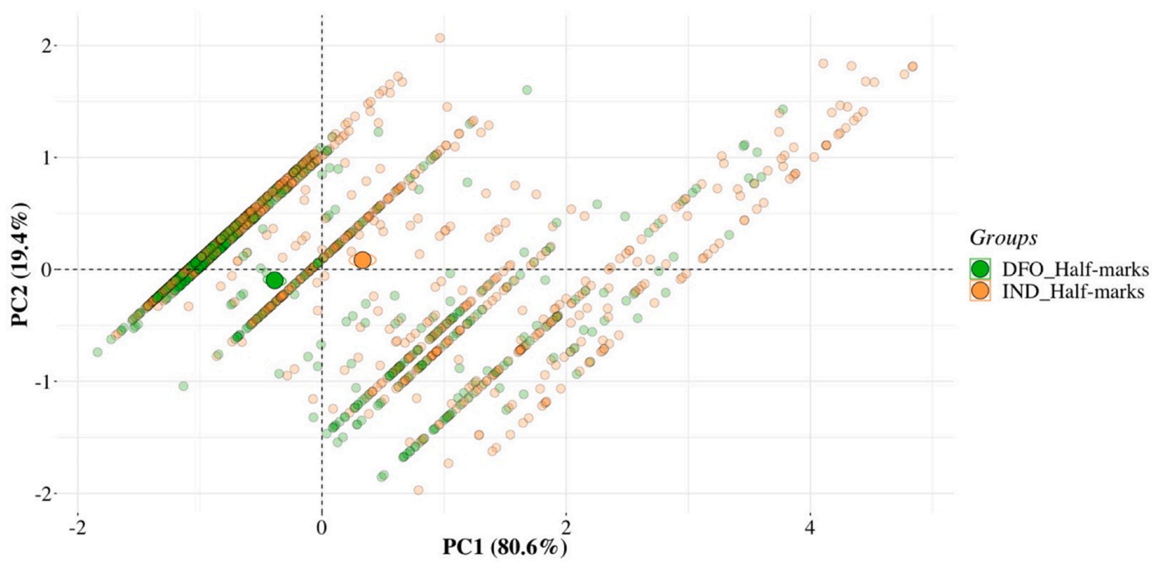

Fig. 15. PCA score plot of ILFQM data for the half-marks according to the two principal components (PC1 and PC2). The green points are associated with fingermarks detected with DFO and the orange ones with those detected with IND/Zn. The average for each group is represented by the two largest points.

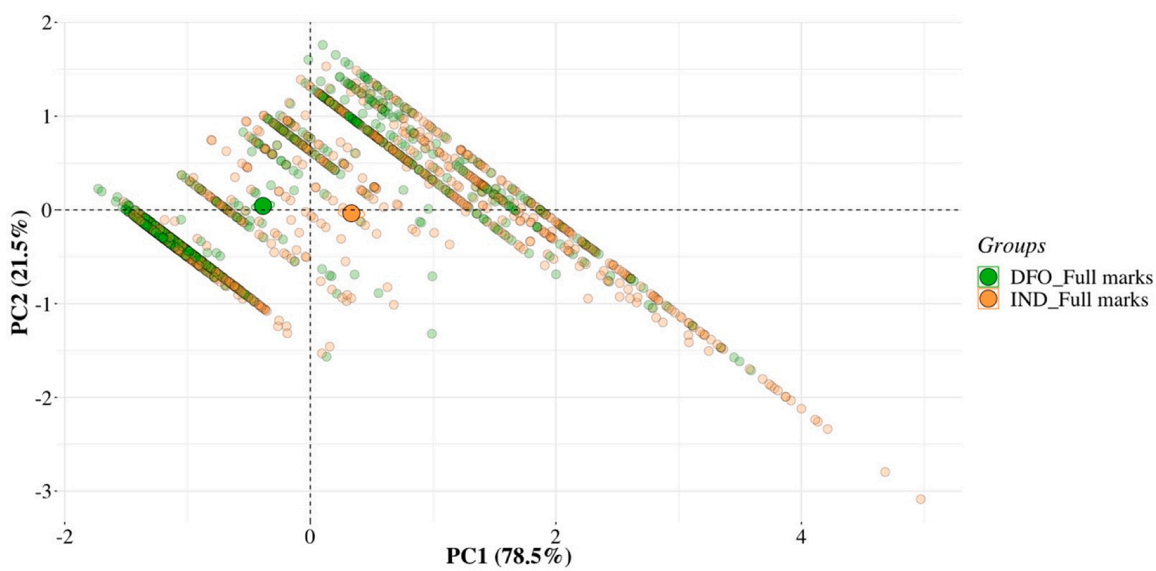

Fig. 16. PCA score plot of ILFQM data for the full marks according to the two principal components (PC1 and PC2). The green points are associated with fingermarks detected with DFO and the orange ones with those detected with IND/Zn. The average for each group is represented by the two largest points.

full and half fingermarks.

The score distributions for these two metrics are shown in Figs. 12 and 13. There is no perfect group separation: there are overlaps between the different boxplots. However, differences are observed between the ILFQM metrics for fingermarks detected with DFO compared to IND/Zn. The metrics for IND/Zn cover higher spreads than those for DFO and this is also apparent in the average values. For the *score* metric, the full marks detected with DFO have an average value of 34.28, while the marks detected with IND/Zn have a value of 53.50. For this same metric, a difference between the half and full marks is also noticeable: the values are one average higher for the latter. However, this trend is less obvious for the *expert_score*. For each detection technique, the ranges of values covered by half-marks (average of 4.89 for DFO and 5.70 for IND/Zn) and full marks (average of 4.90 for DFO and 5.96 for IND/Zn) are close.

Comparing more specifically the corresponding (half-)marks, more than two thirds of the fingermarks received higher values in favor of IND/Zn for the *score* metric (62.5% for half-marks and 68.9% for full marks). This proportion reaches 80% for the *expert_score* metric (82.6% for half-marks and 79.7% for full marks) (Table 3). Fig. 14 shows examples of corresponding half and full fingermarks for each of the three trends documented in Table 3. The *score* and *expert_score* values are given for each mark and the clarity maps provide a visual clue of the areas considered for the quality assessment.

As for LQM fingermarks that were not detected were taken into consideration in this comparison process: if one of the marks was not detected, but the corresponding mark was, then the result of the comparison was considered to be in favor of the detected marks. In the case where both marks were not detected, the performance was classified as equivalent between the two techniques. The « IND/Zn = DFO » trend contains mainly fingermarks that were not detected, except for three cases (two for half-marks and one for full marks) where the marks received very low *score* values (Fig. 14 – middle). The low number of cases in this category can be explained by the fact that the values assigned to the metrics are decimal rather than integers. When the

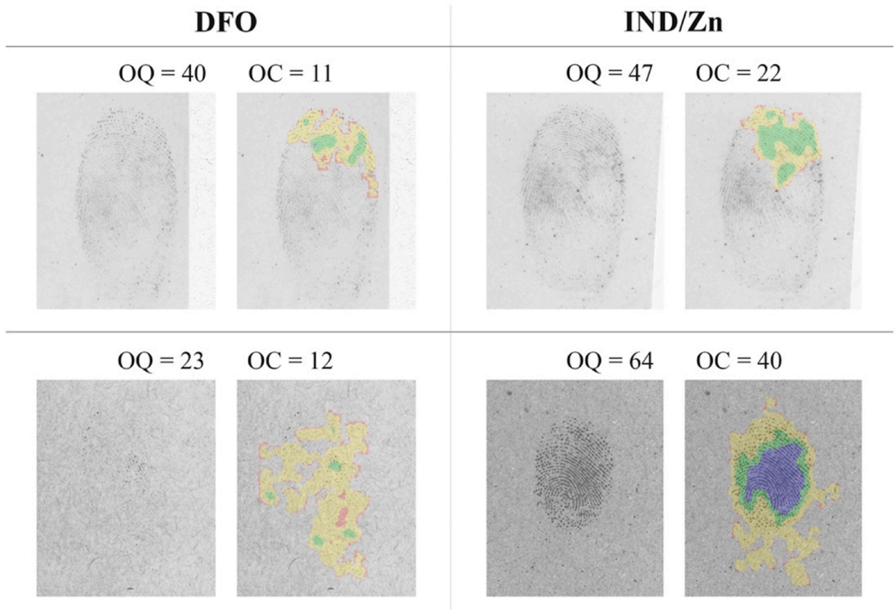

Fig. 17. Comparison of two pairs of fingermarks characterized by counter-intuitive results generated with LQM.

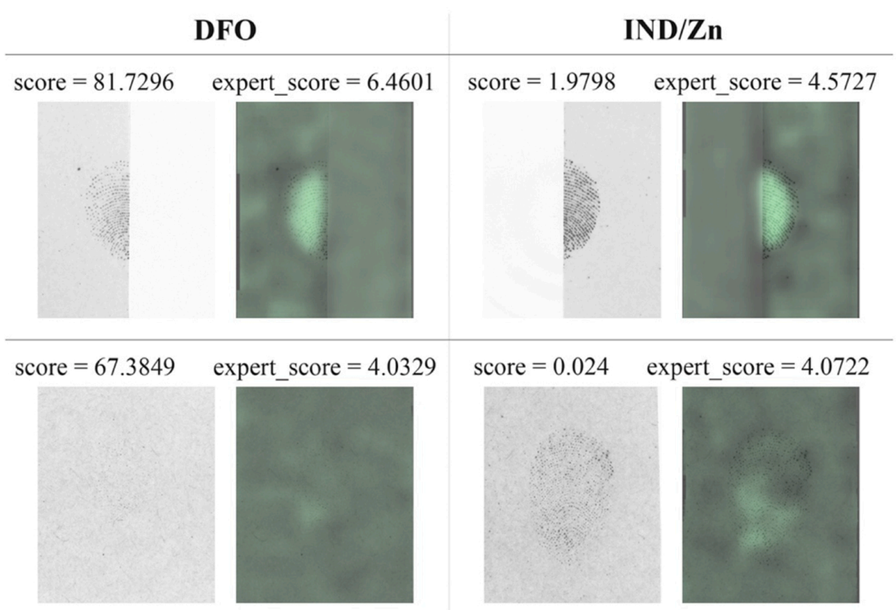

Fig. 18. Comparison of two pairs of fingermarks characterized by counter-intuitive results generated with ILFQM.

fingermark detections are visually close, the values generated for the *score* and *expert_score* will also be close and not necessarily equal.

A Wilcoxon test was carried out to compare the average *score* and *expert_score* values obtained for the two techniques (detailed results not shown). The obtained *p*-values (i.e., $< 2.2\times10^{-16}$) being lower than the threshold (i.e., 0.05), the differences observed between IND/Zn and DFO are hence considered significant on a variable-by-variable basis.

A PCA was carried out on the data obtained for the assessment of half and full marks. The score plots generated (Figs. 15 and 16) show a separation between the fingermarks detected with IND/Zn and DFO through the average points, despite the overlap between the data from the two groups. To check whether this separation is significant, a MANOVA (Wilks' lambda) is applied. Given that the *score* and *expert_score* values do not follow a normal distribution, a PerMANOVA (Mahalanobis distance) is more appropriate and is thus also applied.

The Pearson correlation was calculated between the two metrics and gave a value of 0.5766. For both half and full marks, the *p*-values calculated by the MANOVA (i.e., $< 2.2\times10^{-16}$) and the PerMANOVA (i.e., 9.99×10^{-5}) are below the significant threshold set at 0.05. Based on these results, it can be concluded that the difference observed between IND/Zn and DFO is significant.

3.3. *Influencing factors for the automatic fingermarks assessment and abnormal results encountered*

For less than 5% of the assessed fingermarks, the results provided by the algorithms were counter-intuitive with regards to their visual quality.

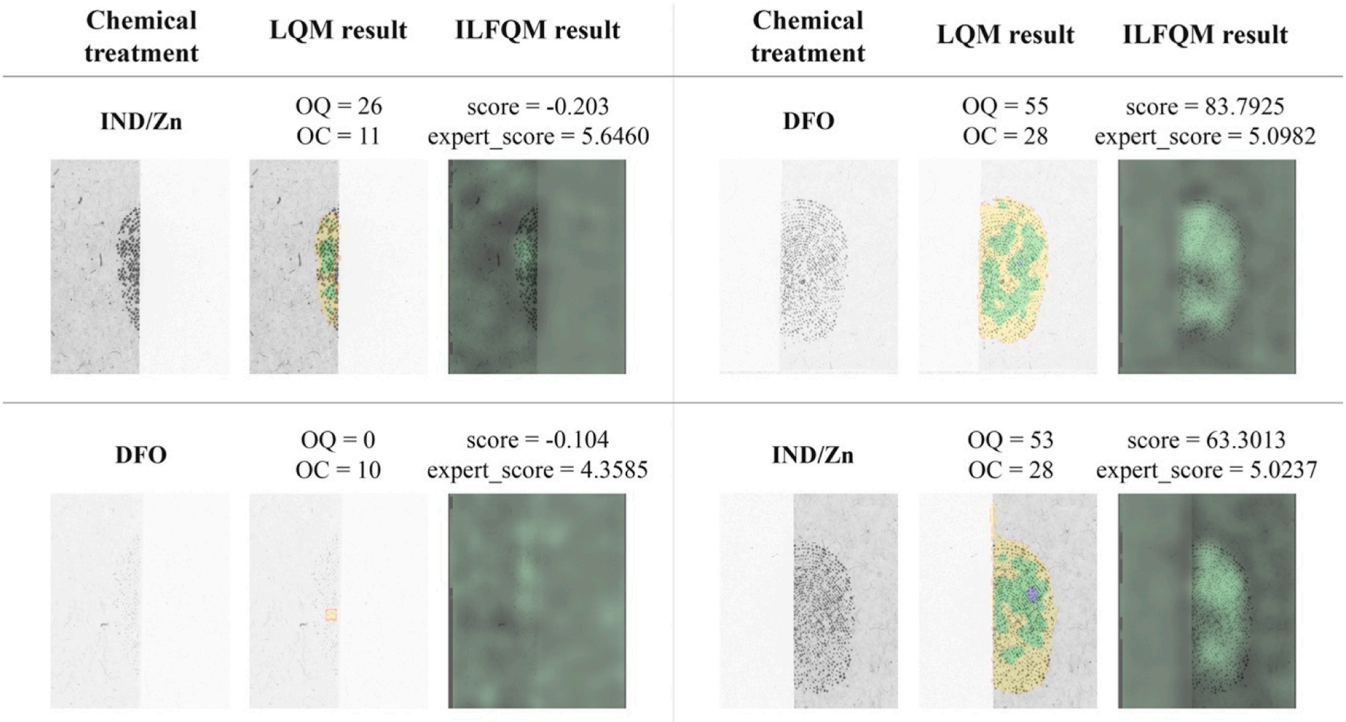

Fig. 19. Assessment of the quality of two successive fingermarks belonging to the same series of depletions, cut in half and with the left and right sides distributed alternately between IND/Zn and DFO.

LQM encountered difficulties when assessing fingermarks presenting a low contrast. In some cases, areas containing ridge details were not considered by the algorithm (Fig. 17 – top), resulting in a partial assessment of the mark quality. On the other hand, in the presence of slight background noise on the substrate, the algorithm erroneously considered elements of the latter to be part of the mark and may, in this case, reduce the quality value if this area is considered of low quality (Fig. 17 – bottom).

For ILFQM, the problematic fingermarks presented recurrent characteristics: low contrast, discontinuous dotted ridges, parallel ridges with few minutiae, etc. Fig. 18 illustrates two examples of corresponding mark comparisons where this type of characteristic is apparent. In the first case, the *score* value obtained for the half-mark detected with IND/Zn (i.e., 1.9798) is much lower than with DFO (i.e., 81.7296), despite an apparent superiority in quality for IND/Zn. In the second case, the *score* value obtained for the mark detected with DFO (i.e., 67.3849) is higher than with IND/Zn (i.e., 0.024), even though the latter is more visible. In the case of the *expert_score*, this value difference is also present, but it is less obvious given the smaller scale of values provided by this metric.

In the same way as the variability existing between the assessments conducted by a group of humans [5], the abnormal results encountered during the study do not influence the overall results obtained. Moreover, these cases are rare given the number of fingermarks examined, so there is no need to identify them for removal, which would also ruin the advantage of the automatic approach followed.

The initial size and quality of the ridge details significantly impact the fingermark quality values. This can already be noted by comparing the results obtained for half-marks and full marks. Given that the area of half-marks is smaller than that of full marks, the average *OQ*, *OC* and *score* metrics were slightly lower. The *expert_score* metric, which mimics the assessment made by human experts, was less affected.

However, despite the difference observed between full and half-marks, the conclusions drawn from the data obtained for the two groups are concordant: LQM and ILFQM evaluated the majority of fingermarks detected with IND/Zn with better scores than those detected with DFO. When setting up a study comparing detection techniques, the choice of using half-marks is therefore not an obstacle to the application of these algorithms.

During the fingermark deposition by donors, unavoidable variations between fingermarks do occur, mostly due to variations in pressure intensity and homogeneity (e.g., a donor pressing a little bit more on the left side of the finger compared to the right one) and differences in contact surfaces. When producing series of half-marks, the fingermarks should ideally be centered on the cutting line to obtain an equivalent ridge surface between both halves. However, this is not always the case because some donors may leave their fingermarks off the line. When the differences are important, the side with a larger area of visible ridges will receive a better score. A way to address this issue consists in alternating the distribution of left and right half-marks between both reagents [2], as illustrated in Fig. 1. In our case, this allowed avoiding any bias due to the unbalanced (mis-centered) halves from a depletion series (Fig. 19).

Alternating distribution of half-marks is also a way to avoid biases caused by the direction of ridge flows. Fingermark ridges are by nature not always symmetrical in their general flow, so it is possible for one side to have more characteristics than the other.

The alternating distribution of full marks is also a way to avoid biases caused by the general pattern quality (i.e., ridges on the fingers). The depletion series illustrated in Fig. 20 shows an example in which the pattern left by the ring finger (images with blue frame) contains some dark areas with little or no ridge detail visible, whereas they are absent from the pattern left by the index finger (images with red frame). It is also possible to see the detection difference between DFO and IND/Zn throughout the sequence, given that the development and observation conditions have been optimized for both techniques. Except for the first two depletions, where the quality attributed to the left mark was higher than the right mark (impact of the dark areas visible on the clarity maps), the fingermarks detected with IND/Zn obtained better results for the *OQ*, *OC* and *expert_score* metrics, while the results for the *score* metric were a little more nuanced.

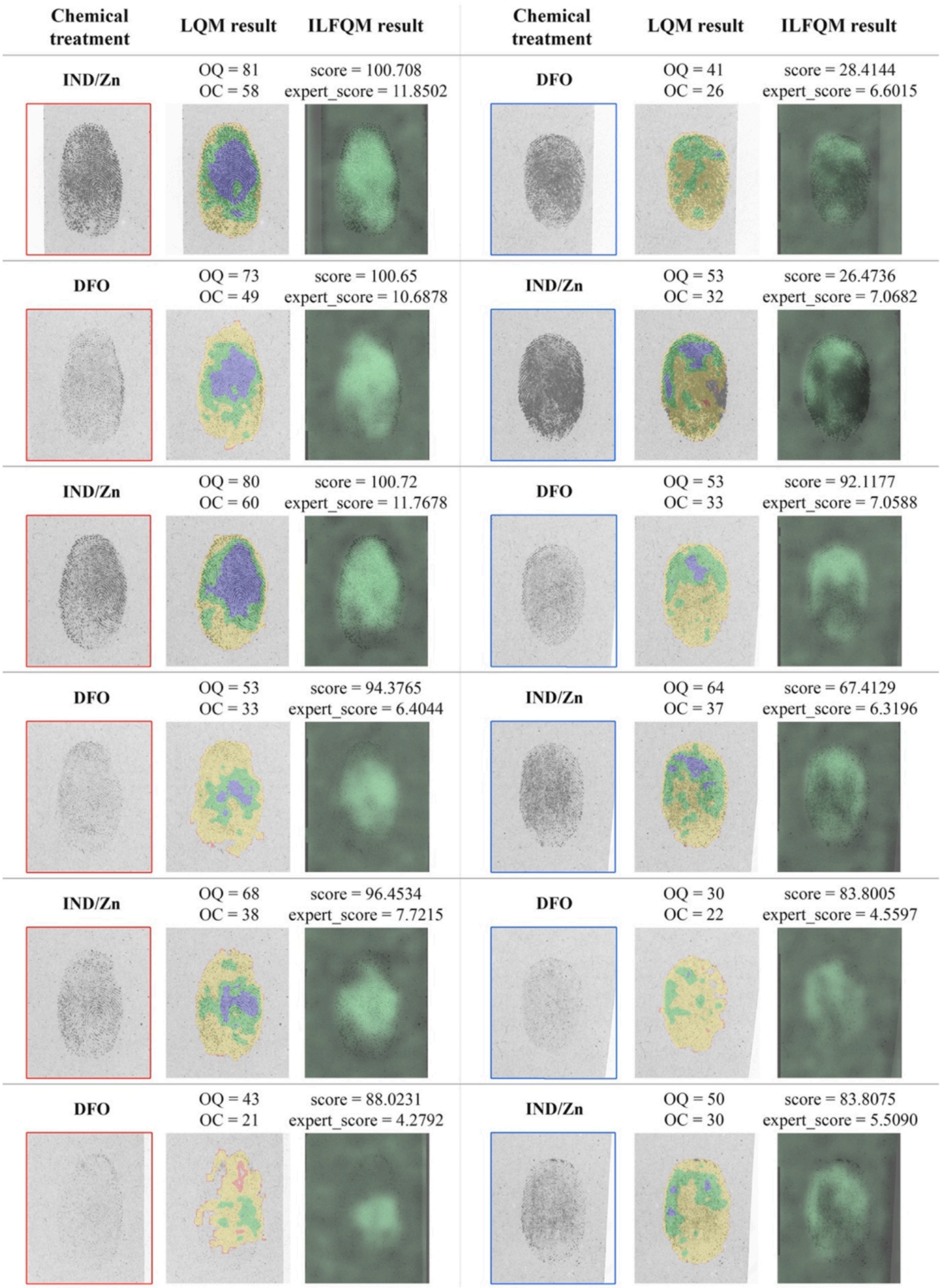

Fig. 20. Illustration of a depletion series of six full fingermarks, using the index (left rows with the red frames) and the ring finger (right rows with the blue frames) of a same donor. The marks were distributed between IND/Zn and DFO in an alternating manner, before being processed and evaluated by LQM and ILFQM.

To compare two detection techniques, the results must objectively represent the detection reality and not be biased towards one of the chemical treatments. So, in conclusion to the observations made and to optimally apply the algorithms in this context, it is necessary to process the fingermarks alternately.

4. Discussion

The objectives of this study were (1) to determine whether two quality metric algorithms, namely LQM and ILFQM, are able to identify a difference in quality for the fingermarks detected with DFO compared with those detected with IND/Zn, and (2) to verify whether the resulting trends are in line with the consensus found in the literature and based on human assessments.

4.1. Relative performance of IND/Zn compared to DFO

In the various papers comparing both reagents, the superiority of IND/Zn is undisputed. In the frame of (pseudo-)operational studies, Marriott *et al.* [23] reported 34–43% more marks detected with IND/Zn, while Olszowska *et al.* [7,24] and Luscombe *et al.* reported 1.2–1.7 times as many marks detected by IND/Zn, respectively. Bicknell and Ramotowski [28] also stated in their paper that IND/Zn allowed the development of additional fingermarks on 72.4% of the naturally-handled substrates they analysed. These values are in line with the study described in this paper, in which around 1.2 times as many marks were detected with IND/Zn (i.e., 1'784 marks for IND/Zn and 1'538 marks for DFO). Other studies have pointed out that IND/Zn produces better quality marks with increased luminescence than DFO [25–27]. The results obtained by using automated algorithms also support those conclusions. On average, the scores generated for the fingermarks processed with IND/Zn were higher in terms of overall clarity, clarity, number of minutiae, ridge area, etc. than with DFO, both for full or half-marks.

Recycled and kraft paper envelopes are the substrates characterized by the highest number of undetected marks. Of the 278 undetected marks (i.e., 16 for IND/Zn and 262 for DFO), 78.8% were deposited on these two substrates (i.e., 120 for the recycled paper and 97 for the kraft paper) and 94.5% of them had been processed with DFO. These findings indicate that DFO was therefore less effective than IND/Zn on these types of substrates. Other studies comparing IND/Zn and DFO also emphasized a lower detection rate and lower quality marks for brown paper compared to white paper [7,27,30]. The study conducted by Mayse *et al.* [26] focused mainly on the detection of marks on brown paper and showed the superiority of IND/Zn for this type of substrate: more fingermarks detected with greater luminescence.

Overall, it can be concluded from this study that (1) IND/Zn detected more marks than DFO, and (2) the results provided by the LQM and ILFQM algorithms undoubtedly point in favor of IND/Zn providing better quality fingermarks than DFO. These observations are in agreement with those documented in the literature.

4.2. Further studies

To confirm that it is possible to use automated algorithms, such as LQM and ILFQM, to assess the quality of series of fingermarks when comparing two detection techniques, it is necessary to pursue the on-going work by applying this approach to other detection techniques (e.g., one-step vs two-step cyanoacrylate fuming, including the dye-staining step) while considering a dataset made of thousands of fingermarks. By gathering more experimental data, it will be possible to further define the scope of application of such an approach, as well as its limitations.

The automated assessment of fingermark quality for detection purposes is still in its early stages. In this study, algorithms that have been initially developed for the identification field were considered. Given the fact that this approach is new, the availability of those algorithms is currently granted through specific agreements with the institutions or companies in charge of their development and distribution. We are aware that some algorithms are less easy to access compared to others. This is why we conducted the study by considering both algorithms independently from the other. This allowed us to validate the overall approach of the automated quality assessment and show that both algorithms reached comparable conclusions. It should be raised that it is in the interest of the community that access to such algorithms is granted to the people willing to proceed with this approach in their research project.

Given the positive results of this study, and the raising interest for automated approaches and AI-based tools, it could also be expected that new algorithms, integrating specificities related to detection purposes, will be developed in a near future.

One could wonder if the conclusions reached using an algorithm would remain valid once a new version of this algorithm is released, or if an original one is launched on the market. It is still too soon to address that question, but we can provide two elements of answer. First, the algorithms used in this study were (partially) developed to mimic human expertise, and hence kept a link with the human way of assessing the quality of fingermarks. This is something that could ease the transition from the current practice to the automated one. Second, benchmarking processes may become a topic of research aiming to validate the performance of new algorithms, in comparison to well-established ones, for example through openly accessible datasets of processed fingermarks to grade. If the algorithms are developed to grade the intrinsic quality of the illustrated ridge pattern and the methodology follows best practices in the field, then conclusions should remain valid with time, exactly like we are considering all the studies published decades ago when investigating the performance of a reagent.

5. Conclusions

The study presented in this paper focused on the use of algorithms to automatically assess the quality of fingermark in the context of a comparison between two detection techniques, namely DFO and IND/Zn. A total of 3'600 fingermarks (i.e., 1'800 full and 1'800 half-marks) were assessed by using LQM (Latent Quality Metric, Noblis/FBI) and ILFQM (Improved Latent Fingerprint Quality Metric, IDEMIA). The scores generated by the algorithms from the fingermarks detected with IND/Zn and DFO were compared. For each metric, the overall score distribution was plotted, and a pair-by-pair comparison of the corresponding fingermarks was also carried out. PCAs were also performed on the data provided by both algorithms. The results emphasized a significant difference in the distribution of scores obtained for both techniques, in favor of IND/Zn versus DFO. By considering the data provided by the algorithms, it could hence be concluded that IND/Zn is significantly more performant than DFO in detecting marks both in quantity and in quality. These conclusions are in line with the consensus found in the forensic literature reached after tedious manual vetting of marks against chosen ordinal scales. This approach opens the possibility to easily carry out comparison studies encompassing thousands of fingermarks, without the limitations induced by the quality assessment by human examiners.

The results obtained in this study are therefore extremely promising. They open the possibility of changing the way a comparison study between two techniques is designed, by no longer requiring the quality assessment step to be carried out by human examiners. Further studies are on-going to assess how this approach extends to other detection techniques, and to define its limits and its scope of application.

CRediT authorship contribution statement

Christophe Champod: Writing – review & editing, Visualization. **Bérénice Bonnaz:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andy Bécue:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of Competing Interest

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

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.forsciint.2024.112069.

References

[1] T. Kent, Standardizing protocols for fingerprint reagent testing, J. Forensic Identif. 60 (3) (2010) 371–379.
[2] V.G. Sears, et al., A methodology for finger mark research, Sci. Justice 52 (2012) 145–160.
[3] International Fingerprint Research Group, (IFRG), *Guidelines for the Assessment of Fingermark Detection Techniques.* J. Forensic Identif. 64 (2) (2014) 174–200.
[4] T. Hanna, S. Chadwick, S. Moret, Fingermark quality assessment, a transversal study of subjective quality scales, Forensic Sci. Int. 350 (2023) 111783.
[5] P. Fritz, et al., Variability and Subjectivity in the Grading Process for Evaluating the Performance of Latent Fingermark Detection Techniques, J. Forensic Identif. 65 (5) (2015) 851–867.
[6] D. Hockey, A. Dove, T. Kent, Guidelines for the use and statistical analysis of the Home Office fingermark grading scheme for comparing fingermark development techniques, Forensic Sci. Int. 318 (2021) 110604.
[7] A. Luscombe, V. Sears, A validation study of the 1,2-indandione reagent for operational use in the UK: Part 3 - Laboratory comparison and pseudo-operational trials on porous items, Forensic Sci. Int. 292 (2018) 254–261.
[8] S. Chadwick, et al., Investigation of some of the factors influencing fingermark detection, Forensic Sci. Int. 289 (2018) 381–389.
[9] B.T. Ulery, et al., Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners, PLoS One 7 (3) (2012) e32800.
[10] B.T. Ulery, et al., Understanding the Sufficiency of Information for Latent Fingerprint Value Determinations, Forensic Sci. Int. 230 (1-3) (2013) 99–106.
[11] B. Bonnaz, et al., Automatic assessment of fingermarks quality: exploration of the possible application in the context of detection and comparison with human examiners, J. Forensic Sci. 66 (3) (2021) 879–889.
[12] Yoon, S., E. Liu, and A.K. Jain, *On Latent Fingerprint Image Quality, in Fifth International Workshop on Computational Forensics.* 2012: Tsukuba, Japan.
[13] S. Yoon, et al., LFIQ: Latent Fingerprint Image Quality, *IEEE Sixth Int. Conf. Biom. Theory, Appl. Syst. (BTAS)* (2013).
[14] R.A. Hicklin, J. Buscaglia, M.A. Roberts, Assessing the Clarity of Friction Ridge Impressions, Forensic Sci. Int. 226 (2013) 106–117.
[15] N.D.B. Kalka, M. Hicklin, R. A, *LQMetric: A Latent Fingerprint Quality Metric for Predicting AFIS Performance and Assessing the Value of Latent Fingerprints*, J. Forensic Identif. 70 (4) (2020) 443–463.
[16] D.A. Stoney, et al., Occurrence and associative value of non-identifiable fingermarks, Forensic Sci. Int. 309 (2020) 110219.
[17] Merkle, J.S, M.; Bausinger, O.; Breitenstein, M.; Elwart, K.; Nuppeney, *M. Towards Improving the NIST Fingerprint Image Quality (NIST) Algorithm (Extended Version). Available from: https://arxiv.org/pdf/1008.0781.pdf.*
[18] R. Grigg, et al., 1,8-diazafluorenone and related compounds. A new reagent for the detection of a-amino acids and latent fingerprints, Tetrahedron Lett. 31 (49) (1990) 7215–7218.
[19] A.C. Pounds, R. Grigg, T. Mongkolaussavaratana, The Use of 1,8-Diazafluoren-9-one (DFO) for the Fluorescent Detection of Latent Fingerprints on Paper. A Preliminary Evaluation, J. Forensic Sci. 35 (1) (1990) 169–175.
[20] Ramotowski, R., et al., *1,2-Indanediones: A Preliminary Evaluation of a New Class of Amino Acid Visualizing Compounds. Fingerprint Whorld,* 1997. *23: p. 131-140 and correction in FW 124(191), p.136.*
[21] D.B. Hauze, et al., 1,2-Indanediones: New Reagents for Visualizing the Amino Acid Components of Latent Prints, J. Forensic Sci. 43 (1998) 744–747.
[22] M. Stoilovic, et al., Evaluation of a 1,2-indanedione formulation containing zinc chloride for improved fingermark detection on paper, J. Forensic Identif. 57 (2007) 4–18.
[23] C. Marriott, et al., Evaluation of fingermark detection sequences on paper substrates, Forensic Sci. Int. 236 (2014) 30–37.
[24] I. Olszowska, et al., An alternative carrier solvent for fingermark enhancement reagents, Forensic Sci. Int. 284 (2018) 53–64.
[25] V. D'Elia, et al., Evaluation and comparison of 1,2-indanedione and 1,8-diazafluoren-9-one solutions for the enhancement of latent fingerprints on porous surfaces, Forensic Sci. Int. 254 (2015) 205–214.
[26] K. Mayse, et al., An evaluation of the effect of incorporating metal salts into 1,8 diazafluoren-9-one (DFO) formulations for fingermark enhancement, Sci. Justice 59 (2019) 349–358.
[27] C.N. Loh, et al., Evaluation of the performance of IND/Zn and DFO on various porous substrates in Singapore context, Forensic Sci. Int. 315 (2020) 110440.
[28] D.E. Bicknell, R.S. Ramotowski, Use of an Optimized 1,2-Indanedione Process for the Development of Latent Prints, J. Forensic Sci. 53 (5) (2008) 1108–1116.
[29] S. Berdejo, M. Rowe, J.W. Bond, Latent Fingermark Development on a Range of Porous Substrates Using Ninhydrin Analogs - A Comparison with Ninhydrin and 1,8-Diazofluoren, J. Forensic Sci. 57 (2) (2012) 509–514.
[30] R. Lam, D. Wilkinson, Forensic Light Source and Environmental Effects on the Performance of 1,2-Indanedione-Zinc Chloride and 1,8-Diazafluoren-9-one for the Recovery of Latent Prints on Porous Substrates, J. Forensic Identif. 61 (6) (2011) 607–620.
[31] T. Fitzi, et al., Fingermark Detection on Thermal Papers: Proposition of an Updated Processing Sequence, J. Forensic Identif. 64 (2014) 329–350.
[32] Champod, C., et al., Fingerprints and Other Ridge Skin Impressions - Second Edition. 2016, Boca Raton, Florida: CRC Press LLC. 427.
[33] Exall, A., et al., *Fingermark Visualisation Source Book v3.0, UK: Home Office Defense Science and Technology Laboratory.* 2022.
[34] Federal Bureau of Investigation (FBI), LQMetrics User Guide. 2015.
[35] H. Wickham, et al., Welcome to the Tidyverse, J. Open Source Softw. 4 (43) (2019) 1686.
[36] S. Lê, J. Josse, F. Husson, FactoMineR: An R Package for Multivariate Analysis, J. Stat. Softw. 25 (1) (2008) 1–18.
[37] R. Core Team, *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*. 2023, *Vienna, Austria.*
[38] Oksanen, J., et al., vegan: Community Ecology Package. 2022.