Artificial Intelligence in Historical Document Analysis

Dhali, Maruf A.

DOI:
10.33612/diss.869247881

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2024

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 09-06-2024
INITIAL STUDIES ON WRITER IDENTIFICATION

This chapter presents the experimental setup of a complete pattern recognition (PR) pipeline and results for initial writer identification tests on the Dead Sea Scrolls images. These pilot tests demonstrate the feasibility and applicability of PR techniques in the DSS images. The chapter lays the foundation for the next parts of this thesis, especially the need for a sophisticated ink-background separation technique presented in Chapter 3. Additionally, knowledge collected within this chapter becomes useful for the more refined and elaborate writer identification work on the Great Isaiah Scrolls in Chapter 4. The work of this chapter also becomes useful in calibrating and evaluating different feature extraction techniques in both writer identification and dating domain in the rest of the thesis.

PROVENANCE
This chapter is an adaptation of an article that has been previously published as:

AUTHOR CONTRIBUTIONS
Dhali conceived the complete experimental setup, completed the tests, performed the quantitative analysis, and wrote the draft and final version of the manuscript with other co-authors.
ABSTRACT To understand the historical context of an ancient manuscript, scholars rely on the prior knowledge of the writer and the date of that document. In this chapter, we study the Dead Sea Scrolls, a collection of ancient manuscripts with immense historical, religious, and linguistic significance, which was discovered in the mid-20th century near the Dead Sea. Most of the manuscripts of this collection have become digitally available only recently, and techniques from the pattern recognition field can be applied to revise existing hypotheses on the writers and dates of these scrolls. This chapter presents our ongoing work which aims to introduce digital palaeography to the field and generate fresh empirical data by means of pattern recognition and artificial intelligence. Challenges in analyzing the Dead Sea Scrolls are highlighted by a pilot experiment identifying the writers using several dedicated features. Finally, we discuss whether to use specifically-designed shape features for writer identification or to use the Deep Learning methods on a relatively limited ancient manuscript collection that is degraded over the course of time and is not labeled, as in the case of the Dead Sea Scrolls.

2.1 INTRODUCTION

This chapter will present the preliminary results of writer identification in the DSS using several hand-crafted features. Although this gives us fast results without lengthy training on the limited labeled data of the DSS, they are certainly not the best results to be expected. We consider the results as a baseline measurement for later experiments. We suggest how to improve the results by exploiting the power of parameter-heavy machine learning methods using this small dataset. In solving this, we make a three-fold proposition of advanced statistical modeling, data augmentation, and the use of pre-trained networks.

The DSS manuscripts, written by many different writers, some of whom may have written multiple manuscripts, display a broad variety and development of different styles of this Hebrew-Aramaic script. The study of ancient handwriting provides the chronological framework, but the typological sequence of writing styles has to date not been systematically assessed for the DSS.

This project carries out the first systematic assessment of the palaeographic framework of the scrolls by combining two approaches. First, we will conduct new radiocarbon (^14C) dating on a number of physical samples of the scrolls, kindly provided to us by the Israel Antiquities Authority (IAA). Second, we will generate for the first time quantitative data for palaeographic handwriting recognition by means of Artificial Intelligence, using the Monk system [37, 252, 320]. The challenging issue of writer identification in the DSS has not been systematically dealt with before. The tools of digital palaeography enable new, significant steps forward. In this chapter, we focus on this second approach of digital palaeography.
2.1.1 Challenges in digital palaeography

In order to achieve both goals, i.e., handwriting recognition and writer identification, specific challenges at several levels of Computer Vision and Artificial Intelligence must be overcome. Initial analyses are needed for the proper extraction of characters (foreground, ink) from the background, which is mostly either animal skin or papyrus in the case of the DSS. Several image processing techniques need to be applied for optimum results of segmentation. Starting with edge detection, morphological operations, filling gaps, and then finding connected components help to automatically segment the hand-written fragments. Then further processing can localize and extract the characters. Due to the difference in the textures of papyrus and animal skin, individual measures must be taken on their distinctive periodic structures.

We have explored different feature-extraction techniques on the images of the DSS. Feature representation maps the raw pixel intensity into a discriminant high-dimensional space [164, 183] in order to capture specific information about the characters, which can be processed by algorithms in computers. This step is an important element in the field of computer vision and pattern recognition. There have been a lot of efforts to design discriminative and powerful features [164]. Though the (Deep) Learning-based feature representation may achieve better results in many cases, hand-crafted features have several advantages in the analysis of handwritten documents, especially for historical manuscripts. This is due to the amount of data in historical manuscript collections, which is usually not big enough to train deep neural networks. In contrast, the ImageNet dataset [63] contains millions of samples for training the network. The challenge becomes even higher when the total number of usable pages comes to a count of hundreds in the DSS. To take the opportunities offered by the Deep Learning methods, the associated challenges need to be overcome in order to analyze the DSS.

2.2 DATA

2.2.1 Manuscript images

We will use digital images of the DSS as our primary data. There are various sources for digital images of the DSS manuscripts. The source used in this study is kindly provided to us by Brill Publishers [165]. There are 2463 images in the Brill collection with varied resolutions from 600 by 600 pixels to 2800 by 3400 pixels, approximately.

The Brill images are single-layered grayscale images with 300 ppi (pixels per inch) on both axes. They have shadows and reflections from external lighting. Additionally, the lighting throughout all the images is not uniform. Among the images, the ones containing several fragments are mostly not aligned in a horizontal way for text reading. This poses the issue of rotation variance in characters. Many of the images
also contain paper calibration strips for scale representation and contemporary hand-written numbers. The digitization noise can also be noticed in many of the images (see Figure 2.1).

Figure 2.1: Two of the Brill images, PAM 40.456 (left) and PAM 40.531 (right); the images show digitization noise, alignment issues with the small fragments, shadows near the border, and lighting problems.

The IAA images are clear, properly aligned, and free from the problem of lighting and shadow, unlike the Brill ones. Additionally, the different exposure bands of the IAA hold important underlying information regarding the fragments providing essential attributes for the scrolls. For example, one particular band provides clear information on the ink (foreground), whereas another one gives more details on the underlying leather/papyrus (background) on the retro side. Some bands are useful for the textual contents, and some other bands give a better understanding of the textural properties of the scroll material. Extraction of this useful information is possible on both single images and multi-spectral-fused images. As a whole, digital image data has provided a new and broader perspective in the quantitative analysis and processing of the scrolls.

The scope of the current chapter is limited to the images in the Brill collection, but in the near future, we expect to publish our results on the digital data from the IAA’s Leon Levy Dead Sea Scrolls Digital Library. The quality and the challenges of the Brill images can not be seen as a setback, rather, it would be a starting benchmark for the robustness of our work. Additionally, the possibility of using the IAA material will improve our results.

2.2.2 Ground truth

Unlike many other historical manuscripts, the DSS collection does not have a structured and complete dataset nor the ground truths for testing. Before diving into any sort of computer-aided writer identification, the ground truths must be there to analyze the results. To establish the ground truths, we need experts in the field
and also their proper access to the data. We have this two-folded advantage in our group: first through the presence of palaeographic experts and second through the Monk system, which is accessible through web browsers. By integrating these, we started to label the DSS image data for ground truths.

We have proposed two different methods for labeling. The first one is to detect the region of interest in the DSS images. The second one is to create the ground truth for character labels. Both these tasks require manual labor from experts with palaeographic background knowledge on the image level to the pixel level. This is the benchmark in identifying the writers and aligning the temporal developments in script style for the DSS. In this chapter, we will only use the labeled regions of interest (we call them FragmROIs, by shortening the term fragment region of interest) in order to build algorithms to extract features (using available methods) and identify (recognition) writers. From the DSS images, FragmROIs were selected and labeled by the palaeographic experts using the Monk system (see Figure 2.2). Those rectangular FragmROIs could consist of the entire text on an image, or of only a section of text selected from an image. Different FragmROIs from one and the same manuscript were labeled as stemming from one writer or scribe unless palaeographers distinguished two scribes as writers of the manuscript.

While labeling the writers, we have set up a provisional naming rule starting with the name scribeAxxx, where xxx are numerical values starting from 001. Each of the human-labeled new writers will be allocated an individual value. The term A is put before the numerical values in order to preserve the tag of original labeling from the palaeographic experts. If, at a later stage of our study, two of the writers are found to be the same one according to the system, then they can be referred to with the new name of scribeBxxx having two child-node of format scribeAxxx, preserving the original label.
The present pilot study is based on two distinct sets of writers. The first set is a limited sample of 323 FragmROIs labeled as having been written by 13 scribes, namely the scribes of 1Qlsa columns 1-27, 1Qlsa columns 28-54, 1QS, 1QSa, 1QSb, 1QM, 1QpHab columns 1-12, 1QpHab columns 12 end-13, 4Q53, 4Q75, 11Q5, 11Q19 columns 2-5, 11Q20. We labeled them from scribeA001 to scribeA013. Distinct manuscripts were labeled as deriving from different writers, even though in several of the manuscripts of the first set, palaeographers think that one and the same writer produced multiple manuscripts [286]. To incorporate the palaeographic opinion, we then merged those 13 scribes into 7 scribes by introducing the scribeBxxx series. Then we took the second set of 13 scribes with a limited sample of 124 FragmROIs labeled as scribeA014 to scribeA026 (the scribes of 4Q266, 4Q504, 11Q10, 1Q22, 4Q209, 4Q167, 4Q6, 4Q286, 4Q381, 4Q405, 4Q491, 4Q431, 4Q525). The main difference between these two sets is the number of characters per scribe. The first set has a higher number of characters than the second set. Thus, for this pilot project, we have 447 FragmROIs labeled as 20 distinguishable scribes according to palaeographic opinion.

2.3 A PILOT EXPERIMENT

The DSS image data has its own distinctive characteristics compared to other historical manuscript datasets. This dataset has quite a different appearance from, e.g., historical manuscripts such as the Medieval Palaeographic Scale dataset [186] from a previous project [108] in three aspects: (1) the number of characters in fragments from some documents can be as low as one; (2) the ink of each character has been faded out over the course of time, making it more difficult to observe and process; (3) the large diversity and lack of uniformity among text blocks, presenting a challenge for analysis. In this section, we will present the methodology used in our pilot project in writer identification to benchmark our works in analyzing the DSS.

2.3.1 Writer identification

Identifying writers using computers has been done for decades [210], which is a problem of recognizing the writer of a given document based on handwriting styles. A number of different features have been proposed and studied for writer identification on scripts from several languages, including Dutch [37], English [254], Indic [2, 144], and Arabic [38].

In the case of the DSS, we will be identifying the scribes behind the scrolls with Hebrew characters, and a hand-crafted feature especially for these characters is yet to be proposed and studied. Instead of designing a new feature, we initially started working with some of the existing textural-based and grapheme-based features. Textural-based features are based on the statistical information about the slant and curvature of the handwritten characters, and grapheme-based features, inspired by
the bag-of-words model, extract local structures and then map them into a common space \cite{114}. We briefly discuss the preprocessing techniques and the features used in this work in the following sections.

2.3.1.1 Preprocessing

As the feature extraction technique is applied to the binarized images, first, we pre-processed the DSS images. Binarizing the Dead Sea Scrolls images is quite challenging, given their diverse intensity, the similarity between ink and background traces, and image quality. We first started with Sobel edge detection \cite{267} and then removed the connected objects on the border to get rid of the markings. The morphological operation was then used, followed by image thresholding. We used the global Otsu threshold selection method \cite{202} as it is efficient and parameter-less (see figure 2.3).

![Figure 2.3](image)

Figure 2.3: The left one is a FragmROI from the Brill collection (PAM 43.787A), and the right one is the binarized image using the Otsu threshold selection method. It should be noted that the images used in this chapter are limited in number and are relatively easy to process using intensity-based methods like Otsu. This, however, can be much more challenging on the complete collection of DSS from IAA where the diversity is large.

2.3.1.2 Feature representation

Previous studies showed that the textural-based feature extraction methods perform better than grapheme-based methods \cite{114, 115}. Additionally, a more powerful approach was introduced using the spatial co-occurrence among features \cite{37, 133, 230}. The latter idea has been extended in a previous work \cite{115} with the introduction of the joint feature distribution principle (JFD principle). By accommodating these facts, we used eight textural-based methods (three of them following the JFD principle) and one grapheme-based method.
**Hinge**: The Hinge feature is the joint probability distribution of orientations of the legs of two contour fragments attached at a common-end pixel on the ink contours [37]. Figure 2.4 shows two examples of the Hinge kernel on contour fragments with leg length \( l \), and the joint probability of the two orientations, \( \alpha \) and \( \beta \) (\( \alpha < \beta \)), are quantized into a 2D histogram. Empirically we have set \( l = 7 \), and the number of bins of \( \alpha \) and \( \beta \) is set to 23. Finally, the dimension of the feature vector is 253.

\[
\begin{align*}
C(F_1) &= 1.06 \\
C(F_2) &= 1.06 \\
\end{align*}
\]

Figure 2.4: The two figures show two contour fragments with the same Hinge kernel (\( \alpha_1 = \alpha_2 \) and \( \beta_1 = \beta_2 \)) but different fragment curvature values \( C(F_c) \).

**Co-occurrence Hinge (CoHinge)**: The CoHinge feature is the joint distribution of the Hinge kernel following the JFD principle on two different points \( x_i \) and \( x_j \) with Manhattan distance \( l \) (Figure 2.6) on the contours as equation 2.1.

\[
\text{CoHinge}(x_i, x_j) = [\text{Hinge}(x_i), \text{Hinge}(x_j)]
\]

Each Hinge kernel has two values \( \alpha \) and \( \beta \), and therefore, the CoHinge kernel has four values \([\alpha(x_i), \beta(x_i), \alpha(x_j), \beta(x_j)]\), which can be quantized into a 4D histogram. The Manhattan distance \( l \) is set to 7 based on our previous study [115]. We set the number of bins of the angle to 10, and finally, the dimension of the CoHinge feature is \( 10 \times 10 \times 10 = 10,000 \).

\[
\begin{align*}
\Delta^n \alpha(x_i) &= \frac{\Delta^{n-1}\alpha(x_i) - \Delta^{n-1}\alpha(x_i + \delta l)}{\delta l} \\
\Delta^n \beta(x_i) &= \frac{\Delta^{n-1}\beta(x_i) - \Delta^{n-1}\beta(x_i + \delta l)}{\delta l}
\end{align*}
\]

where \((\alpha, \beta)\) is the Hinge kernel and \( n \) is the order of the differential operator. Although many different features can be generated based on the feature network with different \( n \), we work with the \( \Delta^1 \) Hinge feature with a feature dimension of 780.

**Quadruple Hinge (QuadHinge)**: QuadHinge is a powerful feature representation following the JFD principle, which incorporates the curvature information of the contour fragments in the Hinge kernel by computing a fragment curvature measurement (FCM) \( C(F_c) \) for contour fragments [19].
Quill and QuillHinge: The Quill feature [28] is the joint probability distribution $p(\alpha, w)$ of the relation between ink direction $\alpha$ and the ink width $w$ characterizing the writing material properties. The QuillHinge is an extension of the Quill and Hinge, which is the probability of $p(\alpha, \beta, w)$, resulting in a 3D histogram. We use the same parameters of the Quill and QuillHinge as the original paper [28], and the dimensions of Quill and QuillHinge are respectively 1600 and 31200.

Triple Chain Code: The triple chain code feature [263] is based on the chain code on a pixel of the writing contours, which is one of the eight directions where the next pixel is on, denoted from 1 to 8.

$$\text{TCC}(x_i, x_{i+l}, x_{i+2l}) = [\text{CC}(x_i), \text{CC}(x_{i+l}), \text{CC}(x_{i+2l})]$$ (2.3)

where $\text{CC}(x_i) \in \{1, 2, \ldots, 8\}$ is the chain code value on position $x_i$, and $l$ is the Manhattan distance along the writing contours. We take the same value of $l = 7$, similar to the CoHinge feature. The feature dimension is 512.

Cloud Of Line Distribution (COLD): COLD is a curvature-free feature designed with the fact that writing contours can be approximated by a set of line segments obtained by the sequential polygonization algorithm [263] and the lengths and
orientations of these straight lines can capture the handwriting styles. The high-ordered curvature points on the writing contours are obtained using the method [219], denoted by \( \mathcal{P} = \{ p_i(x_i, y_i), i = 0, 1, 2, \ldots, n \} \), where \((x_i, y_i)\) is the coordinate of the point \( p_i \) (see Figure 2.5). The line segments can be obtained between any pair of the dominant points \((p_i, p_{i+k})\), where \(k\) is the parameter that denotes the distance on the dominant sequence \( \mathcal{P} \). Each line can be measured by a pair \((\theta, \rho)\) in the polar coordinate space, where \(\theta\) is the line orientation and \(\rho\) is the line length. All the lines in a given handwritten document can form a distribution in the polar coordinate space and can be quantized into a log-polar histogram inspired by the Shape Context [18]. The features obtained with \(k = 1, 2, 3\) in the log-polar space with the radius 7 and the angular intervals 12 are concatenated into one feature vector with the dimension: \(7 \times 12 \times 3 = 252\).

**Junction features:** Junclets [120], a grapheme-based feature, is the stroke-length distribution in every direction from 0 to \(2\pi\) around a reference point (see Figure 2.6) inside the ink trace. When the center point lies on the junction points, such as the fork points and high curvature points on the skeleton line of the ink strokes, the corresponding feature is the junction feature, which contains the junction information around the joint point. We have taken the stroke length distribution in 120 directions equidistantly sampled from 0 to \(2\pi\), and the feature dimension of each junction is \(120\).

![Figure 2.6: Left: Co-occurrence patterns on ink contours. Right: An illustration of the stroke-length distribution on a reference point (the blue point in the center). The green rays are the partial length in each direction, and the yellow curve is the distribution of the partial length in the polar space. The red line is the skeleton line of the stroke ink. m is the maximum measurable stroke length [115].](image)

2.3.1.3 **Identification methodology**

Writer identification is simply answering the who question. For a query document \(Q^\text{script}_i\), where \(\text{script}_i\) is the script of the handwritten manuscript, and \(\text{scribe}A_x\) is the writer which we want to identify, all the documents in the database \(dss^\text{script}_i\) in \(DSS^\text{script}_i\) are sorted according to the feature distance between \(Q^\text{script}_i\) and \(dss^\text{script}_i\), to produce a hit-list where the writer of the top document is assigned to \(\text{scribe}A_x\). Here \(\text{scribe}A_x\) is the label of all the writers, and for our case, \(\text{script}_i\) is a single script of Hebrew. The nearest neighbor classification method is performed
using the leave-one-out [28, 263] strategy. We take the query document out and sort
the remaining documents according to their distance function to an output hit list.
For the distance function of the feature vectors, we have taken the $\chi^2$ (chi-squared)
distance for its better performance [37].

2.4 RESULTS

In this section we present the performance of writer identification based on the
features and methodology explained in section 2.3.1. 447 FragmROIs were used
for the pilot test. In the first set we took 323 FragmROIs with 13 writers labelled
from scribeA001 to scribeA013 having 74, 33, 14, 13, 26, 37, 58, 3, 25, 24, 4, 10 and
2 FragmROIs respectively. The first set consists of writers with a large number of
characters in their corresponding FragmROIs.

We first calculated the feature vectors for all the FragmROIs. Then we performed
the writer identification using the methodology explained in 2.3.1.3. We produce the
output hit list of all the FragmROIs sorted out in accordance with their distance to
the input FragmROI. The top-$n$ performance is calculated when the query FragmROI
is recognized as the writer of the FragmROI on the top $n$ of the hit list. For example,
the top-10 hit list signifies the overall percentage of finding the same writer as input
within the first ten candidates (shortest distance) of the output hit list. Similarly, the
top-1 means the top-most candidate in the output hit list corresponds to the same
writer as the input. The performance of the top-1 and top-10 hit-list for the first set is
presented in Table 2.1. According to the majority palaeographic opinion, scribeA001

<table>
<thead>
<tr>
<th>Feature</th>
<th>Top-1</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hinge</td>
<td>87.61</td>
<td>97.83</td>
</tr>
<tr>
<td>CoHinge</td>
<td>81.11</td>
<td>95.97</td>
</tr>
<tr>
<td>$\Delta^1$Hinge</td>
<td>79.87</td>
<td>94.73</td>
</tr>
<tr>
<td>QuadHinge</td>
<td>89.47</td>
<td>96.59</td>
</tr>
<tr>
<td>Quill</td>
<td>80.80</td>
<td>93.80</td>
</tr>
<tr>
<td>QuillHinge</td>
<td>76.78</td>
<td>89.78</td>
</tr>
<tr>
<td>TripleChainCode</td>
<td>84.82</td>
<td>96.59</td>
</tr>
<tr>
<td>COLD</td>
<td>82.35</td>
<td>94.42</td>
</tr>
<tr>
<td>Junclet</td>
<td>81.42</td>
<td>95.04</td>
</tr>
</tbody>
</table>

Table 2.1: The top-1 and top-10 performance (in percentage) of writer identification for 13
scribes from scribeA001 to scribeA013.

and scribeA002 are the same scribe, and so also scribeA003, 004, 005, 008, 009 and
scribeA010, 012, which are then labelled as scribeB001, B002 and B003 respectively.
The result is presented in Table 2.2 for these seven scribes. Then we took the second
Table 2.2: The top-1 and top-10 performance (in percentage) of writer identification for 7 scribes: *scribeB001, B002, A006, A007, B003, A011* and *A012*.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Top-1</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hinge</td>
<td>92.26</td>
<td>98.76</td>
</tr>
<tr>
<td>CoHinge</td>
<td><strong>93.80</strong></td>
<td>97.52</td>
</tr>
<tr>
<td>(\Delta^1\text{Hinge})</td>
<td>90.71</td>
<td>96.28</td>
</tr>
<tr>
<td>QuadHinge</td>
<td>93.50</td>
<td>96.90</td>
</tr>
<tr>
<td>Quill</td>
<td>88.54</td>
<td>96.28</td>
</tr>
<tr>
<td>QuillHinge</td>
<td>88.85</td>
<td>96.90</td>
</tr>
<tr>
<td>TripleChainCode</td>
<td>92.26</td>
<td>98.14</td>
</tr>
<tr>
<td>COLD</td>
<td>91.33</td>
<td>96.28</td>
</tr>
<tr>
<td>Junclet</td>
<td>56.03</td>
<td>88.85</td>
</tr>
</tbody>
</table>

set of 13 scribes with 124 FragmROIs. The amount of text is lower in this set than in the first one. The result is shown in Table 2.3. Finally, we took all the scribes together for testing. Table 2.4 presents the result of these 20 scribes together (i.e., *scribeB001, B002, B003, A006, A007, A011, A013, A014* to *A026*). We briefly discuss the results and our propositions in the next section (Section 2.5).

Table 2.3: The top-1 and top-10 performance (in percentage) of writer identification for another 13 scribes from *scribeA014* to *scribeA026*, with limited text fragments.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Top-1</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hinge</td>
<td>61.90</td>
<td><strong>92.06</strong></td>
</tr>
<tr>
<td>CoHinge</td>
<td>62.69</td>
<td>89.68</td>
</tr>
<tr>
<td>(\Delta^1\text{Hinge})</td>
<td>43.65</td>
<td>85.71</td>
</tr>
<tr>
<td>QuadHinge</td>
<td><strong>63.49</strong></td>
<td>90.47</td>
</tr>
<tr>
<td>Quill</td>
<td>48.38</td>
<td>89.51</td>
</tr>
<tr>
<td>QuillHinge</td>
<td>45.16</td>
<td>74.19</td>
</tr>
<tr>
<td>TripleChainCode</td>
<td>61.90</td>
<td>88.09</td>
</tr>
<tr>
<td>COLD</td>
<td>58.87</td>
<td>88.71</td>
</tr>
<tr>
<td>Junclet</td>
<td>31.45</td>
<td>76.74</td>
</tr>
</tbody>
</table>
Table 2.4: The top-1 and top-10 performance (in percentage) of writer identification for all 20 scribes.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Top-1</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hinge</td>
<td>78.30</td>
<td>94.40</td>
</tr>
<tr>
<td>CoHinge</td>
<td>79.19</td>
<td>89.93</td>
</tr>
<tr>
<td>$\Delta^1$Hinge</td>
<td>68.23</td>
<td>85.48</td>
</tr>
<tr>
<td>QuadHinge</td>
<td>79.64</td>
<td>89.04</td>
</tr>
<tr>
<td>Quill</td>
<td>71.58</td>
<td>86.57</td>
</tr>
<tr>
<td>QuillHinge</td>
<td>69.57</td>
<td>82.10</td>
</tr>
<tr>
<td>TripleChainCode</td>
<td>79.19</td>
<td>91.28</td>
</tr>
<tr>
<td>COLD</td>
<td>76.95</td>
<td>88.37</td>
</tr>
<tr>
<td>Junclet</td>
<td>49.71</td>
<td>72.93</td>
</tr>
</tbody>
</table>

2.5 DISCUSSIONS

2.5.1 Performance evaluation

We presented the results of writer identification on a limited number of scribes. Of the shape-based methods, the QuadHinge performs the best in the top-1 hit-lists for three out of four cases (only for the case of seven scribes in Table 2.2, CoHinge performs better with a small difference of 0.30% than QuadHinge), whereas the Hinge feature gives better result in all the top-10 hit-lists. The reason for this performance can be deduced from the design criteria of the features themselves. The Hinge feature takes into account the joint probability distribution of the orientations of legs of two contour fragments from a common end pixel on ink contours, which proves to be a strong identical property for individual scribes of these ancient manuscripts. Additionally, the incorporation of FCM to the Hinge feature following the JFD principle gives the QuadHinge feature a boosted performance.

The directional measurement of the ink-trace width makes the Quill feature, which is quite informative on quill-based medieval scripts, a weak candidate for the DSS. This is due to the uniformity of the ink trace in these documents coming from a probably fairly blunt tip of the ancient writing equipment. Consequently, the QuillHinge fails to provide higher performance in this test set. The $\Delta^1$Hinge has limited performance, indicating that on Hebrew characters, loss of the angle with respect to the horizontal removes too much of the writer-specific information.

The grapheme-based feature, Junclets, gives lower performance than the cross-script writer identification [120] due to the lower variability in the stroke-length distribution in every direction around a reference point inside the ink of the DSS’ Hebrew characters.
2.5.2 Propositions

The challenges in analyzing the DSS are unique and unprecedented. Using the dedicated features (in 2.3.1.2), we found fast results without lengthy training on the limited labeled data of the DSS. But they are certainly not the best results to be expected. Especially when the amount of data is small with large variability (as in Table 2.3), the performance becomes lower. To overcome this situation, we need to consider a pragmatic approach incorporating several propositions.

1) Statistical modeling can be used in the case of the DSS, where the sample size is low, and there are differences in the scholarly opinion of writers as well. We can use the differences in writing attributes of a set of different manuscripts to build a population model. A writer model can be built using the query manuscript. The classification is then carried out by evaluating the similarity of a further manuscript sample with respect to the models. We can build our provisional model, similar to the work of speaker identification [163], as follows:

\[
\Lambda(d(W_i, W_j)) = \frac{p_b(d(W_i, W_j))}{p_w(d(W_i, W_j))}
\]  

(2.4)

Here, \(d(W_i, W_j)\) is the distance computed from \(W_i\), the query writer, to \(W_j\), the suspected writer. \(\Lambda\) denotes the likelihood ratio over \(d(W_i, W_j)\). The distribution of distances between the suspected writer and the population is denoted by \(p_b(d(W_i, W_j))\), which can be referred to as the between-group distance among writers. \(p_w(d(W_i, W_j))\) is the distribution of distances taken within different instances of the suspected writer (within-group distance). The collection of statistical models [82], analysis of variance (ANOVA), can be used to analyze the within-group and between-group variances of the writers.

2) Another possibility is transfer learning [170]. It starts with the use of pre-trained networks on massive not-labeled handwriting collections. Such networks are trained to reconstruct images over (via) a very limited number of values (hidden units). After training, such a network implicitly knows a lot about historical handwriting in general. In a second stage, such a network is then applied to the DSS, using those hidden unit vectors as feature descriptors.

3) Data augmentation can be utilized in the processing of the DSS. If there is a believable random transformation of the DSS’ text patterns, i.e., one that remains legible by humans, then for each natural sample of a character, a number of \(N\) derived random versions of it may be added to the training set, effectively enlarging the amount of labeled data. Known already in the nineties [13], this was later made popular in handwriting recognition later by the use of hidden Markov models [97, 299].
2.5.3 Conclusions

In this chapter, we have introduced the digital palaeography of the DSS by presenting a pilot experiment, which is part of a pioneering multi-disciplinary project that brings together the natural sciences, artificial intelligence, and the humanities. By introducing the rule to establish ground truths, we performed writer-identification tests using dedicated features on provisionally labeled data. The varying performance of results for different sets of writers led us to the propositions of statistical modeling, transfer learning, and data augmentation for this largely diverse collection of manuscripts.

We consider the results of this chapter as a baseline measurement for our later experiments. We will combine both the aspect of specifically-designed shape features and the Deep Learning methods to produce fresh empirical data for the study of the DSS. Additionally, we will conduct new radiocarbon ($^{14}$C) dating on a number of physical samples of the scrolls. The outcome of $^{14}$C dating will then be subjected to Bayesian statistical methods in combination with the results from the temporal alignment using pattern recognition to reach more accurate and precise dating of the DSS.

In the end, it is evident from this chapter that a sophisticated character extraction technique is paramount for robust and accurate feature calculations, especially on the larger dataset of IAA images compared to Brill images. For the limited number of Brill images, we can carefully crop the region of interest, adjust the intensity threshold and then binarize them using traditional techniques. However, this is not suitable for the extensive IAA collection with diverse image types. Therefore, we need more than intensity-based techniques to perform character extraction robustly. This necessity leads us to the next chapter, where we introduce BiNet, a new binarization technique.

**Acknowledgements for Chapter 2**

The authors (from the original article) would like to thank Ruwan van der Iest (research assistant for the ERC project at the Qumran Institute) for his valuable input in labeling the regions of interest through the Monk system.

The works of this chapter have been supported by an ERC Starting Grant of the European Research Council (EU Horizon 2020): The Hands that Wrote the Bible: Digital Palaeography and Scribal Culture of the DSS (HandsandBible # 640497). Additional support comes from NWO (Netherlands Organisation for Scientific Research) and FWO (the Research Foundation Flanders): Models of Textual Communities and Digital Palaeography of the DSS (# 326-25-001).