Computational Analysis of Scribal Hands in Huntington 200

Berat Kurar-Barakat[1], Daria Vasyutinsky-Shapira[1], Nachum Dershowitz[1], and Judith Olszowy-Schlanger[2]

[1] Tel Aviv University
[2] École Pratique des Hautes Études

Abstract. Identifying scribal hands is a central task in manuscript studies, especially when several scribes write in the same script type and within the same scribal tradition. Traditionally, such attributions are made by scholars through detailed palaeographical and codicological analysis, including the comparison of letter forms, ductus, page layout, ruling, catchwords, corrections, ink, and other material features. While this expert analysis is essential, it is also time consuming and difficult to scale to large manuscript collections. We study this problem through Bodleian Library MS Huntington 200, a medieval Hebrew manuscript copied in 1279 in Oriental non-square script, for which scholarly attributions disagree: the SfarData dataset assigns the manuscript to eleven scribes, whereas Judith Olszowy-Schlanger identifies six principal scribes working collaboratively.

We propose an unsupervised computational pipeline for supporting scribal hand analysis from manuscript images. The method first extracts connected components from each page, treating them as local handwritten forms. These components are encoded using a sparse autoencoder trained without character or scribe labels. The resulting descriptors are then aggregated into page-level representations using a bag-of-visual-words model. Finally, spectral clustering is applied to infer scribal groups, with the number of clusters selected using the silhouette criterion.

On Huntington 200, the proposed method identifies four computational scribal groups. Although the number of groups is smaller than both scholarly attribution systems, the resulting division aligns strongly with Olszowy-Schlanger's six scribe analysis, reaching an adjusted Rand index of 0.98, and shows lower agreement with the eleven scribe SfarData division, with an adjusted Rand index of 0.72. These results suggest that unsupervised component based representation learning can recover meaningful scribal structure and provide independent computational evidence for evaluating competing palaeographical attributions.

Keywords: Computational paleography · Handwriting style features · Script classification.

1 Introduction

The identification of scribal hands is a fundamental task in manuscript studies. It supports the reconstruction of production processes, the study of scribal collaboration, the dating and localization of manuscripts, and the analysis of textual transmission. The task becomes especially difficult when several scribes write in the same script type and within the same scribal tradition. In such cases, the differences between hands may be subtle and may coexist with non-scribal sources of variation, such as ink quality, page condition, layout, ruling, and changes in writing density.

Traditional scribal attribution is carried out through expert palaeographical and codicological analysis. Scholars compare letter forms, proportions, ductus, inclination, spacing, line management, catchwords, corrections, marginal additions, and other material and graphic features. This form of analysis is indispensable, since scribal identity is not determined by shape alone but by the interaction between handwriting, page organization, and manuscript production. At the same time, expert attribution is labor intensive and difficult to scale. Computational methods can therefore provide a useful complementary perspective, especially when they operate without assuming a predefined number of scribes or requiring labelled training data.

In this paper, we study the problem through Bodleian Library MS Huntington 200, a medieval Hebrew manuscript copied in 1279, probably in Egypt, in Oriental non-square script. Existing scholarly attributions of Huntington 200 differ in their granularity. The SfarData database assigns the manuscript to eleven scribes.[3] By contrast, Judith Olszowy-Schlanger's codicological and palaeographical study [12] identifies six principal scribes, arguing that some of the finer distinctions in SfarData correspond not to separate hands but to variation in ink colour, ink quality, or related material factors. In addition, an earlier work by Malachi Beit-Arié [1], suggests seven scribes; the difference between Beit-Arié and Olszowy-Schlanger is that the former suggests that thirteen lines on page 62v and until 63v were written by a separate scribe (scribe 8 of SfarData). In Olszowy-Schlanger's classification, this is part of the work of the scribe no. 4. The original questionnaire, filled in situ for Huntington 200 and available on the SfarData page of the manuscript, is full of corrections and remarks by different researchers; obviously, there was a lot of discussion about this manuscript. The division into eleven hands was, we assume, not suggested by Malachi himself.

We propose an unsupervised component based pipeline for scribal hand analysis. First, connected components (cc) are extracted from page images. In Hebrew manuscript pages, these components often correspond to letters or letter fragments, and therefore provide local units for modelling handwriting morphology. Second, a sparse autoencoder is trained on the connected components without character or scribe labels, producing compact descriptors of local handwritten forms. Third, the component descriptors are aggregated into fixed length page representations using a bag-of-words model [9]. Finally, spectral clustering is applied to the page vectors, and the number of clusters is selected using the silhouette criterion.

Our experiments on Huntington 200 show that the proposed method identifies four computational scribal groups. Although this number is smaller than

[3] https://sfardata.nli.org.il/#/manuscript/0C077

both scholarly attribution systems, the resulting clustering aligns very strongly with Olszowy-Schlanger's six scribe division, obtaining an adjusted Rand index of 0.98. Its agreement with the eleven scribe SfarData attribution is lower, with an adjusted Rand index of 0.72.

2 Related Work

Computational palaeography has often treated scribal attribution as a visual similarity problem, where handwriting samples are represented by image descriptors and then compared, classified, or clustered. A central family of methods is based on the bag of visual words. Wolf et al. [14] used local descriptors, bag of features representations, and discriminative metric learning to identify candidate joins among Cairo Genizah fragments according to their handwriting similarity. Fiel and Sablatnig [8] used visual vocabularies with Fisher vectors for writer identification and writer retrieval, showing how local handwriting descriptors can be encoded into global document level representations. Christlein et al. [2] later described writer identification methods as falling partly into codebook based approaches, where local descriptors such as SIFT are encoded using dictionaries, Fisher vectors, VLAD, or related visual vocabulary representations. Bag-of-visual-words methods have also been used beyond writer identification, for example in computational script classification of medieval manuscripts [11]. These approaches are relevant to palaeography because they do not require explicit segmentation into complete letters; instead, they summarize recurring local visual patterns in the writing.

A second line of work replaces hand crafted descriptors and visual vocabularies with learned representations. Christlein et al. [2]proposed an unsupervised feature learning method for writer identification and retrieval, where SIFT descriptors are clustered and the resulting cluster labels are used as surrogate classes for CNN training. Hosoe et al. [10] used a conditional autoencoder for offline text independent writer identification, showing that autoencoder based representations can be used to model writer specific handwriting information. In Hebrew palaeography, Droby et al. [5,4] introduced deep learning models for classifying medieval Hebrew manuscripts into script types and modes, and the VML-HP dataset provided a benchmark for Hebrew script subtype classification. These works show the move from explicit palaeographic features toward learned image representations, although such models often require labelled data or proxy tasks.

A third relevant direction concerns the estimation of the number of scribes or writers. In ancient Hebrew inscriptions, Faigenbaum-Golovin et al. [6] used image processing and machine learning methods to infer the presence of multiple authors in the Arad corpus, estimating at least six writers. Later work on the Samaria inscriptions estimated that 31 inscriptions were most likely written by two scribes [7]. For the Dead Sea Scrolls, Dhali et al. [3] tested writer identification features on several sets of scribes and compared hand crafted descriptors such as Hinge, QuadHinge, Quill, and COLD. Popovic et al. [13] then applied

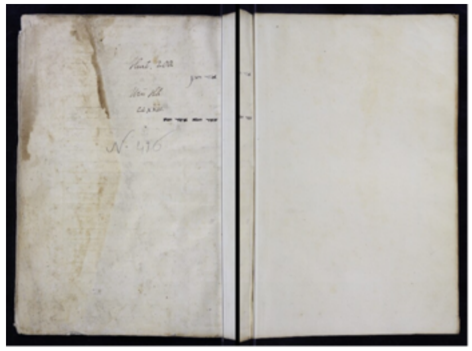

Fig. 1. Opening pages of Huntington 200.

artificial intelligence based writer identification to the Great Isaiah Scroll and reported evidence for two main scribes, with a transition around the middle of the column sequence.

3 Dataset

We evaluate our method on Bodleian Library MS Huntington 200, hereafter Huntington 200, a medieval Hebrew manuscript preserved at the Bodleian Library in Oxford (Figure 1). The manuscript was acquired in Aleppo in the 1670s by Robert Huntington. Although its colophon names Ezra ben Nathanael as the scribe, palaeographical analysis shows that this attribution is only partial: Ezra was one of several scribes who participated in the production of the manuscript. Huntington 200 is therefore a useful case study for computational scribal attribution, since the number and distribution of its scribal hands remain debated.

The manuscript was copied in 1279, probably in Egypt. This localization is supported by palaeographical evidence, including Cairo Genizah fragments written by one of the scribes of Huntington 200 (Figure 2). This scribe is identified as Ezra ben Nathanael. According to Olszowy-Schlanger, Ezra copied a large part of the manuscript and also played an important role in collating and correcting the codex as a whole. His script is more decorative than that of the other scribes, and his role in the manuscript suggests that he was not merely one copyist among others, but a central figure in its production [12].

Codicologically, Huntington 200 is not a luxury manuscript (Figure 3). It is a medium sized codex of approximately 24×16 cm, written on Oriental paper and ruled with a *misṭara*. The number of lines is not consistent across the manuscript and may vary between scribes, and even between the recto and verso of the same leaf. The manuscript was written with iron gall ink, whose colour and quality vary across the codex [12]. In our computational experiments, we use 447 page images from Huntington 200.

Palaeographically, Huntington 200 is written in Oriental non-square script (Figure 4). The manuscript was copied by several scribes who worked separately

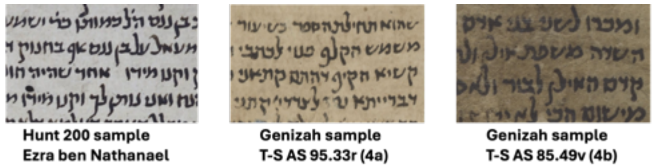

Fig. 2. Comparison between a Huntington 200 sample attributed to Ezra ben Nathanael and two Cairo Genizah samples attributed to the same hand.

Fig. 3. Example pages from Huntington 200, showing the use of Oriental paper, iron gall ink, and ruled page layout.

but toward the production of a single codex. Their contributions are unequal: some scribes copied large portions of the manuscript, while others contributed only a small number of folios.

Two existing attribution systems are used as reference points in this study. The SfarData database assigns Huntington 200 to eleven scribes, whereas Olszowy-Schlanger's study identifies six principal scribes. Since the present experiments operate at page level, pages containing more than one attributed scribe were assigned the label of the majority scribe on that page. The two systems agree on several major divisions (Figure 5) but differ in their level of granularity (Figure 6). In particular, SfarData divides some groups that Olszowy-Schlanger attributes to a single scribe. Olszowy-Schlanger argues that at least some of these finer distinctions are related to ink colour and ink quality rather than to distinct scribal hands (Figure 7).

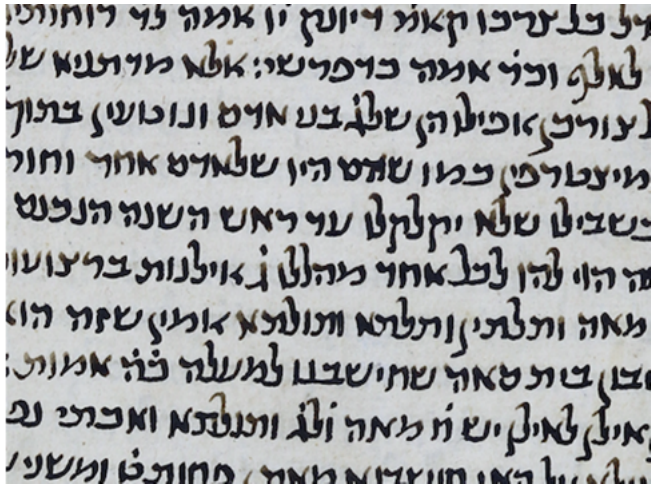

Fig. 4. Example of the Oriental non-square script used in Huntington 200.

4 Method

We investigate whether unsupervised computational analysis can identify meaningful scribal groups in Huntington 200. The aim is not to train a classifier from existing scholarly labels, but to derive a computational scribal division directly from the manuscript images and then compare this division with the attributions proposed by Olszowy-Schlanger and SfarData. Figure 8 illustrates this setting: different attribution systems may assign different scribal boundaries to the same manuscript, while the historical ground truth remains inaccessible.

The proposed pipeline consists of three stages, shown in Figure 9. First, connected components are extracted from the page images and represented by learned visual descriptors. Second, the connected component descriptors from each page are aggregated into a fixed length page representation. Third, the page-level vectors are clustered without using scribal labels or assuming a known scribal division.

4.1 Connected Component Feature Extraction

For local handwriting representation, we use connected components extracted from the manuscript page images. In Hebrew manuscript images, connected components often correspond to individual letters or letter fragments, although they may also include touching characters, punctuation, or noise. We train a sparse autoencoder to extract features from the connected component images, as illustrated in Figure 10. The autoencoder is trained in an unsupervised manner: it receives a connected component image as input, compresses it into a low dimensional latent representation, and reconstructs the original input from this

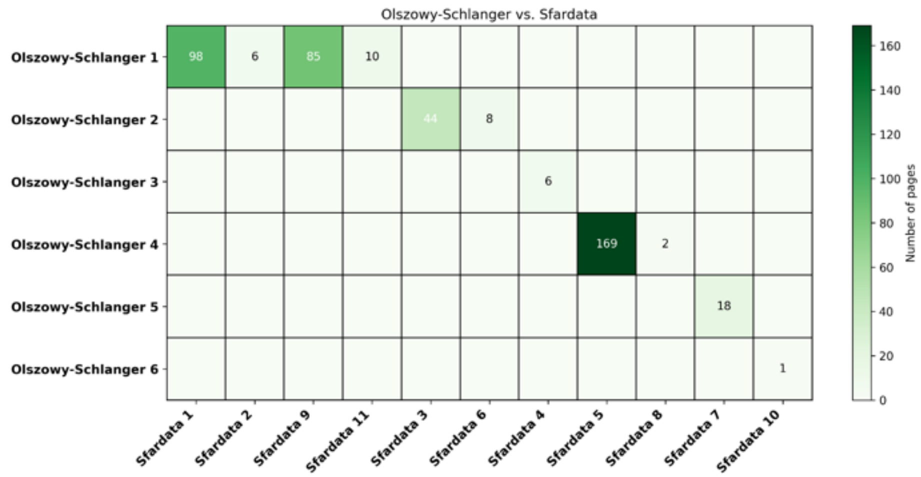

Fig. 5. Correspondence between the Olszowy-Schlanger and SfarData scribal attributions. Each cell gives the number of pages shared by one Olszowy-Schlanger scribe and one SfarData scribe.

representation. After training, the encoder part of the model is used to produce a feature vector for each connected component.

For training the sparse autoencoder, we use approximately 400 connected components per page. Since the dataset contains 447 page images, this gives approximately 180,000 connected components in total. The model is trained on 160,000 connected components and validated on 20,000 connected components. Training is stopped when the validation reconstruction loss reaches 0.0666. Figure 11 shows examples of original connected components and their reconstructions produced by the trained autoencoder.

4.2 Page Level Feature Aggregation

The sparse autoencoder produces a feature vector for each connected component, but each manuscript page contains a variable number of components. To compare pages, these local descriptors must be converted into fixed-length page-level representations. We therefore use a bag-of-visual-words model, as shown in Figure 12. The connected component descriptors are clustered into a visual vocabulary of 1000 words. Each page is then represented by the distribution of visual words appearing among its connected components. This produces one fixed length vector per page, summarizing the local handwritten forms present in that page image.

Figure 13 shows example visual words learned by the bag-of-visual-words-model. Each row contains connected components assigned to the same visual word. The examples indicate that the vocabulary captures recurring local handwritten forms. In some cases, visually similar letter shapes are assigned to dif-

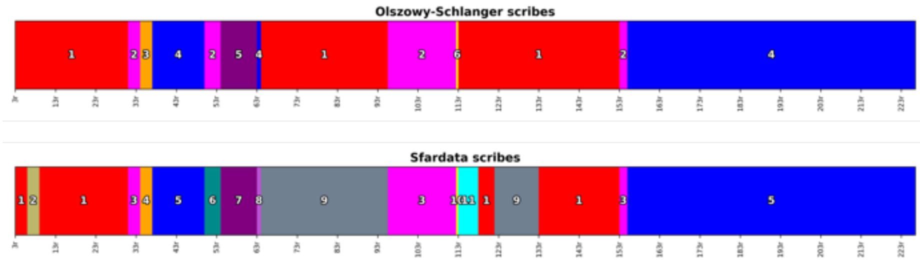

Fig. 6. Barcode comparison of the two scholarly attribution systems. Each vertical bar represents one page. The upper strip shows the Olszowy-Schlanger division, and the lower strip shows the SfarData division.

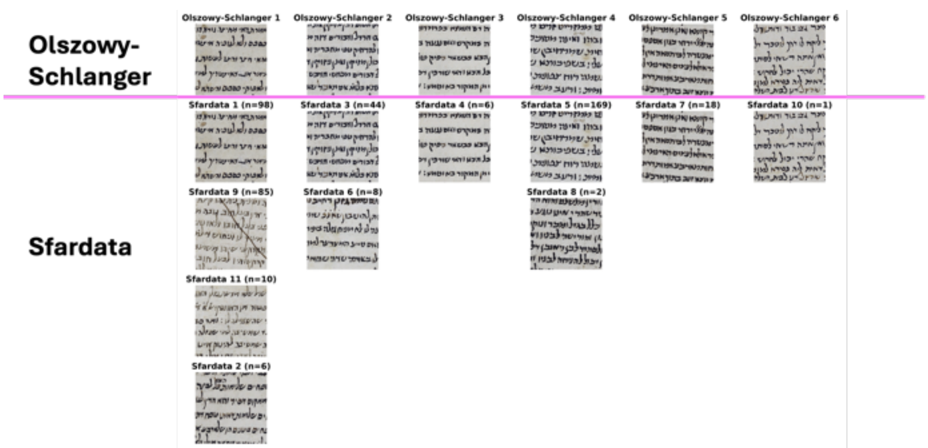

Fig. 7. Sample patches illustrating the relationship between the Olszowy-Schlanger and SfarData divisions. The top row shows samples from the Olszowy-Schlanger scribes, and the corresponding SfarData scribes are shown below.

ferent visual words, suggesting that the representation is sensitive not only to broad letter morphology but also to finer stylistic variation.

The learned page representations are inspected through a three dimensional projection of the feature space, shown in Figure 14. Each point corresponds to one page of Huntington 200 represented by its bag-of-visual-words vector. In this unlabeled visualization, no scribal attribution from Olszowy-Schlanger, SfarData, or the computational clustering is shown. This allows the spatial organization of the page features to be examined before applying any external interpretation.

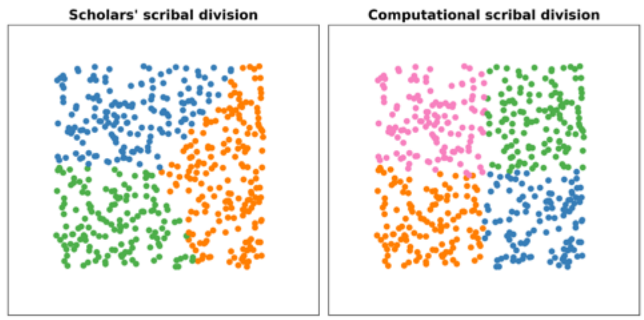

Fig. 8. Illustration of the scribal attribution problem. Scholarly and computational divisions may assign different boundaries to the same manuscript pages, while the historical ground truth is unknown.

Fig. 9. Overview of the proposed pipeline: connected component feature extraction, page-level feature aggregation, and unsupervised clustering of page images.

4.3 Clustering page images

After the bag-of-visual-words stage, each page is represented by a fixed length feature vector. These vectors are clustered in an unsupervised setting, without using the scribal labels proposed by Olszowy-Schlanger or SfarData.

We use spectral clustering, which is appropriate for data that may not form linearly separable groups in the original feature space. Since the number of scribes is not assumed to be known in advance, we evaluate different numbers of clusters. In our experiments, spectral clustering is run for values of k from 3 to 15. For each value of k, we compute the silhouette score, which measures how close each page is to pages in its own cluster compared with pages in the nearest other cluster.

As shown in Figure 15, the highest silhouette score is obtained for $k = 4$. Therefore, the final computational division contains four scribal groups.

5 Results

The results show that the learned page-level features capture a structured distribution of the pages of Huntington 200. Several regions of the projected feature space are compact, indicating that the bag-of-visual-words representations preserve meaningful variation in the local shapes of connected components. We

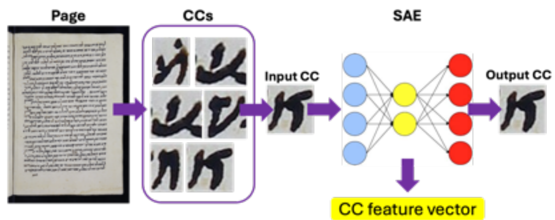

Fig. 10. Feature extraction from connected components using a sparse autoencoder. The encoder maps each connected component image to a compact descriptor.

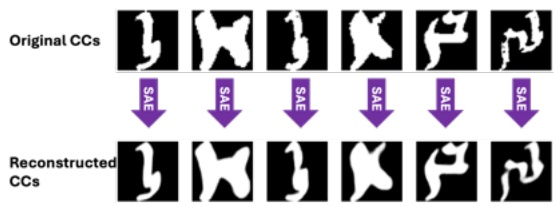

Fig. 11. Examples of connected component reconstruction by the sparse autoencoder. The top row shows input connected components, and the bottom row shows the corresponding reconstructions.

analyze this structure by first visualizing the computational scribal division, and then comparing the same feature space with the scribal attributions of Olszowy-Schlanger and SfarData.

5.1 Computational Scribes

Figure 16 shows the three dimensional projection of the page-level feature vectors colored according to the computational scribal division. Each point represents one page, and the color indicates the scribal group assigned by the clustering method. The visualization shows that several computational scribes form compact regions in the feature space, suggesting that the learned representations separate some scribal groups clearly.

5.2 Olszowy-Schlanger Scribes

Figure 17 shows the same three dimensional page-feature space, but with pages colored according to the scribal attribution proposed by Olszowy-Schlanger. This visualization allows us to examine how the scholarly scribal labels are distributed in the learned computational space. Several of the Olszowy-Schlanger scribes occupy compact regions. For example, Scribe 1 and Scribe 4 form dense groups, indicating a strong correspondence between these scholarly labels and the structure of the learned feature space.

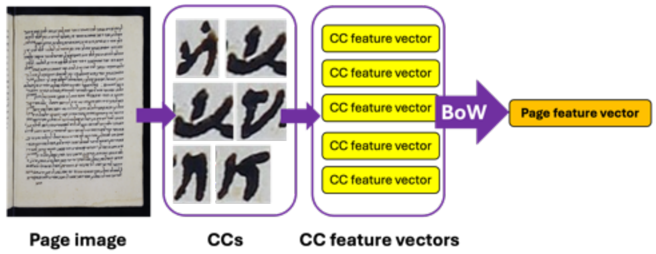

Fig. 12. Aggregation of connected component descriptors into a page-level bag-of-visual-words representation.

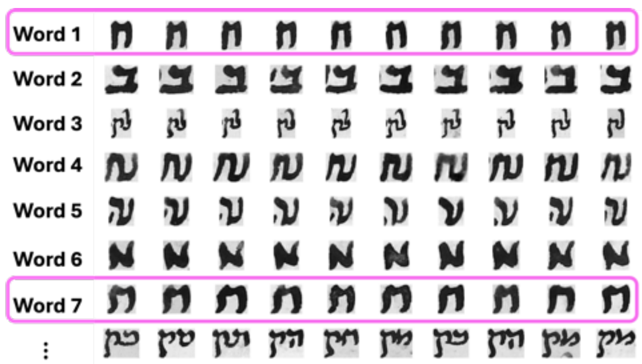

Fig. 13. Examples of visual words learned by the bag-of-visual-words-model. Each row shows connected components assigned to the same visual word.

One notable case is folio 63r, which is attributed by Olszowy-Schlanger to Scribe 4 but appears close in the projected feature space to the region of Scribe 5, especially near folio 55r. Figure 18 compares these two pages visually. Although they belong to different palaeographical scribes, their proximity in the computational space suggests that they share similar local handwritten forms in some connected components.

5.3 SfarData Scribes

Figure 19 shows the page feature space colored according to the scribal labels recorded in SfarData. This comparison highlights the finer granularity of the SfarData attribution. While some SfarData scribes form compact regions, the main difference appears in the large cluster on the right side of the visualization. This region corresponds largely to Olszowy-Schlanger's Scribe 1, but SfarData divides it into several scribes, including Scribes 1, 2, 9, and 11. Olszowy-Schlanger argues that these folios were copied by the same scribe and that the observed differences are limited to the color and quality of the ink. The computational

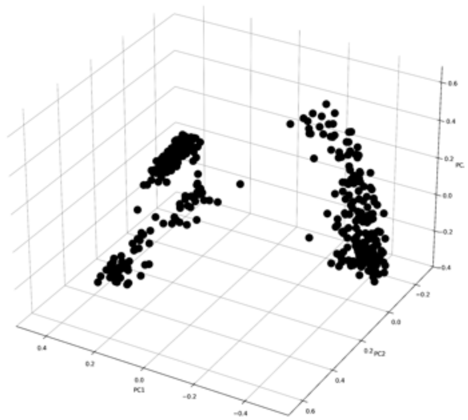

Fig. 14. Three dimensional visualization of the Huntington 200 page feature space. Each black point represents one manuscript page.

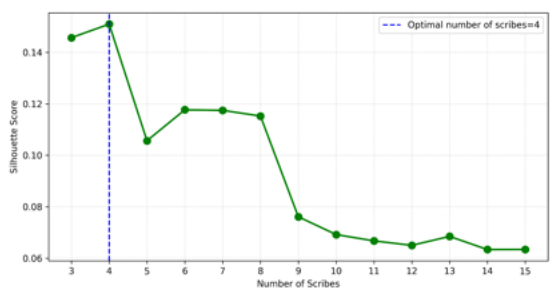

Fig. 15. Silhouette scores obtained for spectral clustering with different numbers of clusters. The highest score is reached at $k = 4$, which is therefore selected as the number of computational scribal groups.

visualization is consistent with this interpretation, since these pages remain close within the same broad region of the feature space.

A representative example is shown in Figure 20. Folio 66r is assigned to SfarData Scribe 9, while folio 10r is assigned to SfarData Scribe 1. Despite this difference in the SfarData attribution, the two pages are located within the same computational region. Their visual comparison suggests that the computational grouping is driven by similarities in local handwritten forms rather than by the finer distinctions introduced in the SfarData labels.

5.4 Olszowy-Schlanger vs. Computational Scribes

We next compare the computational scribal division with the scribal attribution proposed by Olszowy-Schlanger. Figure 21 shows the overlap between the two

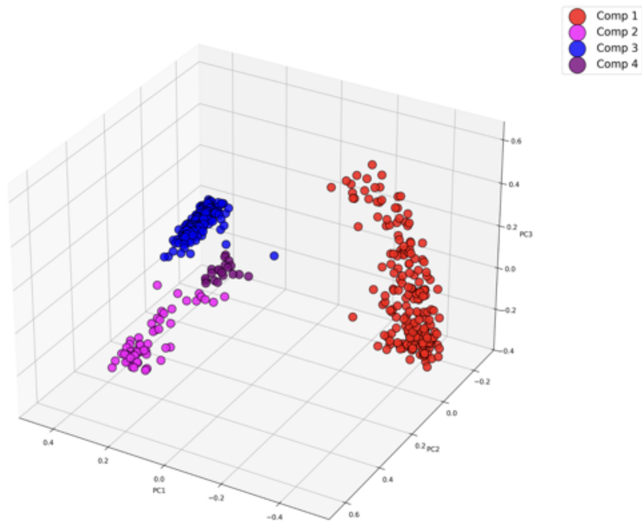

Fig. 16. Three dimensional projection of the Huntington 200 page-level feature vectors colored according to the computational scribal division. Each point represents one page, and the four colors correspond to the four computational scribes.

systems. Each row corresponds to one Olszowy-Schlanger scribe, and each column corresponds to one computational scribe. The value in each cell indicates the number of pages shared by the two assignments. The heatmap shows a strong correspondence between the computational division and Olszowy-Schlanger's attribution. Computational Scribe 1 fully matches Olszowy-Schlanger Scribe 1. Computational Scribe 2 merges Olszowy-Schlanger Scribes 2, 3, and 6, which is expected given that the computational method selected four clusters whereas Olszowy-Schlanger identifies six scribes. Computational Scribe 3 corresponds mainly to Olszowy-Schlanger Scribe 4, while the remaining computational group aligns with the remaining Olszowy-Schlanger attribution. To quantify this agreement, we compute the adjusted Rand index, which measures the similarity between two partitions while correcting for chance. The value between the computational division and Olszowy-Schlanger's attribution is 0.98, indicating very strong agreement despite the smaller number of computational groups.

Figure 22 shows the same comparison as a page order barcode. Each vertical bar represents one page of the manuscript. The upper strip shows Olszowy-Schlanger's attribution, and the lower strip shows the computational division. The barcode confirms the strong agreement observed in the heatmap, with most page ranges assigned consistently across the two systems.

5.5 SfarData vs. Computational Scribes

We also compare the computational scribal division with the scribal labels recorded in SfarData. Figure 23 shows the overlap between the two systems.

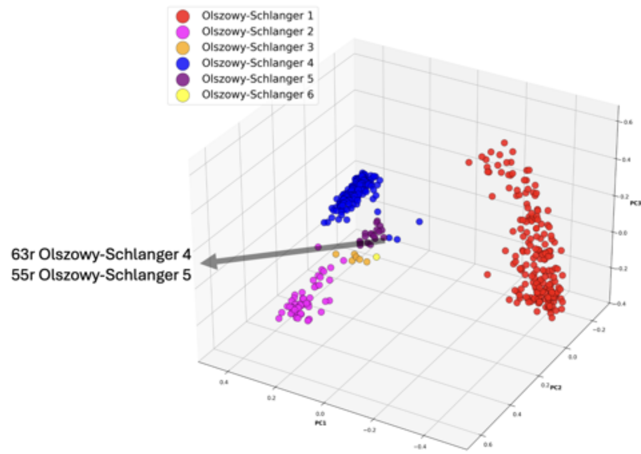

Fig. 17. Three dimensional projection of the Huntington 200 page-level feature vectors colored according to Olszowy-Schlanger's scribal attribution. Each point represents one page, and the colors indicate the palaeographical scribe labels.

Each row corresponds to one SfarData scribe, each column corresponds to one computational scribe, and each cell gives the number of pages shared by the two assignments. The agreement with SfarData is lower than the agreement with Olszowy-Schlanger. A major difference is that Computational Scribe 1 merges SfarData Scribes 1, 2, 9, and 11, corresponding to the broader grouping identified by Olszowy-Schlanger as Scribe 1. This suggests that the computational division does not support the finer subdivision of this group in SfarData. The adjusted Rand index between the computational division and SfarData is 0.72, indicating a weaker correspondence than the comparison with Olszowy-Schlanger.

Figure 24 shows the page order barcode comparison between SfarData and the computational division. The two systems agree over several major manuscript regions, but clear differences appear in the pages corresponding to the finer SfarData subdivisions, especially around pages 63–93 and pages 113–133. These differences are consistent with the heatmap, where several SfarData scribes are absorbed into broader computational groups.

6 Conclusion

We have proposed an unsupervised computational method for exploring scribal grouping in manuscripts like Huntington 200. The method learns local connected component representations using a sparse autoencoder, aggregates them into page -level bag-of-visual-words features, and clusters the pages without using any ground-truth scribal labels. The resulting computational division suggests four scribal groups.

The comparison with scholarly attributions shows that the computational division agrees strongly with Olszowy-Schlanger's palaeographical division, reach-

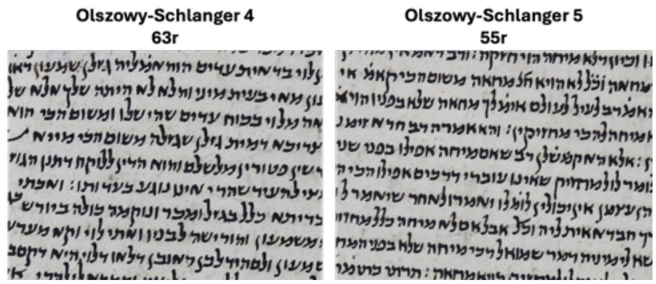

Fig. 18. Visual comparison of folios 63r and 55r. Although Olszowy-Schlanger assigns these pages to different scribes, they appear close to each other in the computational feature space, suggesting similarities in some local letter forms.

ing an adjusted Rand index of 0.98. The agreement with the SfarData division is lower, with an adjusted Rand index of 0.72. This difference suggests that the learned page level representations support the broader scribal grouping proposed by Olszowy-Schlanger more strongly than the finer division recorded in SfarData.

Future work will focus on improving the explainability of the computational grouping. In particular, we aim to identify which visual patterns and connected component forms drive the clustering decisions. At present, the model can group similar connected components together, but it does not explicitly explain whether this grouping is driven by *morphological similarity* or by *stylistic similarity*.

References

1. Beit-Arié, M.: Description of MS Oxford, Hunt. 200 (Catalogue Neubauer n. 416). Kiryat Sefer **53**, 365–367 (1978), (in Hebrew)
2. Christlein, V., Gropp, M., Fiel, S., Maier, A.: Unsupervised feature learning for writer identification and writer retrieval. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). pp. 991–997. IEEE (2017). https://doi.org/10.1109/ICDAR.2017.165
3. Dhali, M.A., He, S., Popović, M., Tigchelaar, E., Schomaker, L.: A digital palaeographic approach towards writer identification in the Dead Sea Scrolls. In: Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods. pp. 693–702. SciTePress (2017). https://doi.org/10.5220/0006249706930702
4. Droby, A., Kurar-Barakat, B., Vasyutinsky-Shapira, D., Rabaev, I., El-Sana, J.: VML-HP: Hebrew paleography dataset. In: Lladós, J., Lopresti, D., Uchida, S. (eds.) Document Analysis and Recognition – ICDAR 2021. Lecture Notes in Computer Science, vol. 12824, pp. 205–220. Springer (2021). https://doi.org/10.1007/978-3-030-86337-1_14
5. Droby, A., Rabaev, I., Vasyutinsky-Shapira, D., Kurar-Barakat, B., El-Sana, J.: Digital Hebrew paleography: Script types and modes. Journal of Imaging **8**(5), 143 (2022). https://doi.org/10.3390/jimaging8050143

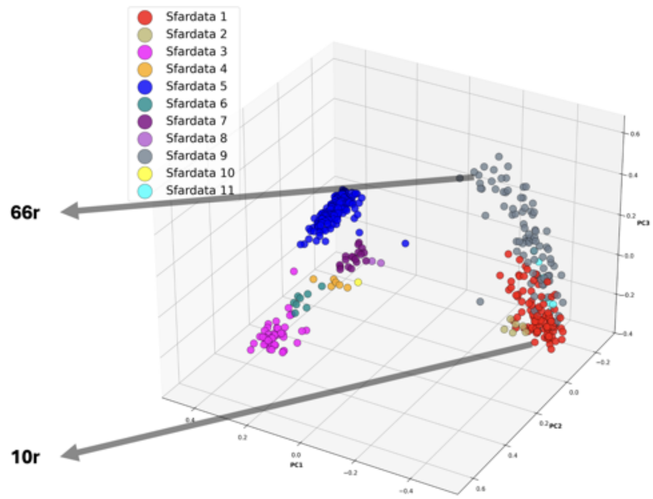

Fig. 19. Three dimensional projection of the Huntington 200 page-level feature vectors colored according to the SfarData scribal attribution. Each point represents one page, and the colors indicate the SfarData scribe labels.

6. Faigenbaum-Golovin, S., Shaus, A., Sober, B., Levin, D., Na'aman, N., Sass, B., Turkel, E., Piasetzky, E., Finkelstein, I.: Algorithmic handwriting analysis of Judah's military correspondence sheds light on composition of biblical texts. Proceedings of the National Academy of Sciences of the United States of America **113**(17), 4664–4669 (2016). https://doi.org/10.1073/pnas.1522200113
7. Faigenbaum-Golovin, S., Shaus, A., Sober, B., Turkel, E., Piasetzky, E., Finkelstein, I.: Algorithmic handwriting analysis of the Samaria inscriptions illuminates bureaucratic apparatus in biblical Israel. PLOS ONE **15**(1), e0227452 (2020). https://doi.org/10.1371/journal.pone.0227452
8. Fiel, S., Sablatnig, R.: Writer identification and writer retrieval using the Fisher Vector on visual vocabularies. In: 2013 12th International Conference on Document Analysis and Recognition. pp. 545–549. IEEE (2013). https://doi.org/10.1109/ICDAR.2013.114
9. Gogawale, S., Grudka, G., Vasyutinsky-Shapira, D., Ventura, O., Kurar-Barakat, B., Dershowitz, N.: Bag of bags: Adaptive visual vocabularies for genizah join image retrieval. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1–9 (2026)
10. Hosoe, M., Yamada, T., Kato, K., Yamamoto, K.: Offline text-independent writer identification based on writer-independent model using Conditional AutoEncoder. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 441–446. IEEE (2018). https://doi.org/10.1109/ICFHR-2018.2018.00083
11. Kestemont, M., Christlein, V., Stutzmann, D.: Artificial paleography: Computational approaches to identifying script types in medieval manuscripts. Speculum **92**(S1), S86–S109 (2017). https://doi.org/10.1086/694112
12. Olszowy-Schlanger, J.: User-production of Hebrew manuscripts revisited: The case of manuscript Oxford, Bodleian Library, Huntington 200. In: Durand-Guédy, D.,

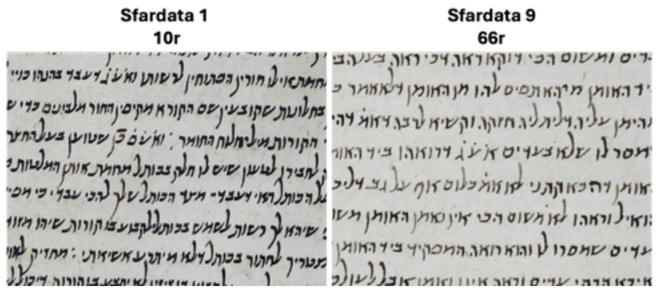

Fig. 20. Visual comparison of folios 66r and 10r. Although SfarData assigns these pages to different scribes, they are located within the same computational region, suggesting similarities in local letter forms.

Paul, J. (eds.) Personal Manuscripts: Copying, Drafting, Taking Notes, pp. 335–358. De Gruyter, Berlin (2023). https://doi.org/10.1515/9783111037196-010
13. Popović, M., Dhali, M.A., Schomaker, L.: Artificial intelligence based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls exemplified by the Great Isaiah Scroll (1QIsa[a]). PLOS ONE **16**(4), e0249769 (2021). https://doi.org/10.1371/journal.pone.0249769
14. Wolf, L., Littman, R., Mayer, N., German, T., Dershowitz, N., Shweka, R., Choueka, Y.: Identifying join candidates in the Cairo Genizah. International Journal of Computer Vision **94**(1), 118–135 (2011)

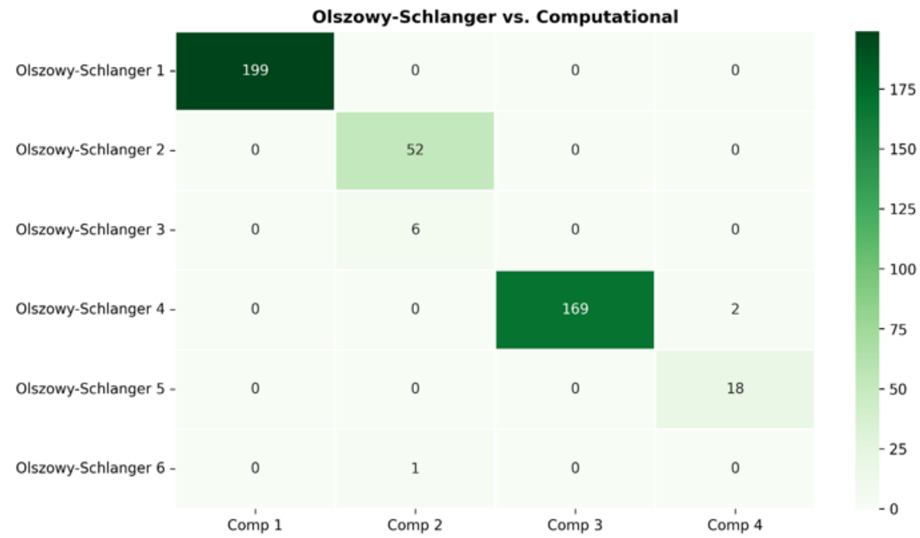

Fig. 21. Overlap between Olszowy-Schlanger's scribal attribution and the computational scribal division. Rows correspond to Olszowy-Schlanger scribes, columns correspond to computational scribes, and each cell gives the number of shared pages.

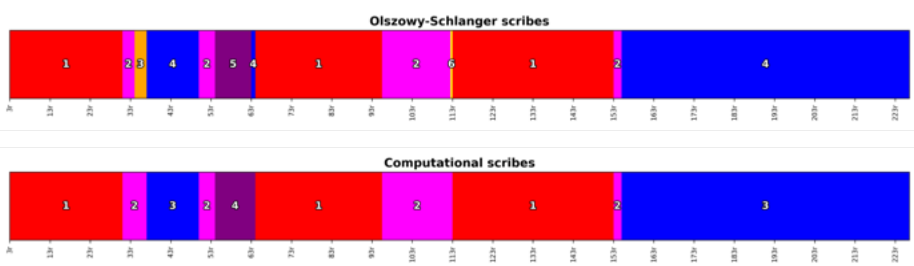

Fig. 22. Barcode comparison between Olszowy-Schlanger's scribal attribution and the computational scribal division. Each vertical bar represents one page; colors encode the assigned scribe label in each system.

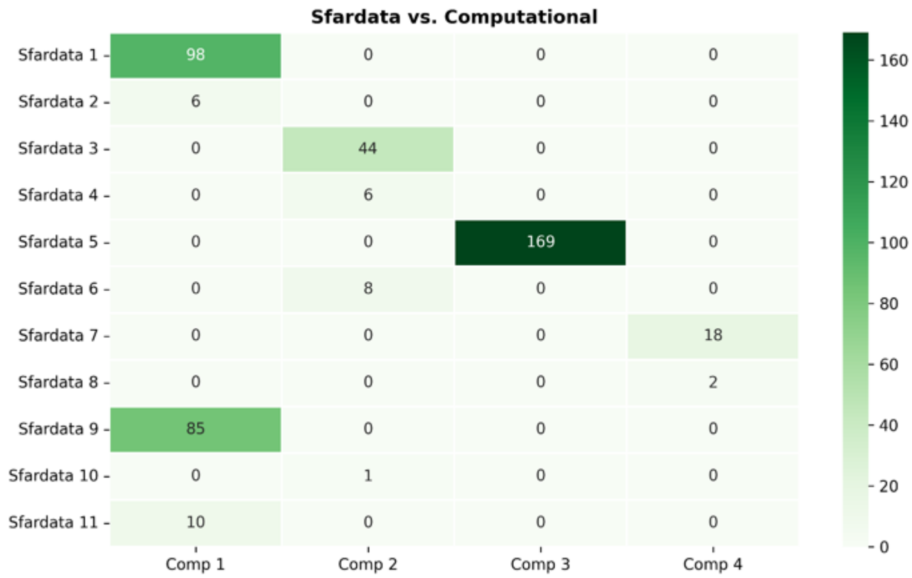

Fig. 23. Overlap between the SfarData scribal attribution and the computational scribal division. Rows correspond to SfarData scribes, columns correspond to computational scribes, and each cell gives the number of shared pages.

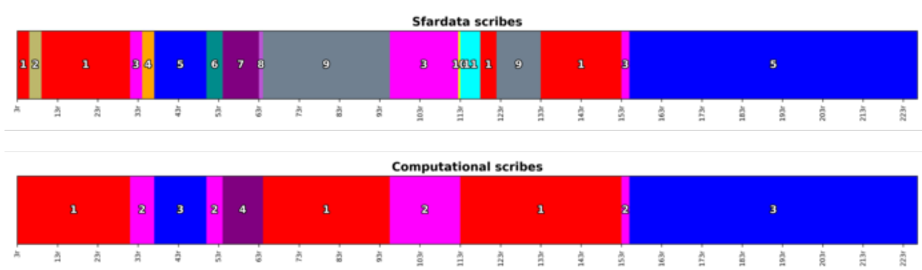

Fig. 24. Barcode comparison between the SfarData scribal attribution and the computational scribal division. Each vertical bar represents one page, colors encode the assigned scribe label in each system.