DCPAPERS

Beyond the Quill: AI-Powered Dating and Typological Analysis of Post-1540 Hebrew Manuscripts

Gila Prebor[1,*]

[1]Department of Information Science, Bar-Ilan University, Israel

Abstract

This project presents an innovative approach to analyzing and dating post-1540 Hebrew manuscripts through artificial intelligence. While traditional dating methods rely on manual paleographic analysis requiring extensive expertise, our research leverages MARC metadata from digital manuscript repositories, particularly the KTIV platform containing over 100,000 Hebrew manuscripts. Building on recent advancements in AI applications for historical manuscripts [1, 5] and Hebrew text analysis [3, 4], we aim to develop systematic methods for manuscript dating and identification of new typological features. Our preliminary analysis of manuscripts from 1600–1700 demonstrates distinct patterns in scribal traditions across different regions and time periods, suggesting the feasibility of AI-powered dating and classification methodologies.

Keywords

MARC metadata, Hebrew manuscripts, artificial intelligence, digital humanities, manuscript dating

1. Introduction

This research proposal aims to develop an artificial intelligence system for the automated dating of undated Hebrew manuscripts produced after 1540. While existing projects have primarily focused on manuscripts up to 1540, a substantial gap remains in the study of later periods.

The foundation of this research is the premise that MARC metadata from Hebrew manuscripts contains distinctive patterns and features that can be effectively analyzed using AI algorithms to determine approximate dating and typological classification. We hypothesize that post-1540 Hebrew manuscripts exhibit systematic variations in their physical and codicological characteristics that correlate with their temporal and geographical origins. The integration of AI analysis of MARC metadata with traditional paleographic knowledge is expected to achieve significantly higher accuracy in manuscript dating than either method alone.

2. Methodology

We use MARC metadata from the KTIV platform as the primary dataset. This metadata includes physical characteristics, scribal features, and provenance details. Our analysis focuses on:

- Supervised learning (e.g., support vector machines) for predicting manuscript dates

DCMI-2025 International Conference on Dublin Core and Metadata Applications

*Corresponding author.

gila.prebor@biu.ac.il (G. Prebor)

0000-0001-6458-0831 (G. Prebor)

https://doi.org/10.23106/dcmi.952549664

- Unsupervised learning (e.g., clustering) for discovering typological groupings
- Statistical analysis and visualization to validate temporal and regional patterns

A preliminary dataset of 1000 dated manuscripts from 1600–1700 was used for initial testing.

3. Research questions

1. How can AI techniques enhance the analysis of MARC metadata for dating post-1540 manuscripts?
2. How can unsupervised learning identify new typological features in manuscript metadata?
3. How can AI algorithms incorporate historical and cultural nuances for accurate dating?

4. Preliminary Results

Our initial analysis revealed:

- A surge in Ashkenazic manuscript production around 1640, with subsequent decline
- Italian scripts maintaining high production levels into the late 17th century
- Distinct subject correlations across script types (e.g., Karaite texts with Karaite script)
- Clear regional patterns, e.g., Yemenite scripts found exclusively in Yemen
- Strong statistical correlation ($r > 0.75$) between certain MARC fields and manuscript dates

These findings confirm the potential of AI models to identify scribal traditions and temporal trends through catalog metadata alone.

5. Contribution and Significance

This project pioneers the application of AI to post-1540 Hebrew manuscripts by leveraging structured MARC metadata. Building on previous applications of AI in manuscript analysis [2], it bridges a crucial gap in Hebrew manuscript studies and offers a replicable framework for large-scale, data-driven dating and typological analysis. By integrating AI methodologies with traditional paleographic scholarship, the project generates innovative tools for cataloging undated manuscripts, reveals insights into regional scribal practices and their historical evolution, and contributes significantly to both Jewish studies and the digital humanities. It also provides practical benefits for libraries and archives and demonstrates how legacy metadata standards such as MARC can be creatively reimagined to support advanced computational analysis.

6. Future Work

Future phases will include expanding the dataset, refining the machine learning models, and developing a user-facing tool for manuscript dating and typology. A key goal is to integrate historical context into the AI models to improve interpretability and scholarly relevance.

https://doi.org/10.23106/dcmi.952549664

References

[1] Dhali, M. A. (2016). Artificial Intelligence in Historical Document Analysis: Pattern recognition and machine learning techniques in the study of ancient manuscripts with a focus on the Dead Sea Scrolls. University of Groningen. https://doi.org/10.33612/diss.869247881

[2] Prebor, G., Zhitomirsky-Geffet, M., & Miller, Y. (2020). A new analytic framework for prediction of migration patterns and locations of historical manuscripts based on their script types. *Digital Scholarship in the Humanities*, 35(2), 441–458.

[3] Suissa, O., Zhitomirsky-Geffet, M., & Elmalech, A. (2022). Toward a period-specific optimized neural network for OCR error correction of historical Hebrew texts. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, 15(2), 1–20.

[4] Suissa, O., Elmalech, A., & Zhitomirsky-Geffet, M. (2023). Optimizing the neural network training for OCR error correction of historical Hebrew texts. arXiv preprint arXiv:2307.16220.

[5] Wahlberg, F., Wilkinson, T., & Brun, A. (2016, October). Historical manuscript production date estimation using deep convolutional neural networks. In *2016 15th International conference on frontiers in handwriting recognition (ICFHR)* (pp. 205–210). IEEE.

[6] Zhitomirsky-Geffet, M., Prebor, G., & Miller, I. (2020). Ontology-based analysis of the large collection of historical Hebrew manuscripts. *Digital Scholarship in the Humanities*, 35(3), 688–719.

https://doi.org/10.23106/dcmi.952549664