Artificial Intelligence in Historical Document Analysis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.
BIBLIOGRAPHY


THE ANALYSIS OF HISTORICAL DOCUMENTS CONTINUES TO HOLD IMMENSE SIGNIFICANCE, EVEN IN THIS MODERN ERA WHERE THE PREVALENCE OF HANDWRITING IS BECOMING LESS COMMON. AS SOCIETY TRANSITIONS TOWARDS DIGITAL COMMUNICATION AND DOCUMENTATION, THE WEALTH OF HISTORICAL RECORDS IN HANDWRITTEN FORM PERSISTS, REPRESENTING INVALUABLE REPOSITORIES OF KNOWLEDGE AND CULTURAL HERITAGE. ARTIFICIAL INTELLIGENCE (AI) AND MACHINE LEARNING (ML) TECHNIQUES ARE PIVOTAL IN DIGITIZING, ANALYZING, AND INTERPRETING THESE DOCUMENTS, ENABLING HISTORIANS, SCHOLARS FROM DIFFERENT DISCIPLINES, AND PALEOGRAPHERS TO UNLOCK INSIGHTS INTO THE PAST USING PREVIOUSLY UNSEEN METHODS. BY UNCOVERING PATTERNS AND CONNECTIONS WITHIN AND BETWEEN DIVERSE DATASETS, THESE TECHNIQUES ARE REVOLUTIONIZING THE STUDY OF HISTORY AND PALAEOGRAPHY.

THE CHAPTERS IN THIS THESIS ARE BASED ON SEVERAL SCIENTIFIC PAPERS PUBLISHED (OR UNDER REVIEW) DURING THE PH.D. TRAJECTORY. THE RESEARCH FOCUSES ON THE PROBLEMS OF WRITER IDENTIFICATION AND DATING OF HISTORICAL MANUSCRIPT IMAGES, ESPECIALLY THE DEAD SEA SCROLLS (DSS). BY USING COMPUTER VISION, MACHINE LEARNING, AND PATTERN RECOGNITION TECHNIQUES, THIS RESEARCH PROVIDES VALUABLE INSIGHTS INTO UNDERSTANDING AND EXTRACTING HANDWRITTEN TEXTS, IDENTIFYING WRITERS, AND DATING HISTORICAL DOCUMENTS. AS PART OF A MULTIDISCIPLINARY PROJECT, THIS THESIS PRESENTS COMPLETE PATTERN-RECOGNITION PROCESSING PIPELINES, PROVIDING NEW EMPirical, QUANTITATIVE FINDINGS ON IMPORTANT ANCIENT IMAGE COLLECTIONS.

A PILOT STUDY ON WRITER IDENTIFICATION ON THE DSS SETS THE STAGE. THIS INITIAL STUDY SHOWS THE APPLICATION OF PATTERN RECOGNITION TECHNIQUES ON DSS, SHOWCASING ITS ABILITY TO DISCERN DISTINCT HANDWRITING STYLES AND POTENTIALLY ATTRIBUTE AUTHORSHIP TO SPECIFIC SECTIONS OF THE SCROLLS BY EXTRACTING TEXTURAL AND ALLOGRAPHIC FEATURES FROM THE HANDWRITTEN TEXTS. THE CONTOUR-BASED JOINT FEATURE DISTRIBUTION PRINCIPLE (JFD PRINCIPLE) IS USED AT THE TEXTURAL LEVEL TO ENCODE ORIENTATION AND CURVATURE INFORMATION OF THE TEXT TO DISCREETLY CHARACTERIZE THE INDIVIDUAL HANDWRITING STYLE. THE WRITER IS FURTHER CHARACTERIZED BY A STOCHASTIC PATTERN GENERATOR OF INK-TRACE FRAGMENTS OR GRAPHEMES AT THE ALLOGRAPH LEVEL. THIS NOVEL STUDY ON DSS ACHIEVES SUCCESS IN WRITER IDENTIFICATION ON A LIMITED NUMBER OF MANUSCRIPTS. THE HINGE SHAPE FEATURE SHOWS PROMISING RESULTS IN ALL THE TOP-10 HIT LISTS OF SEVERAL POSSIBLE-WRITER COMBINATIONS, WITH ACCURACIES FROM 92.06% UP TO 98.76%. THE HINGE FEATURE CONSIDERS 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 SOLID IDENTICAL PROPERTY FOR INDIVIDUAL SCRIBES OF THE DSS.

THE PILOT STUDY SERVES AS A BASELINE MEASUREMENT FOR FUTURE EXPERIMENTS IN THE DIGITAL PALAEOGRAPHY OF DSS. THIS STUDY ALSO REVEALS A CRITICAL NEED FOR IMPROVED BINARIZATION, A SEMANTIC SEGMENTATION PROCESS TO DIFFERENTIATE INK PIXELS FROM BACK-
ground pixels, for extracting characters from the manuscript images. Addressing this challenge becomes instrumental in enabling more accurate writer identification and refining the dating of the documents by extracting features from handwriting only. It is important to note that although many methods can be employed directly using color or grayscale images without binarization, such methods are not preferable for the DSS if one aims at avoiding faulty decisions based on superficial correlations with the texture of the parchment or papyrus instead of the ink trace.

By developing a new optimized binarization process, BiNet, for the DSS, this thesis opens up many doors for further analysis of writer identification and dating techniques. The binarization study explores different traditional intensity-based approaches (Otsu, Niblack, Sauvola, etc.) and deep learning-based models, e.g., CNNs, U-Nets, GANs, and CGANs. Additionally, the study introduces new image-fusion techniques that utilize multi-spectral and full-spectrum color images to produce pseudo-color images with high contrast between ink and background, making the binarization more explicit: The proposed BiNet method vastly outperforms not only the traditional models but also the next-best contender, i.e., with an F-score of 86.7% vs. 68.4% for current state-of-the-art CGANs. This success makes enhanced writer identification possible, leading to the discovery of multiple authors for the Great Isaiah Scroll, one of the longest scrolls in the DSS collection.

In the case of the Great Isaiah Scroll, without the prior assumption of writer identity, based on point clouds of the reduced-dimensionality feature space, the study finds that columns from the first and second halves of the manuscript ended up in two distinct zones of such scatter plots, each addressing very different featural aspects of the script samples. A switching point is found in the column series in a secondary, independent analysis, assuming writer differences and using another independent feature method and several different types of statistical testing. A clear phase transition is apparent in columns 27–29 (out of 54 columns in total). A difference in distance variances is demonstrated such that the variance is higher in the second part of the manuscript. Given the statistically significant differences between the two halves, a tertiary, post-hoc analysis was performed using visual inspection of character heat maps and the most discriminative Fraglet sets in the script. The preferred explanation is that two main scribes were responsible for copying the scroll. This insight challenges previously established assumptions about the scroll. Instead of asking whether traditional palaeography captures everything, this study shows the need for and added value of collaboration between the disciplines for the first time in both qualitative and quantitative ways.

This thesis then moves on to the time period estimation of the DSS manuscripts to set the ground for the final date prediction model. Several dedicated feature-extraction techniques have been explored to extract handwriting styles for different time periods. Additionally, a self-organizing time map is used as a codebook. Support vector regression (SVR) is used to estimate significant periods (Archaic, Hasmonean, and Herodian) based on the feature vector of a manuscript. It archives a Mean
Absolute Error (MAE) of 23.4% (±6.6) years. This study provides a comprehensive overview of the current state of dating the DSS and highlights the remaining problems and obstacles. Once the groundwork is done, the thesis focuses on the final date prediction model.

Enoch, the final date prediction model developed for the DSS collection, was designed by combining AI with radiocarbon dating. The study focused on collecting and preparing data, utilizing neural network architectures for preprocessing and feature extraction, and employing a Bayesian regression approach for date prediction. The limited data were insufficient for fully deploying deep learning in the date prediction task found with tests using the Progressive Neural Architecture Search (PNASNet). Enoch is the first complete AI-based methodology for estimating the dates of manuscripts from their raw image inputs. In contrast to earlier models, Enoch uses a probability-based strategy, utilizing palaeographic input and the entire probability distribution from $^{14}$C output to ensure transparency and interoperability. With a date prediction success rate of 67%, Enoch shows promise when working with completely undated documents. Better image treatment of the test data suggests the possibility of increased success rates. Future research needs to address the problems of sparse labeling and high dimensionality in solving the date predictions using deep learning. It is to be expected that new solutions will appear here because these problems are encountered in many application domains.

The discussions in Chapter 7 suggest several future research directions, including complete page transcription, character reconstruction, material knowledge incorporation, etc. (some exploratory works are added in the appendix). This thesis’s interdisciplinary knowledge fusion with AI will enrich the understanding of the writers’ identities and dating of ancient manuscripts, along with material properties and historical context, leading to enhanced interpretations of the past and opening new doors for future research.
De analyse van historische documenten blijft van grote betekenis, zelfs in dit moderne tijdperk waarin het gebruik van handschrift steeds minder gangbaar is geworden. Terwijl de samenleving overgaat naar digitale communicatie en documentatie, blijft er een rijkdom aan historische documenten in handgeschreven vorm bestaan als waardevolle bronnen van kennis en cultureel erfgoed. Kunstmatige intelligentie (AI) en machine learning (ML)-technieken zijn cruciaal bij het digitaliseren, analyseren en interpreteren van deze documenten. Historici, wetenschappers uit verschillende disciplines en paleografen kunnen nieuwe inzichten over het verleden ontsluiten met behulp van de nieuwe methoden. Met behulp van beeldanalyse kunnen patronen binnen en tussen diverse datasets ontdekt worden. Dit leidt momenteel tot een revolutie in de studie van geschiedenis en paleografie met digitale methoden.

De hoofdstukken in dit proefschrift zijn gebaseerd op verschillende wetenschappelijke artikelen die zijn gepubliceerd (of momenteel worden beoordeeld) tijdens het doctoraal traject. Het onderzoek richt zich op schrijversidentificatie en datering van gescande historische manuscripten, met name de Dode Zeerollen (DZR). Als onderdeel van een multidisciplinair project presenteert dit proefschrift een complete computacionele verwerkingslijn, welke nieuwe empirische, kwantitatieve bevindingen oplevert over belangrijke beeldcollecties uit de antieke historie.

Een pilot-studie naar schrijversidentificatie toont de toepassing van patroonherkennig op de DZR. Hierbij wordt het vermogen gedemonstreerd om verschillende handschriftstijlen te onderscheiden en mogelijk auteurschap toe te kennen aan specifieke delen van de rollen door textuur- en allografische kenmerken van de handgeschreven teksten. Het algemene principe van de methodes berust op het simultaan voorkomen van meerdere vormkenmerken (Joint-Feature-Distribution principe) in het schriftbeeld. Hierbij gaat het om oriëntatie- en krommingsinformatie van de geschreven tekens te coderen (textuur) en om de vormkenmerken van individuele letterfragmenten om een handschriftstijl te karakteriseren. De schrijver wordt beschouwd als een stochastische patroongenerator van inktspoorfragmenten op letter- of grafeen niveau. Deze nieuwe studie over de DZR boekt succes bij het identificeren van schrijvers en is ook toepasbaar indien een beperkt aantal manuscripten beschikbaar is, in tegenstelling tot veel andere huidige methoden. De tevens gebruikte ‘Hinge’-methode laat veelbelovende resultaten zien in alle top-10 schattingen van verschillende mogelijke schrijvers, met nauwkeurigheden van 92,06% tot 98,76%. Deze Hinge-methode is een efficiente methode voor het beschrijven van de lokale kromming op de rand van het inktspoor. De kansverdeling van krommingshoeken is een solide eigenschap voor het identificeren van individuele schrijvers in de DZR.
De pilotstudie dient als nulmeting voor toekomstige experimenten in de digitale paleografie van de DZR. Deze studie onthult ook een cruciale behoefte aan verbeterde binarisatie, een semantisch segmentatieproces om in de hele manuscriptafbeelding inktpixels te onderscheiden van achtergrondpixels, met als doel het beter extraheren van karakters. Deze aanpak speelt een belangrijke rol bij het mogelijk maken van een nauwkeurigere identificatie van de schrijver en het verifen van de datering van de documenten door kenmerken alleen uit het handschrift te halen, en niet uit andere vlekken of patronen op het papier. Het is belangrijk op te merken dat hoewel veel methoden rechtstreeks kunnen worden toegepast met behulp van kleuren- of grijswaardenafbeeldingen zonder binarisatie, dergelijke methoden worden beïnvloed door oppervlakkige correlaties met de textuur van het perkament of papyrus. Doordat deze methoden niet alleen maar letten op de inktsporen, zijn ze meer geneigd om fouten te maken, zeker op de DZR collectie waar het papier ook veel textuur heeft.

Door het ontwikkelen van een nieuw geoptimaliseerd binarisatieproces, BiNet, opent dit proefschrift vele deuren voor verdere analyse van schrijversidentificatie en dateringstechnieken voor de DZR. De binarisatiestudie onderzoekt verschillende traditionele, op intensiteit gebaseerde, benaderingen (Otsu, Niblack, Sauvola, enz.) en op machine learning gebaseerde modellen, bijvoorbeeld CNN’s, U-Nets, GAN’s en CGAN’s. Bovendien introduceert de studie nieuwe beeldfusietechnieken die gebruik maken van multispectrale en traditionele kleurenbeelden om pseudo-kleurenbeelden te produceren met een hoog contrast tussen inkt en achtergrond. Hierdoor wordt de binarisatie explicieter: de voorgestelde BiNet-methode presteert enorm beter dan de traditionele modellen, en ook beter dan de huidige state-of-the-art CGAN, met een F-score van 86,7% voor BiNet versus 68,4% voor CGAN. Dit succes maakt verbeterde identificatie van schrijvers mogelijk, wat heeft geleid tot de ontdekking van meerdere auteurs voor de Grote Jesajarol, een van de langste boekrollen in de DZR collectie.

In het geval van de Grote Jesajarol, constateert de studie dat kolommen uit de eerste en tweede helft van het manuscript in twee afzonderlijke zones van het manuscript terechtkwamen. In een secundaire onafhankelijke analyse, wordt aangetoond dat de berekende statistische kenmerken van het schrift significant verschillen in het tweede deel van het manuscript. Een duidelijke faseovergang is zichtbaar in de kolommen 27–29 (van de in totaal 54 kolommen). Gezien de statistisch significante verschillen tussen de twee helften werd een tertiaire, post-hoc analyse uitgevoerd met behulp van visuele inspectie van karakter-heatmaps en de meest onderscheidende inktfragmenten in het script. De meest aannemelijke verklaring van deze vinding is dat er twee hoofdschrijvers verantwoordelijk waren voor het kopieren van de boekrol. Dit inzicht is in tegenstelling tot eerder gevestigde aanname over de boekrol. In plaats van te beslechten of de traditionele paleografie alle informatie omvat, toont dit onderzoek voor het eerst de noodzaak en meerwaarde van samenwerking tussen de disciplines, zowel kwalitatief als kwantitatief.
Dit proefschrift gaat vervolgens verder met de schatting van de tijdsperiode van de DZR-manuscripten om de basis te leggen voor het inschatten van de definitieve datum. Er zijn verschillende speciale technieken gebruikt voor het vinden van kenmerken in handschriftstijlen met als doel verschillende tijdsperioden te identificeren. Een vector-regressiemethode (SVR) wordt gebruikt om significante perioden (Archaïsch, Hasmonees en Herodiaans) te identificeren op basis van de kenmerken in handschriftstijlen in een manuscript. De schattingen die dit systeem maakt zijn vrij accuraat, met een absolute foutmarge (MAE) van 23.4% (±6.6) jaar. Er volgt een uitgebreid overzicht van de huidige stand van zaken bij het dateren van de DZR en de resterende problemen en obstakels worden belicht. Nu deze basis is gelegd voor het inschatten van een periode, concentreert het proefschrift zich op het inschatten van de exacte datum waarop het werk geschreven is.

Enoch is het dateringsmodel voor de definitieve datum dat is ontwikkeld voor de DZR collectie. Enoch is ontworpen door AI technieken en methoden te combineren met koolstofdatering (¹⁴C-datering). Het onderzoek concentreerde zich op het verzamelen en voorbereiden van gegevens, het gebruik van neurale netwerkarchitecturen voor verwerking van de afbeeldingen, en het gebruik van een Bayesiaanse regressie voor het inschatten van de datum van schrijven. De beperkte gegevens waren onvoldoende om zwaardere zogenaamde deep learning technieken volledig in te zetten, zo blijkt uit tests met behulp van de Progressive Neural Architecture Search (PNASNet) tool. In tegenstelling tot eerdere modellen gebruikt Enoch een op waarschijnlijkheid gebaseerde strategie, waarbij gebruik wordt gemaakt van paleografische input en de volledige waarschijnlijkheidsverdeling van koolstofdatering om inzichtelijkheid en interoperabiliteit te garanderen. Experts in de paleografie geven aan dat de geschatte schrijfdatum voor 67% van de documenten in de DZR collectie accuraat is. Enoch is met deze prestatie veelbelovend bij het werken met volledig ongedateerde documenten. Het is te verwachten dat hier nieuwe oplossingen zullen verschijnen omdat deze specifieke problemen met deep learning in veel toepassingsdomeinen voorkomen.

De discussie in Hoofdstuk 7 beschrijft verschillende toekomstige onderzoeksrichtingen, waaronder volledige paginatranscriptie, karakterreconstructie, integratie van materiële kennis, enz. (enkele verkennende werken zijn toegevoegd in de bijlage). Dit proefschrift toont aan dat de interdisciplinaire combinatie van AI met de kennis van experts meer begrip oplevert over de auteur en schrijfdatum van oude manuscripten. Deze combinatie van AI met expertise over de materiaaleigenschappen en historische context maakt nieuwe en betere interpretaties mogelijk over ons verleden en zal nieuwe deuren openen voor toekomstig onderzoek.
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full form</th>
</tr>
</thead>
<tbody>
<tr>
<td>2DFT</td>
<td>2-Dimensional Fourier Transform</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>AMS</td>
<td>Accelerator Mass Spectrometry</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>BCE</td>
<td>Before the Common Era</td>
</tr>
<tr>
<td>CE</td>
<td>Common Era</td>
</tr>
<tr>
<td>CNN</td>
<td>Convolutional Neural Network</td>
</tr>
<tr>
<td>CO(_3)</td>
<td>Connected Component Contour</td>
</tr>
<tr>
<td>COLD</td>
<td>Cloud Of Line Distribution</td>
</tr>
<tr>
<td>CS</td>
<td>Cumulative Score</td>
</tr>
<tr>
<td>CV</td>
<td>Computer Vision</td>
</tr>
<tr>
<td>DIBCO</td>
<td>Document Image Binarization Competition</td>
</tr>
<tr>
<td>DL</td>
<td>Deep Learning</td>
</tr>
<tr>
<td>DRD</td>
<td>Distance Reciprocal Distortion</td>
</tr>
<tr>
<td>DSS</td>
<td>Dead Sea Scrolls</td>
</tr>
<tr>
<td>DZR</td>
<td>Dode Zeerollen</td>
</tr>
<tr>
<td>EA</td>
<td>Early Aramaic</td>
</tr>
<tr>
<td>ERC</td>
<td>European Research Council</td>
</tr>
<tr>
<td>FAR</td>
<td>False Acceptance Rate</td>
</tr>
<tr>
<td>FCM</td>
<td>Fragment Curvature Measurement</td>
</tr>
<tr>
<td>FCO(_3)</td>
<td>Fragmented-Connected Component Contour</td>
</tr>
<tr>
<td>FRC</td>
<td>French Royal Chancery</td>
</tr>
<tr>
<td>FRR</td>
<td>False Reject Rate</td>
</tr>
<tr>
<td>FWO</td>
<td>Research Foundation Flanders</td>
</tr>
<tr>
<td>GAN</td>
<td>Generative Adversarial Network</td>
</tr>
<tr>
<td>IAA</td>
<td>Israel Antiquities Authority</td>
</tr>
<tr>
<td>JFD</td>
<td>Joint Feature Distribution</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>KN</td>
<td>Kohonen Network</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean Absolute Error</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>MPS</td>
<td>Medieval Palaeographic Scale</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Network</td>
</tr>
<tr>
<td>NWO</td>
<td>Netherlands Organisation for Scientific Research</td>
</tr>
<tr>
<td>OCR</td>
<td>Optical Character Recognition</td>
</tr>
<tr>
<td>PNAS</td>
<td>Progressive Neural Architecture Search</td>
</tr>
<tr>
<td>PR</td>
<td>Pattern Recognition</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>ROI</td>
<td>Region Of Interest</td>
</tr>
<tr>
<td>SOFM</td>
<td>Self-Organizing Feature Map</td>
</tr>
<tr>
<td>SOM</td>
<td>Self-Organizing Map</td>
</tr>
<tr>
<td>SOTM</td>
<td>Self-Organizing Time Map</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>SVR</td>
<td>Support Vector Regression</td>
</tr>
<tr>
<td>TCC</td>
<td>Triple Chain Code</td>
</tr>
</tbody>
</table>
PUBLICATIONS BY THE AUTHOR

JOURNAL ARTICLES (PEER-REVIEWED)

Under peer-review process in - 
DOI: 10.48550/arXiv.1911.07930

Published in - 
Pattern Recognition Letters, 131, 413-420. 
DOI: 10.1016/j.patrec.2020.01.027

Published in - 
PloS ONE, 16(4), e0249769. 
DOI: 10.1371/journal.pone.0249769

Camera-ready processing in - 
Invited submission.

Under peer-review process in - 
Nature (Manuscript ID: 2023-07-12799)
**BOOK CHAPTER (PEER-REVIEWED)**

*Camera-ready processing in - Lecture Notes in Computer Science (LNCS),* Springer Book (Title: Pattern Recognition Applications and Methods), June 2023.  
*Invited submission.*

**CONFERENCE PROCEEDINGS (PEER-REVIEWED)**

ISBN: 978-989-758-222-6, DOI: 10.5220/0006249706930702


**AWARDS RELATED TO PUBLICATIONS**

**Best poster** award 2019  
by the Centre for Data Science and System Complexity (DSSC), Groningen, the Netherlands.

**Best student paper** award 2023  
by the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM), Lisbon, Portugal.
NOTABLE MEDIA APPEARANCES

**BBC News:** AI unlocks ancient Dead Sea Scrolls mystery.
[link, web only]

**ABC Australia:** Who wrote the Dead Sea Scrolls? Digital handwriting analysis and artificial intelligence offer new clues.
[link, web only]

**LiveScience:** Mysterious second writer of Dead Sea Scroll uncovered by AI.
[link, web only]

**Medium:** AI Unravels a Dead Sea Scroll Mystery.
[link, web only]

**WIRED:** AI Helps Prove Two Scribes Wrote Text of a Dead Sea Scroll.
[link, web only]

**DhakaTribune:** Bangladeshi among researchers who unlocked ancient Dead Sea Scrolls mystery.
[link, web only]

**DeMorgen:** Wetenschappers kunnen nu ‘handen schudden’ met Bijbelse schrijvers.
[link, web only]

**El Pais:** Tras las manos que escribieron la Biblia.
[link, web only]

**NVIDIA developer:** Deep Learning Helps Demystify Authorship of a Dead Sea Scroll.
[link, web only]

**Scienmag:** Cracking the code of the Dead Sea Scrolls.
[link, web only]

**The Jerusalem post:** Dead Sea Scrolls: Artificial Intelligence sheds new light on their authors.
[link, web only]
ACKNOWLEDGMENTS

Like many who followed a similar path, my Ph.D. life was also full of surprises, challenges, failures, growths, and, obviously, some incredible experiences. I am deeply grateful for all the amazing people instrumental in this exceptional journey.

I extend my sincere appreciation to my supervisors, Lambert Schomaker and Mladen Popović. Their support, guidance, and belief in my potential have been the cornerstone of this academic journey. Lambert is one of those persons who seem reserved and distant at first. But over time, I realized how amazing of a mentor he is. His vast experience, wisdom, and passion for handwriting analysis always directed our research to a better solution! Discussing ideas with him and receiving alternative views on every possible topic has always been a great experience. Mladen is a people person. His incredible energy and dedication to work always kept me motivated. Mladen and Lambert were always there to extend their assistance on every little issue. I am incredibly fortunate to have them as supervisors and receive their support in all different obstacles during our project, and trust me; we had many!

I thank my thesis assessment committee members—Rolf Ingold, Herbert Jaeger, and Cristoph Brune- for taking the time to read the manuscript, for their enthusiasm, and their insightful feedback. I would also like to thank Hans Plicht, Eibert Tigchelaar, Kaare Lund Rasmussen, and the late Marco Wiering for their support and trust in me throughout this journey. Special thanks to Fokie Cnossen for trusting me and providing me the opportunity to teach and also finish the dissertation.

The support provided by our current and old office secretaries, Elina, Carlijne, Sarah, Jan, Ineke, Magda, and Debbie, has been a consistent source of reassurance. Their willingness to address any query or offer valuable support has created a friendly environment in the institute. I thank Jan for not getting annoyed with the numerous meeting room requests and our friendly chats now and then. I also extend a special thanks to Elina for going above and beyond to help and always being there to listen to every issue and magically solve it right away. Special recognition is due to our system administrator, Remco, who seamlessly handled all computer-related challenges, ensuring a smooth research process.

To my colleagues and friends—Sheng, JP, Olarik, Amir, Harmen, Hamidreza, Ana, Pry, and Asmaa—your camaraderie and shared experiences have made the Ph.D. years memorable. My special gratitude extends to my two paranymphs and friends, Ben and Ayhan, for fulfilling their official roles and being steadfast companions on this academic journey. Ben, with whom I now share my office, had to put with me during the last days of printing, frustrations, and quite some effort in helping with the summary in Dutch. And with Ayhan (and also Gemma), I share many memories,
including roaming the streets of Jerusalem for hours and finding restaurants with good food. Though Drew and Michelle have left Groningen, their impact on my journey remains significant. I appreciate their continued support for me and my family when we needed them.

I want to thank all the students whom I got the chance to supervise during the PhD trajectory, including but not limited to Nino, Daan, Maaike, Thomas, Stephan, Aylar, and Lisa. Working with them improved my skills and experiences. I am especially grateful to Jan Willem, Tobias, and Andrei, who went above and beyond, working overtime alongside me on the DSS project.

To Reehan and Kabir, my two old friends who made the transition to the Netherlands easier, and to Anas and Ridoy, my friends from Belgium and Germany, for their constant motivation—I am grateful for the strength and support you all brought to my life. Fahim Bhai and Tamanna Apu, thank you for always being our guardians in Groningen and offering a sense of familiarity and comfort. Zobaer and Tumpa apu, thank you for always being there for my family and me in the last few years. Furthermore, Morshed Bhai and Kabir bhai thank you for your continuous wisdom and positivity during challenging times. In addition, my heartfelt thanks to my parents-in-law, especially my mother-in-law, for her constant trust and prayers.

My deepest gratitude goes to my parents and sister. I would not be here where I am now without them. My sister, Maisoon, way younger than me, went on to become a doctor before me and encouraged me to finish mine! My father and uncle, Dhali al Mamoon, have always encouraged me to pursue what I like most. Being from Bangladesh, it is sometimes a luxury to let your kids do whatever they want. I am thankful for them. I especially want to mention my mother; all she sacrificed for my comfort and betterment is beyond gratitude. Her steadfast trust in me and daily phone calls were something that always kept me going.

To my wife, Sal Sabil, your continuous support, understanding, sacrifice, and encouragement sustained me through the challenges of this doctoral research. It is never easy to raise two kids and a PhD candidate simultaneously. I can not thank you enough for bearing with us. To my two children, Faatimah and Hamzah, born during this Ph.D. journey, your presence brought joy even in the most challenging times. My apologies to both of you for not always being there and for coming late at night!

To each person mentioned and those whose names may not appear but whose influence is woven into every page of this dissertation, thank you for being an integral part of this journey. Your impact extends far beyond the academic realm. I am profoundly grateful for the collective support that has made this achievement possible.