



# Opening Ottoman Miniature Collections Through Computational Access: Methodological Lessons for Digital Humanities

DISCUSSION PAPER

ESRA CANSU ÇAN BETON

ZEHRA TAŞKIN

\*Author affiliations can be found in the back matter of this article

ju[ubiquity press

## ABSTRACT

Ottoman miniature collections represent a layered cultural heritage that combines artistic, historical, social, and intellectual dimensions. Yet, access to their content is often limited by catalogue records that provide only minimal descriptive information. This discussion paper explores how computational access can expand the study and reuse of such collections by presenting a dataset of 435 Ottoman miniatures, enriched with detailed bibliographic metadata and expert-assigned thematic tags. An original multi-level classification scheme was developed. Tagging was carried out through a custom web interface, preparing the dataset for analysis through thematic network analysis, supervised learning, and unsupervised learning.

Thematic co-occurrence analysis revealed clusters such as political figures, religious motifs, and urban settings, demonstrating the interpretive value of tagging beyond traditional cataloguing. Supervised learning with MobileNetV2 achieved 63% training accuracy and 49% validation accuracy, showing the model's capacity to identify dominant themes while also exposing the limits of small and imbalanced datasets. Unsupervised clustering produced visually coherent groups, but low Silhouette scores and high Davies-Bouldin indices confirmed that thematic boundaries were not strongly separable through visual features alone.

The findings highlight the need to combine computational methods with expert knowledge. While supervised approaches support targeted classification and unsupervised approaches enable exploratory insights, both benefit from integration with rich metadata. This study demonstrates how Ottoman miniatures can serve as a model for applying machine learning and network analysis to cultural heritage collections, offering transferable lessons for digital humanities research.

## CORRESPONDING AUTHOR:

Zehra Taşkın

Department of Information Management, Hacettepe University, Ankara, Turkey; Scholarly Communication Research Group, Adam Mickiewicz University, Poznan, Poland

[ztaskin@hacettepe.edu.tr](mailto:ztaskin@hacettepe.edu.tr)

## KEYWORDS:

Ottoman miniatures; digital humanities; computational access; supervised learning; unsupervised learning; metadata

## TO CITE THIS ARTICLE:

Beton, E. C. Ç., & Taşkın, Z. (2025). Opening Ottoman Miniature Collections Through Computational Access: Methodological Lessons for Digital Humanities. *Journal of Open Humanities Data*, 11: 55, pp. 1–13. DOI: <https://doi.org/10.5334/johd.381>

## (1) CONTEXT AND MOTIVATION

Ottoman miniature collections are among the most significant visual records of cultural heritage, bridging Europe, Asia, and the Islamic world. Emerging from a synthesis of Persian traditions and Ottoman innovations, these works flourished between the fifteenth and seventeenth centuries as central elements of manuscript culture (Aslanapa, 1999; Atasoy & Çağman, 1974). Miniatures were created in imperial ateliers under the patronage of sultans and statesmen, where artists known as *nakkaş* produced highly detailed depictions of battles, coronations, ceremonies, cityscapes, and scenes of daily life. Beyond their artistic quality, these works served political and cultural purposes by reinforcing dynastic legitimacy, documenting imperial campaigns, and preserving courtly rituals (Atıl, 1987; Duran, 1998).

The cultural impact of Ottoman miniatures extends beyond their original imperial function. As historical documents, they provide insights into architecture, costume, medicine, astrology, geography, and urban planning (Çelik, 2022; Gürsoy, 2013; Yıldırım, 2022). *Matrakçı Nasuh's* panoramic depictions of cities, for example, are still consulted by historians and architects for their detailed representation of urban landscapes (Çetin, 2018). Medical manuscripts illustrate diagnostic and surgical practices, while cosmological and astrological miniatures reflect intellectual traditions that shaped the early modern Islamic world. In this way, miniature painting bridges multiple disciplines, allowing researchers from history, sociology, art history, and the history of science to draw on the same visual sources. UNESCO's recognition of miniature art as an element of Intangible Cultural Heritage in 2020 further emphasizes its lasting global significance (UNESCO Türkiye Millî Komisyonu, n.d.).

Despite this richness, miniature collections often remain hidden behind shallow metadata and limited access points in existing catalogues. Many library records identify only the manuscript title, with little or no description of the visual content. Researchers searching for specific themes, figures, or motifs frequently encounter incomplete or inconsistent entries, which prevents systematic comparison or large-scale analysis (Lang & Ommer, 2021). As a result, the interpretive potential of these collections is constrained, and their reuse in interdisciplinary research is limited.

The rise of digital humanities provides an opportunity to address these challenges. Computational access, defined as enabling computers to retrieve, process, and analyze digital materials in structured ways, offers new opportunities for preservation, exploration, and reuse (Digital Preservation Coalition, 2022). For visual collections such as Ottoman miniatures, computational methods can capture details that escape traditional cataloguing, such as recurring figures, compositional patterns, and thematic associations. Techniques including image classification, clustering, and thematic network analysis allow researchers to reveal hidden relationships, group visually similar works, and trace cultural or stylistic continuities across time (Bishop, 2017; Doğan & Türkoğlu, 2019).

Internationally, projects in digital art history and cultural heritage informatics have demonstrated the potential of machine learning for visual analysis (Crowley & Zisserman, 2016; Seguin et al., 2016). Convolutional neural networks and other deep learning approaches have achieved notable success in classifying paintings, detecting stylistic features, and mapping iconographic motifs across centuries (Krizhevsky et al., 2012; LeCun et al., 2015). Yet, applications to Ottoman miniature collections remain scarce. In Turkey, digital initiatives have focused largely on digitization and cataloguing, with less emphasis on reusability, interoperability, and computational modelling (Yıldırım, 2022). Systems such as the Digital Library, Kâşif, and the Manuscripts Database of the Turkish Manuscripts Institution provide important infrastructure but do not yet support the kind of structured access needed for large-scale computational analysis.

This discussion paper responds to these gaps by reflecting on the methodological challenges and opportunities of enabling computational access to Ottoman miniature collections. It draws on a dataset of 435 miniatures enriched with content metadata, and evaluates supervised and unsupervised machine learning approaches for classification, alongside thematic network analysis to map conceptual relationships. The aim is not only to test technical performance but also to distill methodological lessons for digital humanities. The discussion highlights how computational tools can complement traditional cataloguing, how metadata design influences analytical outcomes, and how cultural heritage collections can be reimagined as reusable data resources. By framing Ottoman miniatures as a case study, the paper contributes to wider debates on how computational access can preserve, open, and reinterpret cultural heritage in the digital age.

## (2) DATASET DESCRIPTION

### REPOSITORY LOCATION

<https://doi.org/10.6084/m9.figshare.29979262> (Beton & Taşkın, 2025)

### REPOSITORY NAME

Figshare

### OBJECT NAME

Dataset of Ottoman Miniatures with Metadata and Expert Tags

### FORMAT NAMES AND VERSIONS

- CSV (UTF-8 encoded): MiniatureCollection.csv — main dataset.
- JPEG (.jpg): miniature image files, standardised resolution.
- JSON (Draft 2020-12 schema): classification\_schema.json — hierarchical classification tree with numeric codes and labels.
- TXT (UTF-8 plain text): README files for dataset, codes, and Figshare package.
- Python scripts (.py): code for supervised and unsupervised experiments, including preprocessing, training, evaluation, clustering, and visualization.

### CREATION DATES

2024-01-01–2025-06-15

### DATASET CREATORS

Esra Cansu Çan Beton (Hacettepe University) — data collection, classification, coding, and tagging; Zehra Taşkın (Hacettepe University) — coding, research data management, and supervision.

### LANGUAGE

English

### LICENSE

CC0

### PUBLICATION DATE

2025-08-25

## (3) METHOD

This study followed a multi-stage workflow: corpus construction and metadata design, classification scheme development, manual annotation, supervised and unsupervised analyses on images, and thematic network analysis with co-occurrence visualization.

### (3.1) CORPUS CONSTRUCTION AND METADATA DESIGN

The corpus comprises 435 Ottoman miniature images selected for implementation. All miniatures were sourced from *Metin And's Osmanlı Tasvir Sanatları 1: Minyatir* (2014).<sup>1</sup> This source was chosen for two reasons. First, it presents a wide range of miniatures that represent

---

<sup>1</sup> WorldCat entry available at: <https://search.worldcat.org/title/1030799856> (accessed: 2025-10-22).

different periods, themes, and styles of Ottoman miniature art, which enabled comprehensive testing of the proposed classification and tagging system. Second, the book provides detailed explanations for each miniature, which supported accurate thematic classification and tagging and allowed more reliable analysis of content, context, and historical background. The images were created as standard digital photographs of the reproductions in And's volume, prepared at a consistent resolution to ensure comparability across the dataset. While the reproductions are of high quality, some fine-grained details may not always be fully captured, which could have influenced the clustering results.

Although reproduced in And's volume, the miniatures themselves are public domain artworks held in collections such as Topkapı Palace Museum Library and Chester Beatty Library. The publisher (Yapı Kredi Yayınları) has confirmed that it does not control reuse rights for these images; accordingly, the public domain status applies to the miniatures, while the dataset's metadata, classifications, and compilation are released under a CC0 license.

Accordingly, the dataset was prepared by considering not only the images but also the accompanying descriptions. The aim was to obtain a qualified and meaningful data corpus that would form the basis of the classification system. For each miniature, detailed bibliographic and content information was collected and recorded to support subsequent classification, tagging, and analysis. The recorded elements include: a systematic identifier for each item (for example, MINYATUR\_001), a short title summarizing the main theme or event depicted, the title of the manuscript to which the miniature belongs, and the holding repository and collection (for example, *Topkapı Palace Museum* or *Chester Beatty Library*). The metadata structure was designed with reference to international standards such as Dublin Core, VRA Core, and CDWA (Baca & Harping, 2022; DCMI Usage Board, 2012; Library of Congress, 2022). These frameworks informed field selection and naming, while the final schema was simplified and adapted to the specific characteristics of Ottoman miniature collections to ensure both interoperability and practical usability.

### (3.2) CLASSIFICATION SCHEME

To design a classification system for Ottoman miniatures, existing academic studies were examined in detail. Literature on the history of miniature art, thematic structures, technical characteristics, color palettes, depicted subjects, and stylistic differences across periods was reviewed. The main reference was Metin And's *Osmanlı Tasvir Sanatları 1: Minyatür* (And, 2014), whose classifications and thematic distinctions guided the preliminary phase of the tagging process. Building on the body of scholarship, a comprehensive thematic classification was created that acknowledges miniatures not only as artistic expressions but also as carriers of knowledge and historical documentation. The hierarchical classification scheme, including numerical codes and category labels, is provided in the Figshare repository together with the dataset (Beton & Taşkın, 2025).

In the first stage, common elements and recurring features across the corpus were identified. Based on these observations, an original classification framework was developed. The system was designed around major subject areas such as historical events, everyday life, and religious themes, as well as distinctive figure types. It was intended to reflect the authenticity of the artworks while also supporting computational access and reuse.

The classification scheme consists of eight main thematic categories:

1. **Literary subjects:** Miniatures depicting mesnevis, epics, or scenes related to literary works and authors.
2. **Historical subjects:** Scenes from the history of religion, wars, political events, and general historical episodes.
3. **Metaphysical and religious subjects:** Depictions of sacred texts, religious figures, and metaphysical events.
4. **Science:** Representations of medicine, astronomy, and scholars in these fields.
5. **Nature, city, space, and daily life:** Scenes of urban life, natural phenomena, plants, animals, and everyday practices.
6. **Art and craft:** Miniatures related to painting, music, and the activities of artists and artisans, such as *nakkaş* or musicians.

7. **Portraits:** Individual representations of sultans, religious figures, prominent personalities, and men, women, or children from different social groups.
8. **Figures:** Political, religious, or everyday figures in individual or group contexts, including professions and social roles.

Each main category is further subdivided into secondary and tertiary levels. For example, within Historical subjects, the “Wars” category was divided into “Land Battles” and “Naval Battles,” which were then refined into specific types such as “Campaign,” “Siege,” and “Conquest.” This hierarchical design allows the system to move from general to specific, supporting both broad thematic categorization and detailed analysis.

Alongside thematic labels, a numbering system was established to maintain consistency during annotation. Each category and subcategory was assigned a unique numerical code, ensuring integrity across the tagging process and facilitating systematic data processing in later stages.

### (3.3) TAGGING

In the third stage, a dedicated web-based interface was developed to operationalise the classification system (Figure 1). The tagging interface was designed to be user-friendly, allowing miniatures to be uploaded, displayed, and systematically tagged. Within the interface, each miniature could be assigned multiple tags, enabling the direct application of the classification criteria on the portal itself. At the initial stage of annotation, the visual features and accompanying descriptions of each miniature were reviewed in order to determine the appropriate main subject categories.

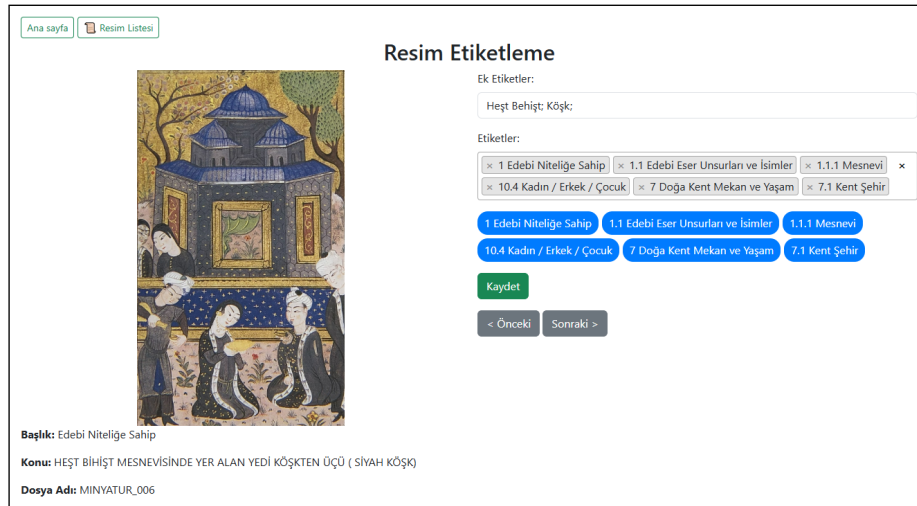


Figure 1 Tagging interface.<sup>2</sup>

Tagging was generally limited to ten tags per miniature, ensuring consistency across the corpus. An additional “supplementary tags” field was created for information that could not be fully captured within the main classification framework. This included proper names, period details, place names, and references to specific objects. For instance, supplementary tags were used to record the names of individuals, dynasties, geographical locations, or artefacts depicted in the miniatures.

In the next step, miniatures were systematically tagged through the web interface according to the hierarchical classification scheme. The process involved assigning each item to categories and subcategories that corresponded to its content. Tagging decisions were made based on direct visual observation of the miniatures, complemented where possible by contextual information contained in their descriptive texts. These descriptions supported the identification of periods, proper names, places, or regional references, thereby enriching the metadata. As a result, each miniature was associated with a detailed metadata profile. For example, the miniature “A Segment of Shopkeepers Passing Through Atmaipazar Square during the Circumcision Feast of Murad III’s Son” (MINYATUR\_048) from the Şehinşehnâme (Topkapı Palace Library), attributed to Seyid Lokman, was classified under the category “Nature, city, space and life.” (see Figure 2). It received tags such as sultan, women/men/

<sup>2</sup> <https://miniature.taskin.web.tr/> (accessed: 2025-10-22).

children, and city, with supplementary tags including feast, hippodrome, and Sultan Murad III. This demonstrates how the classification framework captures not only broad thematic categories but also contextual and historical details, highlighting the intellectual significance of miniatures as visual records of ceremonial and urban life.



**Figure 2** “A Segment of Shopkeepers Passing Through Atmaipazar Square during the Circumcision Feast of Murad III’s Son” (*Şehinşehnâme*, Topkapı Palace Library, attributed to Seyid Lokman).

Finally, the annotated dataset was exported from the tagging interface for subsequent analyses. Frequency analyses, co-occurrence analyses, and category-based distribution analyses were conducted on the exported data. The resulting tag set also served as training input for the supervised machine learning process, enabling an assessment of annotation accuracy. These analyses revealed dominant themes across the corpus, thematic variations across historical periods, and patterns such as trends in color use, thereby demonstrating the added value of systematic tagging for computational access.

### (3.4) ANALYSIS

This study included three types of analysis. First, a thematic network analysis was conducted to identify clusters and co-occurrences across the miniature collection. Second, a supervised learning experiment tested the classification accuracy of tagged miniatures. Third, an unsupervised learning experiment explored clustering patterns in the dataset. Together, these approaches provided complementary perspectives on the dataset and the methodological opportunities of computational access. For the thematic network analysis, only the tag annotations exported from the interface were used. For supervised learning, the image files were paired with their expert-assigned main class labels, while unsupervised learning relied solely on the visual features of the images without metadata.

#### (3.4.1) Thematic network analysis and visualization

The tagged dataset was subjected to a thematic co-occurrence analysis using VOSviewer software. VOSviewer enables the visualization of conceptual relationships by representing terms as nodes and their co-occurrences as edges (van Eck & Waltman, 2017). Node size corresponds to the frequency of a given tag, edge thickness indicates the strength of co-occurrence between two terms, and colors show thematic groups detected by the clustering algorithm. The VOSviewer co-occurrence map serves as a descriptive visualization of the tag distributions and thematic clusters, rather than as a central analytical component of the study. No additional data cleaning

beyond exporting the annotated tags from the interface was required, and reproduction simply involves importing the tag list into VOSviewer in CSV format.

This analysis revealed distinct thematic clusters in Ottoman miniatures, including political figures and portraits, literary and intellectual themes, and representations of daily life. The results highlighted both the internal cohesion of categories and the bridging role of certain motifs, which linked different domains. In this way, network analysis demonstrated how tagging enables the discovery of conceptual connections that are not easily visible through traditional catalogue metadata.

### (3.4.2) Supervised learning

For image classification, the MobileNetV2 deep learning model was applied to the tagged miniature dataset. Training was implemented in Python using the TensorFlow library on the Google Colab platform. Images were divided into 80% training and 20% test sets, automatically distributed across eight thematic categories.

MobileNetV2 was chosen because it is a pre-trained convolutional neural network originally trained on ImageNet, which makes it effective for feature extraction in cultural heritage datasets (Rachburee & Punlumjeak, 2022). In this study, the base layers of MobileNetV2 were retained, while the final layers were replaced with new ones tailored to the miniature classification task.

The training process was run for 10 epochs. Model performance was evaluated through accuracy and loss scores. Classification reports provided further detail on per-class precision, recall, and F1-scores. The supervised pipeline included:

- Assigning images to folders based on theme labels from miniatures.csv.
- Splitting images into training and validation sets using `image_dataset_from_directory` with an 80/20 ratio.
- Training the MobileNetV2 model with the new output layer for eight classes.
- Evaluating accuracy and loss across epochs.

### (3.4.3) Unsupervised learning

The second analysis used unsupervised learning methods to identify clusters of miniatures based on visual similarity. The process was implemented in Python within Google Colab.

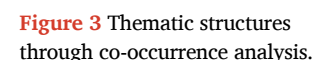
- **Feature extraction:** MobileNetV2, pre-trained on ImageNet, was used to generate feature vectors for each image. These vectors encoded dominant colors, shapes, and structural elements, effectively transforming each image into a numerical representation (Medjahed, 2015).
- **Image hashing:** To compare visual similarity, ImageHash was used. Image hashing generates compact codes representing key image characteristics, enabling fast and efficient comparisons (Akkaya, 2024).
- **Dimensionality reduction:** As each image produced hundreds of numerical features, Principal Component Analysis (PCA) was applied to reduce dimensionality while preserving variance (Yildiz et al., 2010).
- **Clustering:** Initial clustering was performed with k-Means and DBSCAN. To refine results, additional methods were tested, including color histograms for dominant color profiles and a PCA-VGG16 approach, which combines feature extraction with dimensionality reduction for efficiency (Zhang et al., 2024).

Cluster quality was evaluated using the Silhouette coefficient and Davies–Bouldin index. Higher silhouette scores indicated well-defined clusters, while lower Davies–Bouldin scores reflected better separation between groups (Subasi, 2020; Xiao et al., 2017).

## (4) RESULTS AND DISCUSSION

### (4.1) MAPPING THEMATIC STRUCTURES THROUGH CO-OCCURRENCE ANALYSIS

The co-occurrence network produced with VOSviewer provides a visualization of the thematic structures within the miniature collection. The map highlights the most frequently used tags



**Figure 3** Thematic structures through co-occurrence analysis.

**Figure 3** Thematic structures through co-occurrence analysis.

**Figure 3** Thematic structures through co-occurrence analysis.

**Figure 3** Thematic structures through co-occurrence analysis.

**Figure 3** Thematic structures through co-occurrence analysis.

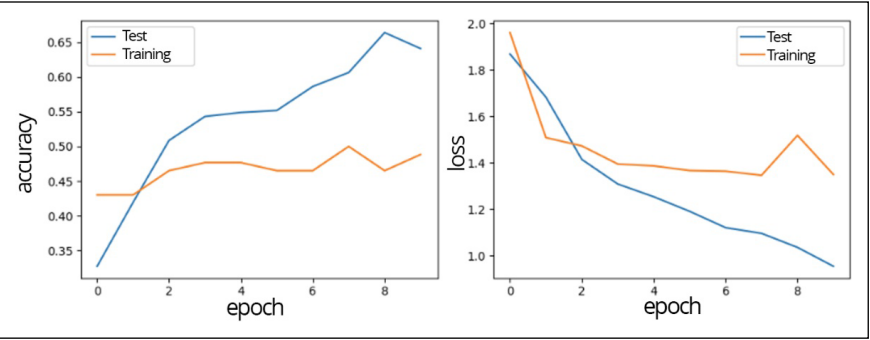
**Figure 3** Thematic structures through co-occurrence analysis.

**Figure 3** Thematic structures through co-occurrence analysis.

| STAGE             | TRAINING ACCURACY | VALIDATION ACCURACY |
|-------------------|-------------------|---------------------|
| Initial (Epoch 1) | 30.5%             | 43.0%               |
| Middle (Epoch 5)  | 57.6%             | 47.6%               |
| Final (Epoch 10)  | 62.5%             | 48.8%               |

**Table 1** Accuracy results of the supervised learning process for training and validation sets.

**Figure 4** Accuracy and loss values during supervised learning training.



The results show that the supervised learning model was able to capture patterns in the miniature images. Training accuracy increased steadily from 30% to over 60%, while training loss decreased from 2.08 to 0.90, indicating that the model was learning meaningful features from the dataset. The improvement in [Figure 4](#) suggests that the thematic variety of miniatures could be distinguished computationally and that themes were represented in a classifiable structure. The proximity of training and validation accuracy during the first six to seven epochs suggests that the model did not overfit at this stage and demonstrated some capacity for generalization.

Validation accuracy, however, did not exceed 50%, which shows that the generalization of the model was limited. The divergence observed at Epoch 9, where training accuracy increased while validation accuracy dropped, indicates the beginning of overfitting. The relatively small dataset, with about 350 training images across ten categories, restricted class representation and reduced the ability of the model to generalize. These limitations also meant that the model’s predictions often diverged from the expert-provided thematic tags, underscoring the importance of qualitative expertise in ensuring accurate classification.

To improve performance, techniques such as data augmentation could be used to increase visual diversity. Enlarging the dataset and balancing class sizes would also support more reliable results in future studies.

#### (4.3) EVALUATING THE PERFORMANCE OF UNSUPERVISED LEARNING

In this stage, miniature images were grouped solely on the basis of visual similarity, without any prior information or category labels. The algorithm was not instructed about the thematic class of each image. Instead, unsupervised learning methods were applied, and clustering experiments were carried out with different algorithms. The results are presented in [Table 2](#) and [Figure 5](#).

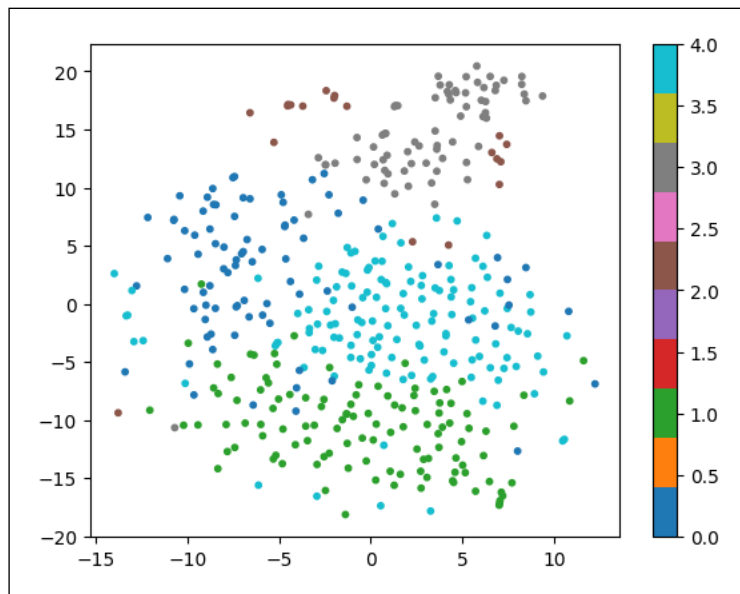
| METRIC               | DESCRIPTION                                           | BEST RESULT |
|----------------------|-------------------------------------------------------|-------------|
| Silhouette score     | Intra-cluster similarity vs. inter-cluster difference | 0.247       |
| Davies–Bouldin index | Cluster separation quality                            | 2.39        |
| Calinski–Harabasz    | Cluster distinction strength                          | 117.5       |

**Table 2** Performance metrics for the unsupervised learning process.

In [Figure 5](#), each point represents an image. Colors indicate the clusters assigned by the KMeans algorithm. Images with similar color palettes, themes, or compositions were placed close to each other in the t-SNE space. Several observations can be made:

- Light blue and green clusters in the lower part of the map contain images with a high degree of similarity. These groups may include miniatures that are thematically or formally distinct but visually consistent.
- Grey and brown clusters at the top consist of fewer images, yet they share distinctive features, suggesting that some themes or visual elements are easier to separate than others.

- The center of the map shows overlapping colors, indicating blurred boundaries between clusters. This suggests that many images share structural features, making clear-cut classification difficult.



**Figure 5** t-SNE visualization of clusters obtained in the unsupervised process.

The proximity of clusters can be explained by recurring elements in nature depictions and daily life scenes. These scenes often include similar color palettes (for example, green, blue, and brown tones), open spaces, and repeated motifs such as trees, people, animals, and architectural elements. The algorithm tended to group such images together, even if they belonged to different thematic categories.

The performance metrics in [Table 2](#) confirm these observations. While some clusters captured coherent similarities, overall the thematic categories did not separate cleanly. This outcome reflects the complexity of miniature art, where visual resemblance does not always correspond to distinct thematic boundaries.

#### (4.4) GENERAL PERFORMANCE EVALUATION

The supervised learning method showed potential for achieving higher accuracy within specific thematic categories, since it allowed the definition of clear targets for classification. During training, the model successfully demonstrated its ability to learn patterns in the data, producing reasonable accuracy in dominant categories such as Nature, City, Space, and Daily Life. However, the large number of classes combined with the limited number of examples reduced the model's generalization power, as reflected in the validation accuracy remaining below 50%.

In contrast, the unsupervised learning approach produced groupings based solely on visual similarity without the use of labels. While some clusters captured thematic proximity effectively, the overall performance was limited. A low Silhouette score (0.247) and a high Davies-Bouldin index (2.39) indicated that separation between clusters was not strong. These results suggest that unsupervised methods can be useful for preliminary analyses such as automatic visual exploration or content recommendation, but conceptually consistent classifications still require expert input.

Taken together, the findings show that both approaches serve different research purposes. Supervised learning can support targeted classification tasks, while unsupervised learning offers exploratory insights. When applied in combination, these methods provide complementary perspectives and can lead to more meaningful outcomes in the computational study of cultural heritage collections.

### (5) IMPLICATIONS/APPLICATIONS

Beyond summarizing our findings, this section outlines new directions for research and reuse that extend the dataset's potential. Ottoman miniatures present a multi-layered visual heritage that combines artistic, historical, social, and cultural dimensions. Systematically classifying, tagging, and analyzing these collections is not only important for their preservation but also

for enabling meaningful reinterpretation in the digital age. This study introduced an original classification system for Ottoman miniatures, applied computational access methods, and compared supervised and unsupervised learning approaches.

A collection of 435 miniatures was created and enriched with detailed bibliographic and content metadata. Tagging was carried out with a multi-level classification scheme, preparing the data for both qualitative and quantitative analysis. Supervised learning with MobileNetV2 achieved 63% training accuracy and 49% validation accuracy, demonstrating the model's capacity to learn but also underlining the need for larger datasets. Unsupervised clustering produced visually coherent groups, yet the metrics confirmed that thematic categories cannot be clearly separated without metadata and expert input. Thematic network analysis further revealed strong conceptual relationships, with tags such as women/men/children, sultan, and city forming central nodes. This shows how computational access can illuminate patterns and connections that remain hidden in traditional catalogues.

The findings highlight several implications. Computational methods extend the role of cataloguing from technical organization toward interpretation and meaning-making. At the same time, algorithms cannot replace human expertise; cultural heritage data requires collaboration between computer scientists, art historians, historians, and cultural heritage specialists. The methodology developed here for Ottoman miniatures can be adapted to other illustrated manuscript traditions, such as Persian, Seljuk, or Western collections, opening possibilities for cross-cultural comparative studies.

To strengthen the discussion of reuse and comparability, this dataset can be situated alongside other digitized collections that have opened significant opportunities for computational analysis. For example, large-scale projects such as the Digital Walters Manuscripts (Walters Art Museum) (*The Digital Walters – Digitized Walters Manuscripts*, 2013), the British Library's Digitised Manuscripts portal (*Digitised Manuscripts and Archives – British Library*, 2025), and the Gallica collection at the Bibliothèque nationale de France (*Our Collection | Gallica*, 2025) all provide high-quality scans of illustrated manuscripts with structured metadata. Methodologically similar initiatives include the Japanese Woodblock Print Portal (*Japanese Print Search and Database*, 2025) and various digital ballad and broadside archives that have enabled systematic computational analysis of cultural artefacts. Integrating Ottoman miniatures into this wider ecosystem would not only augment training datasets for machine learning but also support comparative inquiries across cultural traditions, periods, and artistic genres. Highlighting such parallels underscores how the methods developed here can be adapted to and enriched by other digitized manuscript traditions.

Practical applications also emerge. The classification system and tagging interface could evolve into a digital archive platform where users filter miniatures by theme, figure, period, or context. Such a portal would provide dynamic search and exploration features, supporting both scholarly research and public engagement. The methods can further be integrated into teaching materials, virtual museum projects, or digital storytelling initiatives, allowing miniature collections to reach new audiences.

Future work should expand the dataset by incorporating examples from different libraries, museums, and digital repositories, which would improve model performance and balance class representation. Semi-supervised or active learning approaches could reduce reliance on large volumes of expert-labelled data, enabling more efficient workflows. Open and participatory tagging models involving volunteer communities could also be explored, fostering inclusive cultural heritage practices.

Finally, ethical and legal issues in digitization and open access need careful attention. In preparing this dataset, we confirmed with the publisher that it does not control rights for the miniature reproductions, which are public domain cultural-heritage works held in collections such as the Topkapı Palace Library and Chester Beatty Library. Accordingly, the images and metadata were released under a CC0 license in the Figshare repository. At the same time, future work that incorporates material from additional collections may face restrictions, since some institutions require permissions or impose conditions even for older works in the public domain. Being transparent about such decisions is therefore crucial for sustainable and trustworthy digital archiving strategies.

In sum, this study demonstrates that computational access offers a powerful complement to traditional scholarship on Ottoman miniatures. By combining algorithmic methods with expert

## ACKNOWLEDGEMENTS

This paper is based on the master's thesis of the first author at Hacettepe University, supervised by the second author (Beton, 2025).

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR CONTRIBUTIONS

Esra Cansu Çan Beton: conceptualization, methodology, data collection, data curation, investigation, writing – original draft

Zehra Taşkın: conceptualization, methodology, writing – original draft, writing – review & editing, supervision

## AUTHOR AFFILIATIONS

Esra Cansu Çan Beton  [orcid.org/0009-0003-2258-8490](https://orcid.org/0009-0003-2258-8490)

Department of Information Management, Hacettepe University, Ankara, Turkey

Zehra Taşkın  [orcid.org/0000-0001-7102-493X](https://orcid.org/0000-0001-7102-493X)

Department of Information Management, Hacettepe University, Ankara, Turkey; Scholarly Communication Research Group, Adam Mickiewicz University, Poznan, Poland

## REFERENCES

- Akkaya, E. (2024). *Derin pekiştirmeli öğrenme yöntemi ile görüntü hash kodlarını oluşturma* [Generating image hash codes using deep reinforcement learning method] [Master's thesis]. Sakarya University.
- And, M. (2014). *Osmanlı Tasvir Sanatları 1: Minyatür* [Ottoman Descriptive Arts 1: Miniature]. Yapı Kredi Kültür Sanat Yayınları.
- Aslanapa, O. (1999). *Osmanlı minyatür sanatı* [Ottoman miniature art]. In *Osmanlı* [Ottoman] (Vol. 12, pp. 151–159).
- Atasoy, N., & Çağman, F. (1974). *Turkish Miniature Painting*. Istanbul: Publications of the R.C.D. Cultural Institute.
- Atıl, E. (1987). *The Age of Sultan Süleyman the Magnificent*. National Gallery of Art.
- Baca, M., & Harping, P. (2022). *Categories for the Description of Works of Art (CDWA)*. [https://www.getty.edu/research/publications/electronic\\_publications/cdwa/definitions.pdf](https://www.getty.edu/research/publications/electronic_publications/cdwa/definitions.pdf) (Accessed: 2025-10-22).
- Beton, E. C. Ç. (2025). *Osmanlı dönemi minyatür koleksiyonlarına bilişimsel erişim* [Computational access to Ottoman miniature collections] [Master's thesis, Hacettepe University].
- Beton, E. C. Ç., & Taşkın, Z. (2025). *Dataset of Ottoman Miniatures with Metadata and Expert Tags* [Dataset]. figshare. <https://doi.org/10.6084/m9.figshare.29979262>
- Bishop, C. (2017, March 9). *Against Art History* [Text]. Franklin Humanities Institute. <https://humanitiesfutures.org/papers/digital-art-history/> (Accessed: 2025-10-22)
- Çelik, S. (2022). *XVI. yüzyıl Osmanlı minyatürlerinde astronomi, kozmoloji ve astroloji* [Astronomy, cosmology and astrology on Ottoman miniatures in XVI. century] [Master's thesis, Ankara Yıldırım Beyazıt University].
- Çetin, A. (2018). *Mimaride Minyatürlerin Belge Niteliğinde Kullanımı Üzerine* [On the Use of Miniatures as Documents in Architecture]. *Yedi*, 20, 121–127. <https://doi.org/10.17484/yedi.435020>
- Crowley, E. J., & Zisserman, A. (2016). *The Art of Detection*. In G. Hua & H. Jégou (Eds.), *Computer Vision – ECCV 2016 Workshops* (pp. 721–737). Springer International Publishing. [https://doi.org/10.1007/978-3-319-46604-0\\_50](https://doi.org/10.1007/978-3-319-46604-0_50)
- DCMI Usage Board. (2012). *Dublin Core™ Metadata Element Set, Version 1.1: Reference Description*. <https://www.dublincore.org/specifications/dublin-core/dces/> (Accessed: 2025-10-22)
- Digital Preservation Coalition. (2022). *Computational Access: A beginner's guide for digital preservation practitioners*. <https://doi.org/10.7207/compaccess22-01>
- Digitised manuscripts and archives—British Library. (2025). <https://www.bl.uk/collection/digitised-manuscripts-archives> (Accessed: 2025-10-22).

- Doğan, F., & Türkoğlu, İ. (2019). Derin Öğrenme Modelleri ve Uygulama Alanlarına İlişkin Bir Derleme. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 10(2), 409–445. <https://doi.org/10.24012/dumf.411130>
- Duran, T. (1998). Minyatür ve Gravürlerle Osmanlı İmparatorluğu [The Ottoman Empire in Miniatures and Engravings]. *Tarihi Araştırmalar ve Dokümantasyon Merkezleri Kurma ve Geliştirme Vakfı*.
- Gürsoy, F. (2013). Minyatür Sanatında Dermatoloji [Dermatology in Miniature Art]. *Türk Dermatoloji Dergisi*, 7(2), 116–118.
- Japanese Print Search and Database. (2025). <https://ukiyo-e.org/> (Accessed: 2025-10-22).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C.J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, 25. [https://papers.nips.cc/paper\\_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html](https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html) (Accessed: 2025-10-22).
- Lang, S., & Ommer, B. (2021). Transforming Information Into Knowledge: How Computational Methods Reshape Art History. *Digital Humanities Quarterly*, 15(3).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Library of Congress. (2022). VRA CORE – a data standard for the description of works of visual culture. <https://www.loc.gov/standards/vracore/> (Accessed: 2025-10-22).
- Medjahed, S. A. (2015). A Comparative Study of Feature Extraction Methods in Images Classification. *International Journal of Image, Graphics and Signal Processing*, 7(3), 16. <https://doi.org/10.5815/ijigsp.2015.03.03>
- Our collection | Gallica. (2025). <https://gallica.bnf.fr/accueil/en/html/our-collection> (Accessed: 2025-10-22).
- Rachburee, N., & Punlumjeak, W. (2022). Lotus species classification using transfer learning based on VGG16, ResNet152V2, and MobileNetV2. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 11(4), 1344–1352. <https://doi.org/10.11591/ijai.v11.i4.pp1344-1352>
- Seguin, B., Striolo, C., diLenardo, I., & Kaplan, F. (2016). Visual Link Retrieval in a Database of Paintings In G. Hua & H. Jégou (Eds), *Computer Vision – ECCV 2016 Workshops*, 9913 767. Springer. [https://doi.org/10.1007/978-3-319-46604-0\\_52](https://doi.org/10.1007/978-3-319-46604-0_52)
- Subasi, A. (2020). Chapter 7—Clustering examples. In A. Subasi (Ed.), *Practical Machine Learning for Data Analysis Using Python* (pp. 465–511). Academic Press. <https://doi.org/10.1016/B978-0-12-821379-7.00007-2>
- The Digital Walters—Digitized Walters Manuscripts. (2013). [https://www.thedigitalwalters.org/04\\_TechnicalReadMe.html](https://www.thedigitalwalters.org/04_TechnicalReadMe.html) (Accessed: 2025-10-22).
- UNESCO Türkiye Millî Komisyonu. (n.d.). *UNESCO Somut Olmayan Kültürel Miras Listeleri [UNESCO Intangible Cultural Heritage Lists]*. <https://www.unesco.org.tr/Pages/126/123/UNESCO-Somut-Olmayan-K%C3%BClt%C3%BCrel-Miras-Listeleri> (Accessed: 2025-10-22).
- van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, 111, 1053–1070. <https://doi.org/10.1007/s11192-017-2300-7>
- Xiao, J., Lu, J., & Li, X. (2017). Davies Bouldin Index based hierarchical initialization K-means. *Intelligent Data Analysis*, 21(6), 1327–1338. <https://doi.org/10.3233/IDA-163129>
- Yıldırım, M. (2022). *Kültürel mirasın aktarımında illüstrasyon sanatının önemi ve minyatür sanatıyla karşılaştırılması [The importance of illustration art in the transfer of cultural heritage and its comparison with miniature art]* [Master's thesis, Marmara University]. <https://www.proquest.com/openview/e6479799c89efff20baf4e44a150c561/1.pdf> (Accessed: 2025-10-22).
- Yıldız, K., Çamurcu, Y., & Doğan, B. (2010). Veri madenciliğinde temel bileşenler analizi ve negatif matris çarpanlarına ayırma tekniklerinin karşılaştırmalı analizi [Comparative analysis of principal component analysis and non-negative matrix factorization techniques in data mining]. 207–213. [https://ab.org.tr/ab10/kitap/yildiz\\_camurcu\\_AB10.pdf](https://ab.org.tr/ab10/kitap/yildiz_camurcu_AB10.pdf) (Accessed: 2025-10-22).
- Zhang, Y., Ye, Y.-L., Guo, D.-J., & Huang, T. (2024). PCA-VGG16 model for classification of rock types. *Earth Science Informatics*, 17(2), 1553–1567. <https://doi.org/10.1007/s12145-023-01217-y>

#### TO CITE THIS ARTICLE:

Beton, E. C. Ç., & Taşkın, Z. (2025). Opening Ottoman Miniature Collections Through Computational Access: Methodological Lessons for Digital Humanities. *Journal of Open Humanities Data*, 11: 55, pp. 1–13. DOI: <https://doi.org/10.5334/johd.381>

Submitted: 25 August 2025

Accepted: 11 September 2025

Published: 06 November 2025

#### COPYRIGHT:

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Journal of Open Humanities Data* is a peer-reviewed open access journal published by Ubiquity Press.