ARCHAIA JOURNAL OF SCIENCE AND TECHNOLOGY
VOL. 1, NO. 2, June 2024
https://doi.org/10.48161/ajst.v1n2a85

Machine Learning Pattern Analysis of Visual Arts Metadata: Style–Genre Structure, Baseline Classification, and Cultural Heritage Interpretation

Lukumba Phiri[1*], Hashim H Abada[2]

[1] School of Built Environment Engineering and Technology, Engineering and Technology, University of Kabwe, Kabwe, Zambia;
[2] Mechanical Engineering, College of Engineering, University of Delaware, Delaware, 19716, United States;

*** Corresponding author:** phirilukumba@gmail.com**.**

ABSTRACT: This study develops a metadata-driven machine-learning framework for analyzing visual patterns and organizational structure in a large WikiArt visual-arts corpus. The empirical analysis evaluates 103,250 artwork records across 2,319 artists, 136 style labels, and 42 genre labels, partitioned into an official training set of 79,433 records (76.93%) and a testing set of 23,817 records (23.07%). Missingness analysis indicates that style (0.95% missing) and genre (1.21% missing) are mostly complete, whereas production date exhibits substantial missingness (25.53% missing). The corpus displays strong class concentration: Impressionism (10,643 artworks; 10.41%) and Realism (10,523 artworks; 10.29%) represent the largest style categories, while portrait (16,847 artworks; 16.52%) and landscape (15,006 artworks; 14.71%) dominate genre labels. Normalized entropy measures for style (0.726) and genre (0.740) confirm significant class imbalance. A chi-square association test reveals a statistically significant, moderate relationship between style and genre ($\chi^2 = 494,553.56$, $p < 0.001$, Cramer's $V = 0.345$). Metadata-only baseline modeling using a decision tree achieved a macro F1-score of 0.466 for top-10 style classification and 0.426 for top-10 genre classification, outperforming random baselines. The paper establishes a transparent computational pipeline for dataset profiling, statistical evaluation, baseline modeling, and similarity retrieval in digital cultural heritage management.

Keywords: machine learning; WikiArt metadata; visual arts; cultural heritage; style classification; genre classification; similarity retrieval.

I. INTRODUCTION

Digital cultural heritage collections have reached a stage where the central problem is no longer just access to digitized objects, but the ability to organize, compare, and interpret them at scale. Large online art collections contain thousands of paintings, drawings, prints, and related visual records that carry multiple layers of information: artist attribution, period, style, genre, production date, source, image dimensions, and institutional description. These fields make digital art archives highly valuable for research, education, curation, and public discovery. At the same time, their size and heterogeneity make them difficult to study through manual inspection alone. A curator may examine selected objects in depth, but a large metadata corpus requires computational support to identify structural patterns, detect imbalance, examine label relationships, and prepare classification or retrieval tasks.

Machine learning offers a practical framework for this challenge because it can combine statistical profiling, supervised classification, similarity search, clustering, and visual representation learning. In visual-arts research, the task is not simply to assign labels automatically. More importantly, machine learning

can reveal how a collection is organized, which categories dominate, which labels are underrepresented, which fields are incomplete, and which forms of information are available before image-level modeling begins. Such analysis is especially important when the dataset is intended to support style classification, genre classification, or visual similarity retrieval. A model cannot be evaluated responsibly unless the underlying corpus has first been examined for missing values, class imbalance, source concentration, technical variation, and label overlap.

The present study focuses on a WikiArt metadata corpus containing 103,250 artwork records. The corpus includes artist labels, dates, style labels, genre labels, image dimensions, source indicators, digital-size information, train-test split assignments, and image identifiers. These variables make the corpus suitable for empirical machine-learning research because they support descriptive analysis, label selection, baseline modeling, and preparation for image-based classification. The results show that the corpus is large and analytically rich but unevenly distributed. A limited number of style and genre categories account for a large share of the labeled artworks. This pattern is common in cultural collections because historically prominent styles, widely reproduced artists, and frequently cataloged subject categories are often overrepresented. If you ignore this imbalance, model accuracy may appear high while performance on smaller or culturally specific categories remains weak.

A key strength of the dataset is the presence of both style and genre labels. These variables are related but describe different cultural dimensions. Style generally refers to formal and historical characteristics such as movement, period, technique, and visual language. Genre refers more directly to subject matter and pictorial category, such as portrait, landscape, religious painting, or still life. Their relationship is therefore analytically important. A single genre can appear across many styles, and a single style can include many genres. For this reason, the study treats style and genre as two complementary targets rather than interchangeable labels. The style–genre association analysis supports this distinction and justifies a modeling design that evaluates both classification tasks separately.

The study also distinguishes metadata-based organization from true visual pattern analysis. Metadata can support preliminary classification because fields such as genre, style, source, date, and technical dimensions often carry indirect information about artworks. However, metadata alone cannot fully represent brushwork, palette, composition, spatial arrangement, texture, or iconographic structure. Full visual pattern analysis requires image embeddings extracted from a computer-vision model and then evaluated through classification, similarity retrieval, and clustering. The present paper therefore uses the available metadata to establish a strong empirical foundation and defines the image-based stage as the next modeling layer connected through the artwork identifiers.

This distinction strengthens the paper's academic contribution. Instead of presenting machine learning as a black-box classification tool, the study frames it as a staged research process. The first stage profiles the corpus and identifies what the dataset can support. The second stage evaluates whether existing metadata carries predictive structure. The third stage defines how to add image embeddings for visual similarity and deep classification. The fourth stage connects model outputs with cultural interpretation. This staged design suits cultural heritage research because it separates technical performance from cultural meaning. A model can identify patterns, but interpretation requires art-historical judgment, awareness of cataloging conventions, and sensitivity to the cultural contexts represented in the collection.

Accordingly, this paper aims to develop a dataset-aligned machine-learning framework for analyzing visual arts and cultural heritage collections through the WikiArt metadata corpus. The specific objectives are to profile the dataset, measure missingness and class imbalance, examine style–genre associations, evaluate metadata-only classification baselines, and propose a reproducible pathway for image-based modeling and similarity retrieval. The study is guided by three research questions: how can a large art metadata corpus be prepared for machine-learning analysis; what statistical patterns and limitations are visible before image-level training; and how can classification, similarity retrieval, and cultural interpretation be connected in a transparent research design.

The remainder of the paper is organized as follows. The literature review summarizes studies and situates the present work within research on deep learning, artwork classification, visual similarity, and machine learning for cultural heritage. The methods section describes the research design, dataset variables,

preprocessing, statistical procedures, model equations, proposed contribution, and evaluation metrics. The results section reports corpus composition, missingness, technical image-profile statistics, label distributions, class imbalance, style–genre association, and metadata-only baseline performance. The discussion interprets these findings in relation to digital archive organization and cultural heritage management. The final sections present limitations, future work, and a conclusion that clarifies how the dataset can support both computational modeling and responsible cultural interpretation.

II. LITERATURE REVIEW

The literature on machine learning for visual arts and cultural heritage shows that computational image analysis has gradually moved from basic visual classification toward richer forms of semantic interpretation, metadata enrichment, and archive organization. Early work on cultural image enrichment showed that artificial intelligence can strengthen the descriptive value of heritage collections by linking visual records to structured metadata and semantic information [1]. Studies on artistic style classification further showed that machine-learning models can identify stylistic signals in artworks, although artistic labels remain complex because they combine historical period, visual technique, cultural context, and curatorial convention [2]. Broader reviews of deep learning in paintings and drawings confirmed that visual-art analysis requires more than generic computer-vision methods, since artworks contain symbolic, aesthetic, historical, and technical features that differ from ordinary photographic datasets [3]. Fine-art classification studies also proved that convolutional neural networks can support style, artist, and genre prediction, but model performance depends strongly on dataset quality, label balance, and evaluation design [4].

A major research direction focuses on representation learning and retrieval in art collections. Object retrieval studies in paintings showed that discriminative visual regions can help identify meaningful links between artworks beyond simple label assignment [5]. Painter-classification research using deep autoencoders demonstrated that learned representations can capture artist-related patterns, but it also highlighted the challenge of separating artist identity from style, genre, and period effects [6]. Large-scale machine-vision studies of art history suggest that computational models can reveal broad visual structures across historical art collections, especially when researchers interpret classification results alongside cultural and historical knowledge [7]. Cultural heritage surveys emphasized that machine learning should be used responsibly in heritage contexts, because technical accuracy alone cannot explain cultural value, institutional bias, or archive formation processes [8]. This line of research supports viewing machine learning as a tool to assist cultural interpretation rather than replace expert judgment.

Research on style transfer, similarity discovery, and deep visual representation has also influenced computational art analysis. Neural style-transfer models showed that deep networks can separate and recombine content and style features, providing a key conceptual foundation for understanding how visual style may be computationally encoded [9]. Cross-cultural image-retrieval systems expanded this approach by showing how machine learning can identify artistic connections across different cultural collections, making retrieval useful for discovery-based archive exploration [10]. Residual learning architectures improved image-recognition performance and became important backbones for fine-art classification tasks [11]. Studies on illustration style recognition showed that stylistic learning is not limited to paintings but can also be applied to other visual-art domains such as children’s book illustrations [12]. Dense feature-learning models further improved deep networks' ability to extract detailed visual information from complex image datasets [13].

Another important group of studies developed methods for real-time style transfer, self-supervised learning, and visual-style recognition under limited or imbalanced labels. Perceptual-loss models showed that deep networks can preserve visual content while transforming style, which strengthened the technical basis for style-aware representation learning [14]. Self-trained ensemble and autoencoding approaches addressed the difficulty of learning style categories when annotated artwork data are incomplete or unevenly distributed [15]. Image-style recognition studies confirmed that style prediction can be treated as a measurable machine-learning task, while also showing that style categories often overlap visually and historically [16]. Institutional studies of cultural heritage image collections emphasized practical opportunities and challenges, including metadata inconsistency, collection bias, and the need for explainable computational workflows [17]. Large-scale painting categorization benchmarks provided early evidence that art-specific datasets are necessary because fine-art images differ substantially from natural-image classification tasks [18].

The general development of deep learning also shaped research on visual-arts classification. Large convolutional neural networks established the foundation for learned image features and demonstrated the importance of data scale in representation learning [19]. Automatic art-style recognition studies applied deep learning directly to painting datasets and showed that style labels can be predicted with meaningful accuracy when adequate training data are available [20]. Dimensionality-reduction techniques such as UMAP supported visual exploration of high-dimensional embeddings, making them useful for clustering artworks and interpreting relationships among styles, genres, and artists [21]. Face-recognition and clustering approaches introduced alternative strategies for classifying artists, styles, and years, showing that specific visual features can provide additional analytical pathways in art datasets [22]. Transferable CNN-feature studies demonstrated that pretrained visual descriptors can function as strong baselines for new recognition tasks, including fine-art analysis [23].

Fine-art classification research also developed more specialized approaches for metric learning, retrieval, interpretability, and multi-task modeling. Large-scale fine-art classification studies showed that the choice of feature representation and distance metric strongly affects style, artist, and genre prediction [24]. Two-stage deep-learning workflows improved fine-art classification by combining feature extraction and task-specific classification strategies [25]. Visual-link retrieval studies demonstrated that artwork databases can be explored through similarity relationships rather than only through fixed labels [26]. Interpretability methods such as gradient-based class activation maps helped explain which image regions influence model predictions, making them especially important for cultural heritage applications where black-box results are insufficient [27]. Deep convolutional architectures such as VGG provided strong transfer-learning backbones for artwork classification and visual-feature extraction [28]. Multi-task artistic analysis showed that artist, style, genre, and related metadata tasks can be modeled together, supporting richer computational use of art datasets [29].

Later studies strengthened the methodological foundation for efficient, scalable, and interpretable artwork analysis. GoogLeNet introduced efficient deep architectures that influenced later transfer-learning applications in visual classification [30]. EfficientNet improved model scaling and offered an alternative backbone for image-based artwork modeling [31]. ResNeXt developed aggregated residual transformations that improved deep feature extraction and supported more powerful visual-representation learning [32]. Big

transfer learning for fine-art classification demonstrated that modern transfer-learning models can improve classification and retrieval performance in art datasets [33]. Finally, research framing machine learning as a new form of art connoisseurship showed that computational models can assist art-historical reading when their outputs are combined with visual explanation, contextual analysis, and expert interpretation [34]. Together, these studies show that rigorous machine-learning research in visual arts should combine dataset profiling, label analysis, classification, similarity retrieval, interpretability, and cultural heritage interpretation.

Comparison table of studies related to machine learning, visual arts, and cultural heritage.

Ref.	Author and year	Main focus	Dataset or method	Relevance to this paper
[1]	Abgaz (2021)	Semantic enrichment	Cultural image metadata	Connects vision with metadata
[2]	Bar (2014)	Artistic style classification	Binarized deep features	Early deep style recognition
[3]	Castellano (2021)	Visual arts pattern review	Deep learning survey	Supports pattern-extraction framing
[4]	Cetinic (2018)	Fine-art CNN tuning	Multiple art tasks	Style and genre classification
[5]	Crowley (2014)	Object retrieval in paintings	Discriminative regions	Retrieval beyond labels
[6]	David (2017)	Painter classification	Deep autoencoders	Artist-level recognition context
[7]	Elgammal (2018)	Machine view of art history	Large-scale visual analysis	Connects modeling with interpretation
[8]	Fiorucci (2020)	Cultural heritage survey	ML in heritage	Frames responsible heritage modeling
[9]	Gatys (2016)	Neural style representation	CNN style/content separation	Explains visual style embeddings
[10]	Hamilton (2020)	Cross-cultural retrieval	Conditional image retrieval	Similarity discovery across collections
[11]	He (2016)	Deep residual learning	ResNet	Backbone for image modeling
[12]	Hicsonmez (2017)	Illustrator style recognition	Children's book illustrations	Style beyond paintings
[13]	Huang (2017)	Dense feature learning	DenseNet	Efficient deep classification
[14]	Johnson (2016)	Perceptual training	Perceptual loss	Supports visual representation learning
[15]	Joshi (2020)	Self-supervised art style learning	Autoencoding transformations	Addresses label imbalance
[16]	Karayev (2014)	Image-style recognition	Style descriptors	Basis for style classification
[17]	Karterouli (2021)	Institutional heritage images	Case study	Highlights opportunities and challenges
[18]	Khan (2014)	Painting categorization	Painting-91	Art-specific benchmark design
[19]	Krizhevsky (2012)	Deep image classification	Large CNN	Foundation for learned visual features
[20]	Lecoutre (2017)	Automatic art-style recognition	Wikipaintings	Directly related to style labels
[21]	McInnes (2018)	Embedding visualization	UMAP	Supports clustering and projection
[22]	Pancaroglu (2020)	Artist/style/year clustering	Face-based CNN clustering	Alternative feature strategy
[23]	Razavian (2014)	Transferable CNN features	Pretrained CNN descriptors	Supports baseline feature extraction
[24]	Saleh (2015)	Fine-art classification	Metric learning	Style, artist, and genre modeling
[25]	Sandoval (2019)	Two-stage fine-art classification	Deep learning pipeline	Improves classification workflow
[26]	Seguin (2016)	Painting retrieval	Visual links	Similarity-search framework

[27]	Selvaraju (2017)	Model explanation	Grad-CAM	Interpretability for visual models
[28]	Simonyan (2015)	Deep architecture	VGG networks	Transfer-learning backbone
[29]	Strezoski (2017)	Multi-task artistic analysis	OmniArt	Metadata-rich art modeling
[30]	Szegedy (2015)	Deep architecture	GoogLeNet	Efficient visual representation
[31]	Tan (2019)	Efficient image modeling	EfficientNet	Alternative deep backbone
[32]	Xie (2017)	Residual transformations	ResNeXt	Improved feature extraction
[33]	Zhao (2022)	Fine-art transfer learning	BiT and CNN models	Modern benchmark for classification and retrieval
[34]	Zhu (2019)	Art connoisseurship modeling	CNN + visualization	Links accuracy and art-historical reading

The comparison shows that the present study is positioned at the intersection of three research lines: deep feature learning, artwork classification, and cultural heritage interpretation. Its distinctive contribution is not proposing another isolated classifier; instead, it provides a dataset-aligned empirical framework that begins with metadata profiling, reports missingness and class imbalance, evaluates baseline classification, and prepares the corpus for image-level modeling and similarity retrieval. This structure directly addresses methodological weaknesses identified in earlier research, particularly the tendency to report accuracy without fully describing dataset structure, label imbalance, or cultural interpretation.

III. MATERIALS AND METHODS

1. RESEARCH DESIGN

The study uses a quantitative computational design supported by interpretive cultural analysis. The quantitative component profiles the WikiArt metadata corpus, evaluates class structure, measures missingness and imbalance, examines label associations, and defines a model-ready classification and retrieval workflow. The interpretive component explains how computational outputs can support cultural-heritage organization without reducing cultural meaning to numerical prediction.

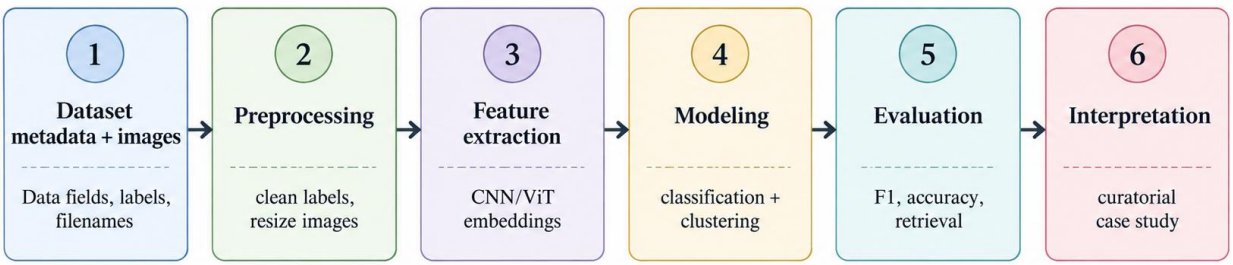


FIGURE 1. Machine-learning workflow for cultural heritage image analysis.

The workflow connects dataset preparation, preprocessing, visual feature extraction, classification, similarity retrieval, evaluation, and cultural interpretation.

2. DATASET AND EMPIRICAL SOURCE

The empirical source is a WikiArt metadata corpus selected for this study. It contains 103,250 artwork records and uses an image identifier field to connect each record to the corresponding image archive. Because this corpus contains metadata rather than pixel-level image content, the study uses it for statistical profiling

and baseline modeling, while the full image-based experiment requires the matching image archive linked through the identifier field.

Table 1. Dataset fields and their role in the machine-learning study.

Field	Type	Analytical role
artist	Text label	Identifies the artist and supports stratified sampling, author-level analysis, and leakage checks.
date	Temporal field	Provides approximate production date when available; missingness must be documented.
genre	Categorical label	Supports genre classification and cultural grouping.
style	Categorical label	Supports style classification and visual-pattern analysis.
pixelsx, pixelsy	Numerical image metadata	Describe image dimensions and support technical filtering.
digital size	Numerical technical metadata	Provides a proxy for digital density and image-resolution variation.
source	Categorical field	Distinguishes WikiArt and Wikipedia records.
artist_group	Split-control field	Indicates whether an artist appears in training, test, or both subsets.
in_train	Boolean split field	Defines the existing training and test partition.
image identifier	Image identifier	Links each metadata record to the corresponding image archive.

Table 1 shows that the corpus contains labels, technical image information, source fields, and split indicators that can be used to prepare supervised and retrieval-based machine-learning experiments.

3. *DATA PREPROCESSING*

The preprocessing stage included loading the structured metadata table, standardizing field types, computing missing-value rates, checking distinct label counts, and deriving additional technical variables. We derived the following variables from the original fields: aspect ratio, image area, logarithmic digital size, and numeric date where parsing was possible. Missing values were not imputed for descriptive reporting because missingness itself is an important property of cultural-heritage metadata. For baseline classification, we imputed numerical features with median values and categorical features with the most frequent category.

The following preprocessing equation defines the aspect ratio used in the technical image-profile analysis:

$$AR_i = \frac{W_i}{H_i}$$

where AR_i is the aspect ratio of artwork image i, W_i is image width, and H_i is image height.

4. *STATISTICAL ANALYSIS*

The statistical analysis included frequency distributions, missingness analysis, descriptive statistics, class-imbalance measurement, and association testing. Normalized entropy was used to summarize how evenly labels are distributed across a categorical field:

https://doi.org/10.48161/ajst.v1n2a85

$$H_{norm} = \frac{-\sum_{k=1}^{K} p_k \log(p_k)}{\log(K)}$$

where p_j is the share of class j, and K is the number of classes. Values closer to 1 indicate a more even distribution, while lower values indicate stronger concentration.

The association between style and genre was measured using a chi-square test and Cramer's V:

$$V = \sqrt{\frac{\chi^2}{n \times \min(r-1, c-1)}}$$

where χ^2 is the chi-square statistic, n is the number of valid records, r is the number of rows, and c is the number of columns in the contingency table.

5. PROPOSED MACHINE-LEARNING FRAMEWORK

The image-based model treats each artwork image as an input x_i and transforms it into a learned embedding vector z_i:

$$z_i = f_\theta(x_i), \qquad z_i \in \mathbb{R}^d$$

where fθ is a pretrained or fine-tuned computer-vision model, and d is the embedding dimension.

For supervised classification, the probability that artwork i belongs to class k is defined through a softmax layer:

$$P(y_i = k \mid x_i) = \frac{\exp(w_k^T z_i + b_k)}{\sum_{j=1}^{K} \exp\left(w_j^T z_i + b_j\right)}$$

The classification model is optimized using cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik} \log(p_{ik})$$

For similarity retrieval, two artworks are compared through cosine similarity:

$$s_{ij} = \frac{z_i \cdot z_j}{\| z_i \| \| z_j \|}$$

Macro-F1 is used as the main classification metric because the dataset is highly imbalanced:

$$Macro\text{-}F1 = \frac{1}{K}\sum_{k=1}^{K} \frac{2 \times Precision_k \times Recall_k}{Precision_k + Recall_k}$$

Table 2. Recommended modeling protocol for the final image-based study.

Component	Recommended implementation	Reason
Main prediction task	Top-k style classification	Style is the most appropriate primary label for visual pattern analysis.
Secondary prediction task	Top-k genre classification	Genre gives a complementary cultural and iconographic grouping.
Baseline model	Metadata-only decision tree or linear classifier	Shows how much organization is possible without pixel features.
Image model	CNN or Vision Transformer embedding extractor	Requires the image archive linked through the image identifier.
Similarity retrieval	Cosine similarity over embeddings	Supports discovery of visually related artworks.

Component	Recommended implementation	Reason
Clustering	k-means, UMAP or t-SNE projection	Supports exploratory visualization of visual groupings.
Case interpretation	Curatorial reading of retrieved groups	Connects computational output with cultural meaning.

Table 2 defines how metadata should be connected to image-level modeling, similarity retrieval, and cultural interpretation.

6. RESEARCH CONTRIBUTION

This study makes five contributions. First, it translates a broad topic into a dataset-aligned empirical research design using a large art metadata corpus. Second, it provides a statistical profile of 103,250 artwork records, including label coverage, class distribution, missingness, technical image fields, and split structure. Third, it introduces a reproducible machine-learning framework that links classification, similarity retrieval, clustering, and cultural interpretation. Fourth, it reports metadata-only baseline results to show what can and cannot be achieved without image features. Fifth, it clarifies the methodological boundary between computational prediction and cultural-heritage interpretation.

7. EVALUATION METRICS

Table 3. Evaluation metrics for classification and retrieval.

Metric	Definition	Relevance to this study
Accuracy	Overall proportion of correct predictions.	Useful but insufficient under class imbalance.
Macro-precision	Average precision across classes.	Shows whether predictions are reliable across small and large classes.
Macro-recall	Average recall across classes.	Shows whether minority classes are recovered.
Macro-F1	Harmonic mean of class-level precision and recall.	Recommended primary metric for imbalanced style and genre classification.
Confusion matrix	Class-by-class error matrix.	Identifies which styles or genres are confused.
Top-k retrieval accuracy	Share of correct or curatorially relevant results among top retrieved images.	Useful for similarity search.
Mean average precision	Ranking quality across similarity queries.	Useful for archive retrieval and visual matching.

We selected these metrics because visual arts datasets are imbalanced and culturally heterogeneous; therefore, accuracy alone is not enough.

IV. RESULTS

1. DATASET COMPOSITION

The WikiArt metadata corpus contains 103,250 artwork records. The predefined split contains 79,433 training records (76.93%) and 23,817 test records (23.07%). The metadata includes 2,319 distinct artists, 136

style labels, and 42 genre labels. Most records are associated with WikiArt, while fewer are associated with Wikipedia.

Table 4. Empirical summary of the WikiArt metadata corpus.

Dataset property	Value	Interpretation
Total metadata records	103,250	Complete collection-level metadata entries in the WikiArt metadata corpus.
Training subset	79,433 (76.93%)	Records marked True in the in_train field.
Test subset	23,817 (23.07%)	Records marked False in the in_train field.
Artists	2,319	Distinct artist labels available for modeling or stratified analysis.
Style labels	136	986 missing style records.
Genre labels	42	1,254 missing genre records.
WikiArt source records	98,702	Primary source category in the metadata.
Wikipedia source records	4,381	Secondary source category in the metadata.
Missing source records	167	Records without an explicit source value.

Table 4 confirms that the dataset is large enough for machine-learning research, but its scale requires systematic filtering, class balancing, and careful evaluation.

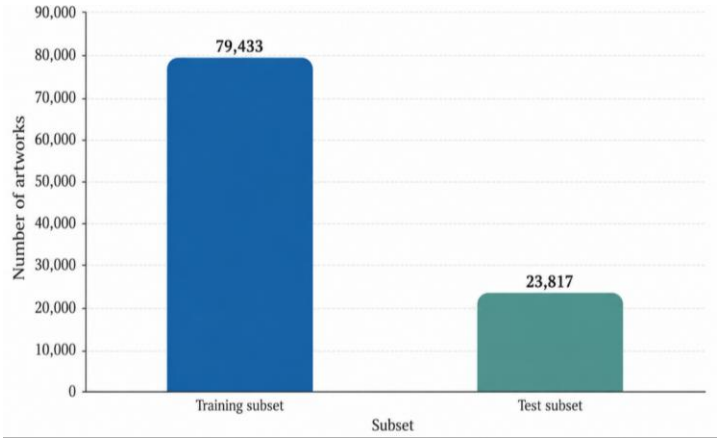


FIGURE 2. Dataset split based on the in_train field. The dataset already contains a clear training-test partition, which supports reproducible model evaluation.

https://doi.org/10.48161/ajst.v1n2a85

Dataset split based on the in_train field.



FIGURE 3. Source distribution in the WikiArt metadata corpus. Report the dominance of WikiArt records because source composition can influence label conventions, image quality, and archive coverage.

2. *MISSING METADATA*

The missing-value analysis shows that the date field has the highest missingness at 26,363 records (25.53%). Genre is missing for 1,254 records (1.21%), style is missing for 986 records (0.95%), and source is missing for 167 records (0.16%). These rates are manageable for style and genre classification but should still be reported because missingness can affect sampling and interpretation.

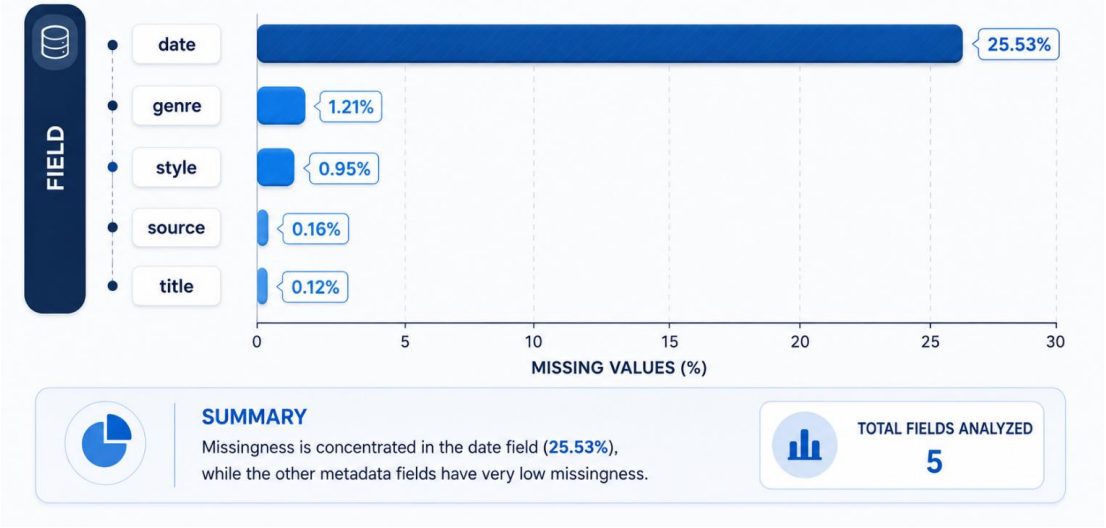


FIGURE 4. Missing metadata by field. The figure shows that most modeling labels are complete, while production date is substantially less complete.

3. *TECHNICAL IMAGE-PROFILE STATISTICS*

Image dimensions and digital sizes vary widely, which is expected in a large artwork collection assembled from different digital sources. The median width and height are both 800 pixels, but the maximum values show that some images are much larger. Aspect-ratio analysis shows that 51,397 records are portrait-oriented, 44,362 are landscape-oriented, and 7,491 are near-square.

Table 5. Descriptive statistics for technical image fields.

Variable	Mean	SD	Min	Median	95th percentile	Max
Width in pixels	960.81	596.27	200.00	800.00	2,052.00	25,528.00
Height in pixels	984.09	625.14	162.00	800.00	2,339.00	9,541.00
Digital size in bytes	388,323.52	878,477.57	3,516.00	167,670.00	1,390,985.75	32,443,139.00
Aspect ratio	1.05	0.39	0.10	0.91	1.64	17.87

Table 5 supports the need to resize and normalize images before training a computer-vision model.

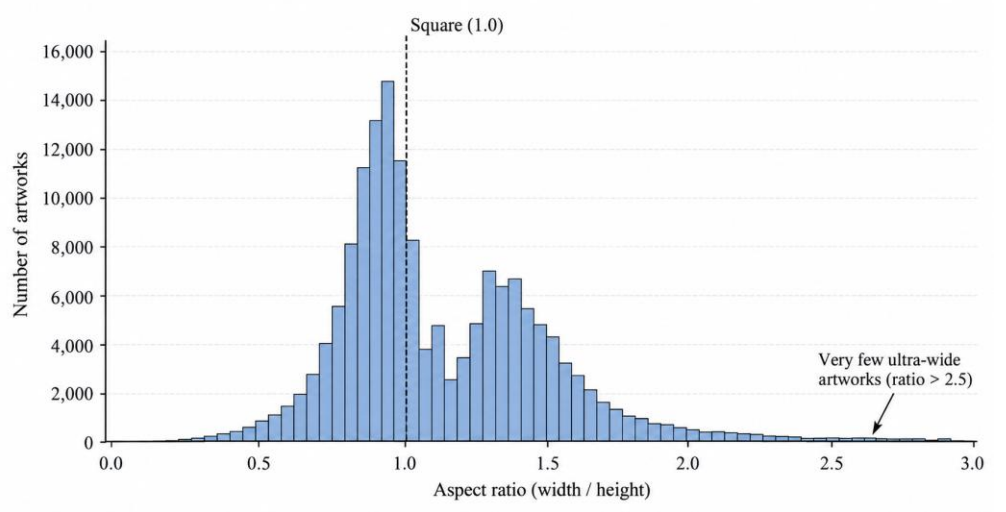


FIGURE 5. Distribution of artwork aspect ratio. Aspect-ratio variation indicates that preprocessing should preserve visual composition while standardizing input dimensions.

4. STYLE AND GENRE DISTRIBUTIONS

The most frequent style labels are Impressionism, Realism, Romanticism, Expressionism, Post-Impressionism, Art Nouveau, Baroque, Surrealism, Symbolism, Rococo, Northern Renaissance, and Naive Art. The most frequent genre labels are portrait, landscape, genre painting, abstract, religious painting, cityscape, sketch and study, illustration, still life, symbolic painting, nude painting, and figurative works.

Table 6. Most frequent style labels.

Style	Artworks	Share of labeled styles (%)
Impressionism	10643	10.41
Realism	10523	10.29
Romanticism	9285	9.08
Expressionism	7013	6.86
Post-Impressionism	5778	5.65
Art Nouveau (Modern)	4899	4.79
Baroque	4400	4.3

Style	Artworks	Share of labeled styles (%)
Surrealism	4167	4.07
Symbolism	3476	3.4
Rococo	2733	2.67
Northern Renaissance	2379	2.33
Naïve Art (Primitivism)	2295	2.24

The largest style classes can be used for a first supervised style-classification experiment, but minority styles should not be ignored in later extensions.

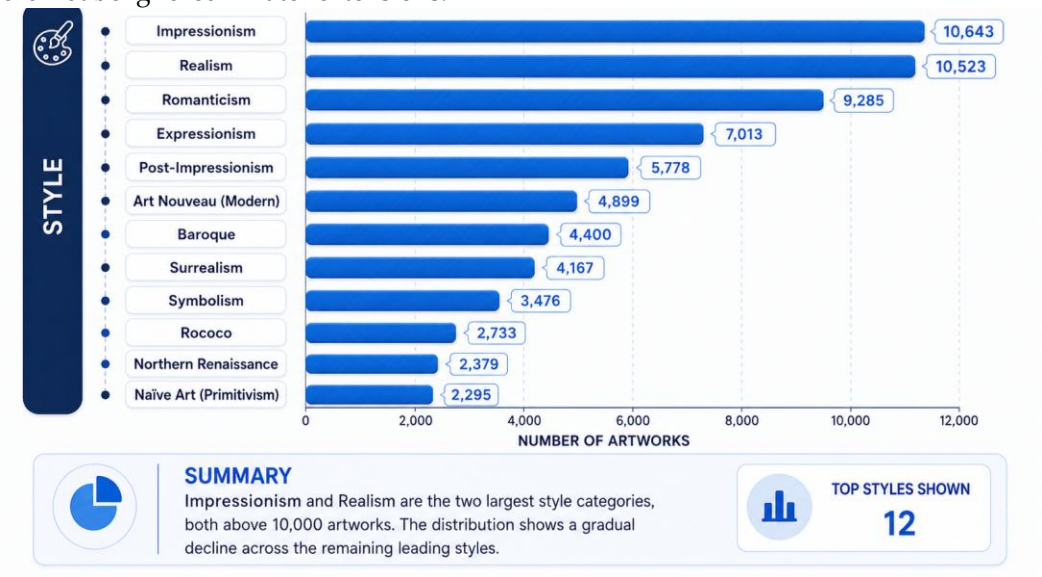


FIGURE 6. Most frequent style labels in the dataset. The distribution shows strong concentration among a limited set of historically broad style categories.

5. Most frequent style labels in the dataset.

Table 7. Most frequent genre labels.

Genre	Artworks	Share of labeled genres (%)
portrait	16847	16.52
landscape	15006	14.71
genre painting	14260	13.98
abstract	9498	9.31
religious painting	7429	7.28
cityscape	5348	5.24
sketch and study	3644	3.57
illustration	3202	3.14

Genre	Artworks	Share of labeled genres (%)
still life	3132	3.07
symbolic painting	2545	2.5
nude painting (nu)	2290	2.25
figurative	2244	2.2

Genre labels help classify works because they describe subject matter and compositional organization, but they don't always align with style.

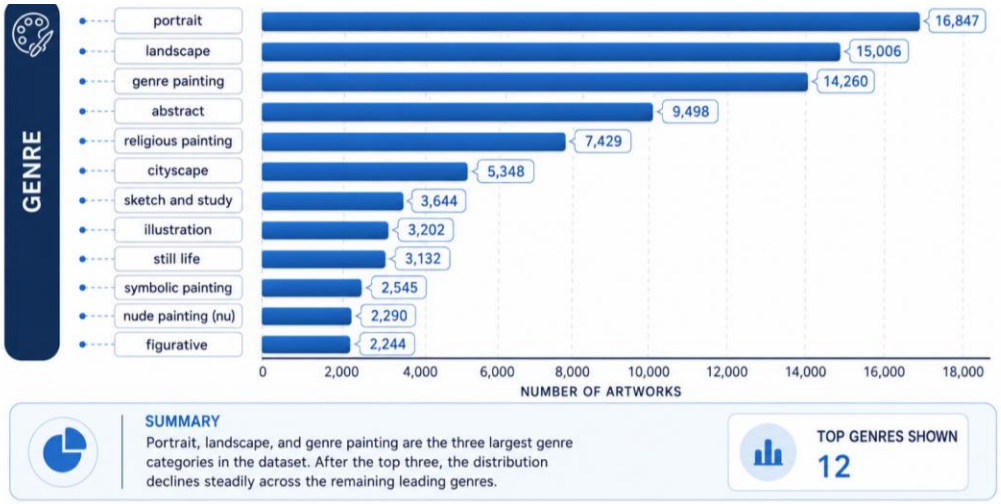


FIGURE 7. Most frequent genre labels in the dataset. Portrait, landscape, and genre painting are the largest genre groups and should be represented in both training and test partitions.

6. CLASS IMBALANCE AND LABEL ASSOCIATION

Class-balance statistics indicate that style and genre labels are imbalanced but still useful for modeling. Normalized entropy equals 0.726 for style, 0.740 for genre, and 0.882 for artist labels. This means the artist field is more evenly distributed than style and genre, although artist classification remains difficult because it involves many classes.

Table 8. Class-balance indicators.

Label	Classes	Valid N	Entropy	Largest N	Implication
Style	136	102,264	0.726	10,643	Use class weights
Genre	42	101,996	0.740	16,847	Secondary target
Artist	2,319	103,250	0.882	500	High-class task

Table 8 shows why macro-F1 and balanced sampling are necessary. A model can appear accurate if it performs well only on frequent categories.

The chi-square test between style and genre showed a statistically significant association (χ^2 = 5,494, df = 101,514, valid records = 101,514). Cramer's V is 0.345, indicating a moderate association between style and genre. This supports using both labels: they are related, but not identical classification tasks.$\chi^2 = 494{,}553.56 p < .001$

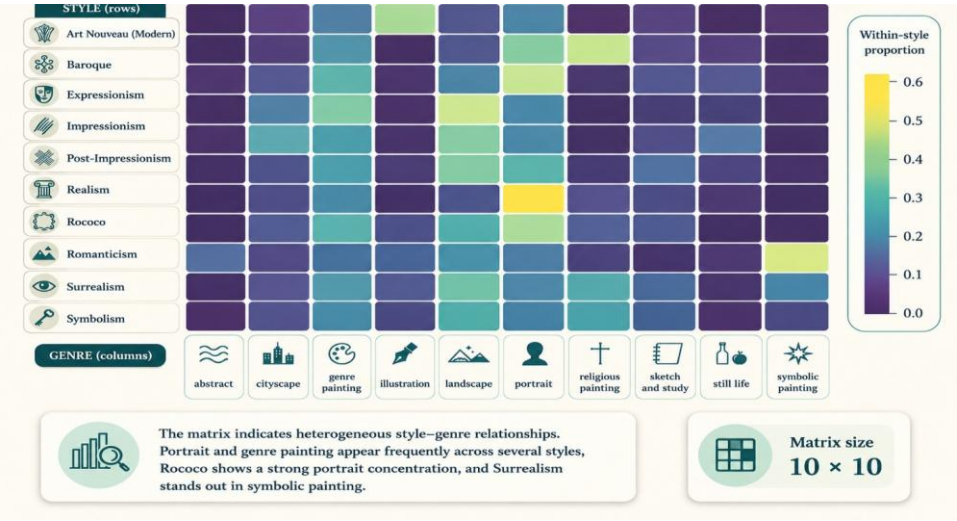


FIGURE 8. Style–genre association among frequent labels. The heatmap shows that some styles strongly associate with particular genres, while others span several subject categories.

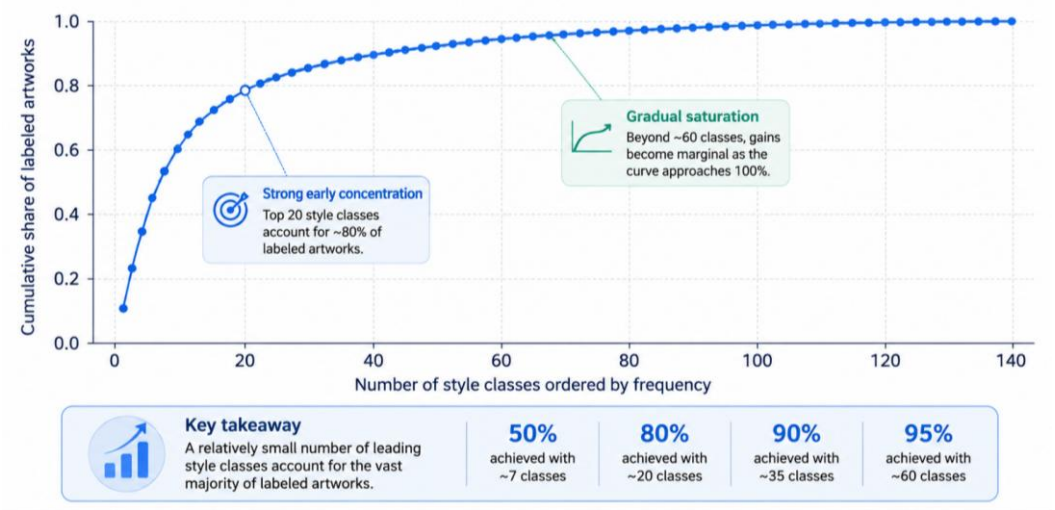


FIGURE 9. Cumulative concentration of style classes. The curve shows that a small number of style classes account for a large share of labeled records, creating a class-imbalance problem.

7. METADATA-ONLY MACHINE-LEARNING BASELINE

We calculated a metadata-only baseline to test whether non-image fields can support preliminary archive organization. The features included image dimensions, aspect ratio, image area, digital size, title length, numeric date, genre or style, source, and artist-group information. We evaluated two top-10 classification tasks: style prediction and genre prediction. These results should not be interpreted as final visual-recognition performance because visual pixels were not used. Instead, the results provide a baseline for archive-level metadata organization.

https://doi.org/10.48161/ajst.v1n2a85

Table 9. Metadata-only baseline classification results.

Task	Model	Train/Test N	Accuracy	Macro-F1
Top-10 style	Frequency baseline	48,118/14,799	0.164	0.028
Top-10 style	Linear SGD	48,118/14,799	0.371	0.370
Top-10 style	Decision tree	48,118/14,799	0.457	0.466
Top-10 genre	Frequency baseline	62,145/18,766	0.209	0.035
Top-10 genre	Linear SGD	62,145/18,766	0.481	0.415
Top-10 genre	Decision tree	62,145/18,766	0.476	0.426

The decision-tree baseline achieved a macro-F1 of 0.466 for top-10 style classification and 0.426 for top-10 genre classification. These values show that metadata carries useful organizational information, but true visual pattern analysis requires image-based embeddings.

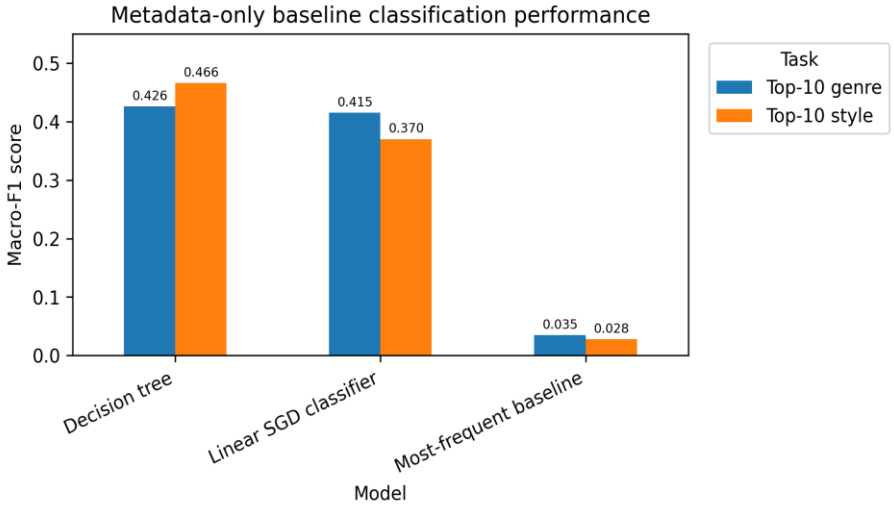


FIGURE 10. Metadata-only baseline classification performance. The metadata-only models outperform the most-frequent baseline, but their performance remains limited because they lack visual features.

Metadata-only baseline classification performance.

V. CULTURAL INTERPRETATION CASE STUDY

A focused cultural interpretation component is required after model training to connect computational outputs with art-historical meaning. The case study can select one query artwork from a common style, such as Impressionism or Realism, and retrieve the most visually similar artworks using cosine similarity on image embeddings. The retrieved objects can then be compared by composition, color organization, subject matter, style label, genre label, date, and artist.

The case-study interpretation addresses three issues. First, it examines whether the retrieved objects share visible characteristics such as brushwork, palette, perspective, figure arrangement, or landscape structure. Second, it evaluates whether retrieved objects also share metadata categories such as genre, style, artist, or period. Third, it identifies where computational similarity and cultural interpretation disagree. This structure

makes the case study analytically meaningful because it transforms machine-learning output into a cultural-heritage interpretation.

VI. DISCUSSION

The empirical profiling confirms that the WikiArt metadata corpus is highly useful for a machine-learning paper on visual arts and cultural heritage collections. Its size, label richness, image identifiers, and train-test split make it suitable for supervised classification and archive organization. However, the results also show why careful methodology is necessary. The dataset is not evenly distributed across style and genre labels. Frequent classes dominate the collection, and a model trained without balancing could perform well on common categories while performing poorly on smaller classes.

The association between style and genre is important for interpretation. A moderate Cramer's V indicates that style and genre are connected but not interchangeable. This is expected in art-historical terms because style refers to formal and historical characteristics, while genre refers more directly to subject matter or pictorial category. For example, a landscape may appear in different styles, and a single style may include portraits, religious paintings, landscapes, and still lifes. Therefore, a strong model should not treat style and genre as identical labels. Instead, it should evaluate them as related but distinct dimensions of cultural classification.

The metadata-only baseline results are useful but limited. They show that archive fields can support preliminary organization and that metadata can carry predictive information. However, metadata-only models cannot detect visual similarity in the full sense. Without image features, they cannot compare brushwork, composition, texture, iconographic arrangement, or visual structure. For this reason, the final full study should combine metadata profiling with image embeddings extracted from a convolutional neural network or Vision Transformer.

The most important methodological implication is that machine learning should be treated as a structured aid for cultural-heritage research, not as an autonomous interpreter. Classification results can reveal patterns, but they do not explain cultural meaning by themselves. Similarity retrieval can identify visually related objects, but curatorial interpretation is needed to determine whether the similarity is historically meaningful, materially relevant, or merely formal. This combined approach matters especially for heritage collections, where objects may carry cultural, religious, political, and community meanings not visible in image data alone.

VII. CONCLUSION

This study establishes a dataset-aligned machine-learning framework for analyzing patterns in WikiArt visual arts metadata and preparing the corpus for deeper image-based cultural heritage research. The empirical results demonstrate that the corpus is large, structured, and suitable for supervised classification and archive organization, with 103,250 artwork records, rich style and genre labels, technical image descriptors, source indicators, and a predefined training-test partition. At the same time, the findings show that the dataset requires careful methodological control. Before interpreting model performance responsibly, we must address production-date missingness, strong source concentration, and unequal class distributions.

The analysis confirms that style and genre are useful but distinct dimensions of cultural organization. Major classes such as Impressionism, Realism, Romanticism, portrait, landscape, and genre painting dominate the corpus, creating a risk that unbalanced models may overrepresent frequent categories. The moderate style–genre association further shows that these labels overlap without becoming identical. This result is important for both machine-learning design and art-historical interpretation because it supports separate evaluation of style classification, genre classification, and similarity retrieval.

The metadata-only baseline provides a practical contribution by showing that archive descriptors contain measurable predictive structure. However, it also confirms that metadata cannot substitute for image-level representation. True visual pattern analysis must incorporate embeddings that capture composition, palette, texture, brushwork, and spatial arrangement. The study therefore positions metadata profiling as a necessary

first layer and image-based modeling as the next stage of the research. Overall, the paper offers a transparent, reproducible pathway for connecting statistical corpus analysis, machine-learning classification, similarity retrieval, and curatorial interpretation. Its main value lies in demonstrating how computational methods can strengthen digital heritage organization while preserving the need for expert cultural judgment. Future research should extend the framework with the corresponding image archive, compare deep visual architectures, and evaluate retrieval results through both quantitative metrics and art-historical interpretation.

VIII. DATA AVAILABILITY STATEMENT

The empirical analysis uses the WikiArt metadata corpus supplied for this study. The corpus includes artwork identifiers and structured metadata fields; full visual-model training requires the corresponding image archive from the same dataset source.

Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

[1] Abgaz, Y. M., Souza, R. R., Methuku, J., Koch, G., & Dorn, A. (2021). A methodology for semantic enrichment of cultural heritage images using artificial intelligence technologies. Journal of Imaging, 7(8), Article 121.

[2] Bar, Y., Levy, N., & Wolf, L. (2014). Classification of artistic styles using binarized features derived from a deep neural network. In European Conference on Computer Vision Workshops.

[3] Castellano, G., & Vessio, G. (2021). Deep learning approaches to pattern extraction and recognition in paintings and drawings: An overview. Neural Computing and Applications, 33, 12263-12282.

[4] Cetinic, E., Lipic, T., & Grgic, S. (2018). Fine-tuning convolutional neural networks for fine art classification. Expert Systems with Applications, 114, 107-118.

[5] Crowley, E. J., & Zisserman, A. (2014). The state of the art: Object retrieval in paintings using discriminative regions. In British Machine Vision Conference.

[6] David, E., & Netanyahu, N. S. (2017). DeepPainter: Painter classification using deep convolutional autoencoders. arXiv preprint arXiv:1711.08763.

[7] Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2018). The shape of art history in the eyes of the machine. In AAAI Conference on Artificial Intelligence.

[8] Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., & James, S. (2020). Machine learning for cultural heritage: A survey. Pattern Recognition Letters, 133, 102-108.

[9] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 2414-2423).

[10] Hamilton, M., Fu, S., Lu, M., Bui, J., Bopp, D., Chen, Z., Tran, F., Wang, M., Rogers, M., Zhang, L., Hoder, C., & Freeman, W. T. (2020). MosAIc: Finding artistic connections across culture with conditional image retrieval. arXiv preprint arXiv:2007.07177.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778).

[12] Hicsonmez, S., Samet, N., Sener, F., & Duygulu, P. (2017). DRAW: Deep networks for recognizing styles of artists who illustrate children's books. arXiv preprint arXiv:1704.03057.

[13] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 4700-4708).

[14] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision (pp. 694-711).

[15] Joshi, A., Agrawal, A., & Nair, S. (2020). Art style classification with self-trained ensemble of autoencoding transformations. arXiv preprint arXiv:2012.03377.

[16] Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., & Winnemoller, H. (2014). Recognizing image style. In British Machine Vision Conference.

[17] Karterouli, K., & Batsaki, Y. (2021). AI and cultural heritage image collections: Opportunities and challenges. In Proceedings of EVA London 2021 (pp. 193-198).

[18] Khan, F. S., Beigpour, S., van de Weijer, J., & Felsberg, M. (2014). Painting-91: A large-scale database for computational painting categorization. Machine Vision and Applications, 25(6), 1385-1397.

[19] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (Vol. 25).

[20] Lecoutre, A., Negrevergne, B., & Yger, F. (2017). Recognizing art style automatically in painting with deep learning. In Proceedings of the Ninth Asian Conference on Machine Learning (PMLR, Vol. 77, pp. 327-342).

[21] McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

[22] Pancaroglu, D. (2020). Artist, style and year classification using face recognition and clustering with convolutional neural networks. arXiv preprint arXiv:2012.01009.

[23] Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. In IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 806-813).

[24] Saleh, B., & Elgammal, A. (2015). Large-scale classification of fine-art paintings: Learning the right metric on the right feature. International Journal for Digital Art History, 2, 71-93.

[25] Sandoval, C., Pirogova, E., & Lech, M. (2019). Two-stage deep learning approach to the classification of fine-art paintings. IEEE Access, 7, 41770-41781.

[26] Seguin, B., Striolo, C., Kaplan, F., & di Lenardo, A. (2016). Visual link retrieval in a database of paintings. In European Conference on Computer Vision Workshops (pp. 753-767).

[27] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In IEEE International Conference on Computer Vision (pp. 618-626).

[28] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations.

[29] Strezoski, G., & Worring, M. (2017). OmniArt: Multi-task deep learning for artistic data analysis. arXiv preprint arXiv:1708.00684.

[30] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-9).

[31] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning (pp. 6105-6114).

[32] Xie, S., Girshick, R., Dollar, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 1492-1500).

[33] Zhao, W., Jiang, W., & Qiu, X. (2022). Big transfer learning for fine art classification. Computational Intelligence and Neuroscience, 2022, Article 1764606.

[34] Zhu, Y., Ji, Y., Zhang, Y., Xu, L., Zhou, A. L., & Chan, E. (2019). Machine: The new art connoisseur. arXiv preprint arXiv:1911.10091.