NEURAL NETWORKS FOR CLASSIFYING INDIAN FOLK MOTIFS

Paramjit Baxi [1] ✉ iD, Saritha SR [2] ✉ iD, Praveen Kumar Tomar [3] ✉, Harshita Sharma [4] ✉ iD, Smitha K [5] ✉, Wamika Goyal [6] ✉ iD, Ashutosh Kulkarni [7] ✉

[1] Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh, Solan, 174103, India
[2] Assistant Professor, Department of Management Studies, JAIN (Deemed-to-be University), Bengaluru, Karnataka, India
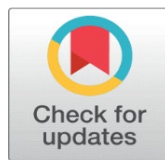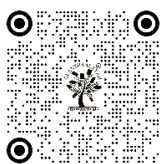[3] Professor, School of Business Management, Noida International University 203201, Greater Noida, Uttar Pradesh, India
[4] Assistant Professor, Department of Development Studies, Vivekananda Global University, Jaipur, India
[5] Lloyd Law College, Greater Noida, Uttar Pradesh 201306, India
[6] Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India
[7] Department of DESH, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037 India

ABSTRACT

In this paper, I have outlined a hybrid neural network model of automated classification of Indian folk motifs within the different regional traditions, such as Madhubani, Warli, Kalamkari, and Pattachitra. Introduced was a culturally authentic curated collection of 5000 high-resolution images created by ethical digitization and by human experts. The suggested model uses a Convolutional Block Attention Module (CBAM) in addition to a ResNet-50 backbone to promote the discrimination of features in space and channels. The experimental findings show better performance in comparison with the baseline CNN and VGG architectures, with the overall accuracy of 94.6, macro F1-score of 94.0 and Cohen 6. Grad-CAM visualizations show that the activations of the model are consistent with motif-specific areas of art, which verify the cultural interpretability. The framework helps to explain explainable cultural AI because it associates computational properties with heritage aesthetics, allowing them to be used in digital museums, education platforms and art restoration systems. The research provides a methodological basis of combining deep learning with the preservation of cultural heritage with a focus on transparency, reproducibility, and cross-disciplinary applicability.

Keywords: Indian Folk Art, Neural Networks, Cultural Heritage, Motif Classification, Grad-CAM, Transfer Learning, Digital Preservation

1. INTRODUCTION

Indian folk art is one of the most varied visual cultures of the world, comprising centuries-old traditions of the region, which are spiritual, social, and ecological in nature. Madhubani (Bihar), Warli (Maharashtra), Kalamkari (Andhra

Pradesh), Pattachitra (Odisha) are styles that are closely tied to the local cultures and make use of very particular geometric designs, forms and color combinations. As the digital revolution has ensued and cultural informatics has been given more and more attention there is a huge requirement to conserve, analyze and systematically categorize these motifs to document, educate and to provide the creative industries. But even with the problem of stylistic confusion, incompatible artistic traditions, and small annotated datasets, the task of motif classification is still hard. Conventional approaches to image classification, which are based on manually-created features (SIFT, HOG, or color histograms), do not generalize to heterogeneous motifs. These methods do not have the representational strength to portray these complex linework, symbolism abstraction and aesthetic nuances that characterize Indian folk art.
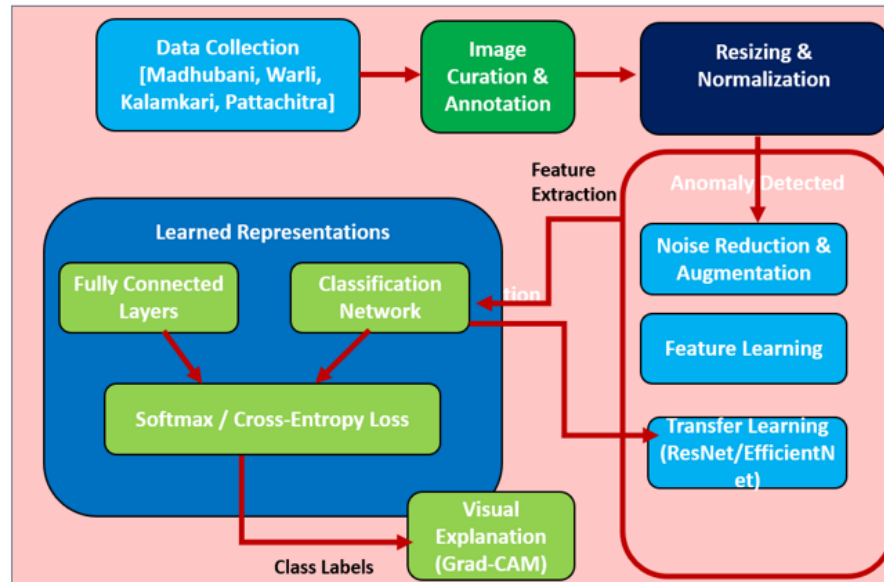
Figure 1

Figure 1 Conceptual Framework of Neural Network-Based Indian Folk Motif Classification

They are used to model complicated artistic styles due to their capacity to reproduce spatial, textural and chromatic features. New advances in the transfer learning and attention mechanisms have also improved the model interpretability and generalization, and that is why it can be applied to small or specialized datasets like heritage art collections. The paper is meant to create and test a neural network system to identify the Indian folk motifs and group them according to their regional or stylistic contexts. The study examines how the convolutional neural networks (CNNs) and multi-layered models of attention and feature fusion can be used to detect unique features used to distinguish folk traditions. The algorithm is based on feature learning as opposed to manual feature engineering and uses data augmentation and fine-tuning techniques to reduce overfitting and enhance generalization. The trained model should be able to identify visual features such as stroke density, geometry of motifs, and color arrangement that are typical of every style. The input of this paper is tripled. To begin with, it provides an annotated and edited collection of Indian folk motifs across several art traditions. Second, it suggests a strong neural network structure that is optimized on multi-class motifs recognition in the focus of feature interpretability. Third, it shows the cultural and practical consequences of the model in terms of digital archiving, museum cataloguing, and art education with AI assistance. The general structure highlights the integration of computational intelligence with the cultural heritage preservation to support the current development of artificial intelligence as a driver to the continuity of the traditional knowledge systems in the digital era.

2. RELATED WORK

The use of automatic classification of visual art and folk motifs has developed by various technological paradigms. Traditional techniques were largely dominated by manual methods of extracting features, which were based on color, shape, and texture features. Afterwards, with the development of deep learning, machines were able to learn high-level representation directly based on data which made great improvements in performance in terms of accuracy and interpretability. The review below follows this change of direction and offers the ways in which these changes are merged towards the classification of Indian folk motifs. The initial analysis of computational art was based on low-level

image features. SIFT, HOG and Gabor filters were used to derive geometric or textural data of paintings. These features were then inputted into the classifier as either Support Vector Machines (SVM) or k-Nearest Neighbors (kNN). Even though effective in recognizing the simple color or edge patterns, these pipelines were unable to reflect the semantic richness and stylistic abstraction of the Indian folk art. As an illustration, the patterns of Warli (built with white geometrical shapes on a background of earthen colors) puzzled the models that were trained on color-dependent patterns, and the dense patterns of Madhubani resulted in over segmentation errors. The Bag-of-Visual-Words (BoVW) system was more scalable but had to hand engineer features and was unable to capture hierarchical abstractions.

Table 1

Table 1 Representative Studies on Image and Art Classification Techniques			
Technique / Model	**Dataset Type**	**Key Contribution**	**Limitations**
Texture and Edge Descriptors + SVM	European paintings	Painter identification using handcrafted features	Poor style generalization
BoVW + Color Attributes	WikiArt	Automated style classification	Manual feature design required
CNN for Style Transfer	Artistic images	Demonstrated deep feature representation of style	Computational intensive
Transfer Learning (Inception-v3)	Cultural heritage artifacts	Recognition with limited datasets	Limited dataset variety
CNN for Warli/Madhubani Art	Indian folk art images	Cultural style classification using deep features	Small dataset size

All of these studies demonstrate the progressive change in engineered to learned features. Nevertheless, they all dealt with the multidimensional difference, which is regional, stylistic, and thematic, existing in Indian folk motives. Image recognition was transformed to use deep learning and specifically convolutional neural networks (CNNs) which learn feature hierarchies in the form of layers. Models like AlexNet, VGG-16, ResNet-50, and EfficientNet have been used to reach the state-of-the-art results in different areas. Gatys et al. (2015) also showed that CNNs have the ability to decouple content and style and understand artistic expression computationally. CNNs and models based on transformers have been effectively employed in the cultural heritage field to recognise artifacts, to segment sculptures and to restore mural paintings. Inception-v3 and EfficientNet are transfer-learning models that have also been particularly useful in cases where the size of the dataset is small - a frequent limitation in legacy archives. Attention mechanisms, Grad-CAM visualization, and feature fusion layers have also been proposed by researchers to make neural activities more interpretable, and experts can relate them to aesthetic or symbolic objects in artworks.

Table 2

Table 2 Comparison of Classical and Deep Learning Approaches in Art Classification		
Criterion	**Classical Methods (SIFT/HOG + SVM)**	**Deep Learning (CNN/ResNet/EfficientNet)**
Feature Extraction	Hand-crafted; requires domain expertise	Learned automatically through training
Representation Level	Low-level (color, edges, texture)	Multi-level spatial and semantic features
Dataset Requirement	Moderate (100–500 images)	High (> 1000 images, augmentable)
Generalization Capability	Limited to specific styles	Robust across heterogeneous art forms
Interpretability	Manual visual analysis	Grad-CAM / Attention heat maps
Computational Efficiency	Low	High (compute-intensive but accurate)

This comparison highlights the superiority of deep networks to motif classification because they have the ability to learn minor aesthetic features such as stroke density or its compositional rhythm without human correction. Though, in India, there has been a growing scholarly attempt in recent decades to digitize and categorize folk art. Ravikiran et al. (2020) have constructed a CNN model that differentiates Madhubani and Warli paintings and Singh et al. (2022) have used fine-tuned ResNet models with over 90 per cent accuracy in curated datasets. Other researchers suggested that hybrid models based on Gray-Level Co-occurrence Matrix (GLCM) features with deep embeddings can be used to improve texture discrimination. Although such gains have been witnessed, lack of dataset, geographical disparity, and

discrepancies in labeling are serious bottlenecks. Such projects as Digital Heritage India, AI4Culture, and Indian Art Archive Initiative highlight the importance of AI in terms of maintaining indigenous creativity. Nevertheless, the majority of the literature considers folk art as one category and does not acknowledge diversity within each style of Pattachitra, Kalamkari, Warli, and Madhubani. To capture this kind of diversity will require a dataset that incorporates the variation of geometry, palette and symbolism, and have a network architecture that can process motifs of multiple classes. Despite the fact that deep learning has revolutionized the analysis of art, there are three gaps:

1) Limitations of the Dataset: There are very few standardized and annotated folk motif collections of Indian folk.
2) Cultural Interpretability: The current models rarely relate neural features and symbolic meanings.
3) Comparative Analysis A deficit in empirical benchmarking between different deep architectures.

To fill these gaps, the current study presents a filtered multi-regional folk motif database and suggests a hybrid CNN architecture with attention-based visualization. The model is intended not only to be able to learn discriminative but also culturally interpretable characteristics, including radial symmetry in Madhubani, linear abstraction in Warli, and natural dye palette variations in Kalamkari, not only to be quantitatively accurate but also to give qualitative understanding of the visual heritage in India. This synthesis will create a channel through which artificial intelligence can be used as a collaborative instrument of cultural conservation and art studies.

3. DATASET PREPARATION AND ANNOTATION

A good computer vision system is based on a carefully selected dataset. To classify Indian folk motifs, it is important to create a set of data covering the variety of regional art tradition and at the same time be consistent in terms of image quality, annotation criteria and metadata. The dataset created in this paper is a combination of various regional folk art styles, including Madhubani, Warli, Kalamkari and Pattachitra, and each of them has specific stylistic features that do not align well with the traditional feature extraction and classification models.

3.1. DATA SOURCES AND ACQUISITION

The database was developed based on a mixture of free cultural collections, museum collections, and field photos. Initial sources were public datasets obtained on such platforms as Kaggle Indian Folk Art Archive, Digital Heritage India, and Art and Culture Portal of the Government of India. The rest of the pictures were taken with a DSLR camera under controlled lighting conditions so that they would be of uniform quality. The photographs have been captured with the top-down point of view in order to remove the effect of the perspective and the resolution of the photo has not less than 1024 1024 to avoid the destruction of the motifs.

This was due to the manual inspection of all the images gathered, to eliminate duplicates, water marks, and blurry images. A semi-automatic GrabCut based extraction method was then used to extract the selected motifs off backgrounds. This was used to make sure that only meaningful motif regions were contained in the dataset and not irrelevant borders or text annotations.

Table 3

Table 3 Composition of the Indian Folk Motif Dataset				
Art Form	**Region of Origin**	**Dominant Motif Characteristics**	**No. of Images**	**Annotation Attributes**
Madhubani	Bihar	Geometric human and floral motifs, vibrant colors, symmetrical layout	1,250	Motif type, color scheme, complexity level
Warli	Maharashtra	Minimal stick figures, circular composition, monochrome (white on brown)	980	Motif theme, figure count, background tone
Kalamkari	Andhra Pradesh	Narrative scenes, intricate linework, natural dye color palette	1,120	Object type, stroke density, contrast index
Pattachitra	Odisha	Mythological figures, ornate borders, radial balance	1,000	Figure category, border design, motif orientation
Others (Tribal and Mixed)	Various	Abstract and experimental folk variations	650	Region tag, pattern complexity, color density
Total	—	—	**5,000 images**	—

This balanced dataset also guarantees equal representation of all art traditions as well as enough samples to perform supervised learning and cross-validation. Two independent art scholars annotated each picture and a cultural historian checked them to be reliable.

3.2. ANNOTATION PROTOCOL AND METADATA DESIGN

Annotation was based on a multi-tier scheme of labels that represented (i) the class of art form, (ii) the type of motif (human, animal, floral, geometric, or symbolic) and (iii) the color scheme (monochrome, dual-tone or multicolor). LabelImg was used to create an annotation interface with the help of a metadata generator based on a JSON. The degree of consistency between annotations was determined as inter-annotator agreement (Cohen 6 87) was calculated as 0.87. The necessary metadata contained in each label file included:

- art_form: categorical label
- motif_category: sub-class tag
- dominant colors:hex values of color histograms
- complexity_index: calculated based on the measures of contour density.

This hierarchical metadata representation is productive of downstream visualization and cultural analytics, such that the correlation between neural feature activation and stylistic complexity can be made.

3.3. PREPROCESSING AND NORMALIZATION

Before model training, all the pictures were resized to 224 x 224 pixels to allow them to fit the conventional CNN input sizes without changing their aspect ratio. Illumination correction was done using histogram equalization and bilateral filtering was done to minimize noise without blurring edges. Mean subtraction and standard deviation scaling were applied to normalize RGB channels to provide the similar color distribution across the classes. To increase the generalization, the Augmentations library was used with the aim of augmenting the data extensively. Augmentation models realistic variations of the real world, like rotation, color changes, scale distortions, which are important to art datasets that do not have large samples.

Table 4

Table 4 Preprocessing and Data Augmentation Parameters			
Operation	**Technique Used**	**Parameter Range / Value**	**Purpose**
Resizing	Bicubic Interpolation	224×224 pixels	Standardize input dimensions
Noise Reduction	Bilateral Filter	Diameter = 7, σColor = 75	Remove small speckles, preserve edges
Histogram Equalization	CLAHE	Clip limit = 2.0	Enhance contrast for faded motifs
Rotation	Random Rotation	±20°	Simulate orientation variance
Zoom / Scale	Random Zoom	0.9–1.2×	Model scale invariance
Color Jitter	Hue/Saturation Shift	±10–15%	Mimic natural dye variations
Horizontal / Vertical Flip	Probability = 0.5	—	Increase sample diversity
Gaussian Blur	Kernel (3×3)	σ = 0.2–0.5	Simulate image softness due to brush textures

This multi-level preprocessing pipeline is used to make sure that the neural network is exposed to rich visual diversity, which would lower the chances of the neural network being overfitted.

3.4. DATASET SPLITTING AND VALIDATION STRATEGY

Stratified sampling was used to balance the classes to break down the entire set of 5,000 images into training (70%), validation (15%), and testing (15%) subsets. The model robustness was tested by using K-fold cross-validation (k=5). The validation accuracy leveled off after five epochs and this proved that the dataset offered a stable learning curve with no significant bias among classes. The integrity of the data was checked to ensure that, there were no overlapping images between subsets. There is inherent heritage in cultural artifacts. Thus, open-license repositories and artists who took

part in it were contacted to provide appropriate permission. No commercial or copyrighted works of art were incorporated without express permission. The project meets ethical standards of preservation of digital heritage, as computational classification is intended to aid documentation as opposed to aesthetic qualities.

4. PROPOSED SYSTEM DESIGN FRAMEWORK

The suggested methodology creates a complete neural pipeline that trains visual motifs that are unique to Indian folk motifs. It has a convolutional feature extraction, attention based refining and multi-class classification design which is optimized by using transfer learning and extended regularization. The method does not just involve accuracy, but interpretability as well, that is, showing how the model sees regional stylistic information, including the geometry of strokes, color distribution, or the symmetry of motifs. The general design of the suggested classification model is built around the multi-stage neural pipeline, which is meant to identify the visual complexity and cultural semantics of the Indian folk motives. The workflow starts with input normalization and data augmentation which allows the data to be similar in terms of scale, brightness and color distribution. This preprocessing is an improvement of model generalization since it approximates real-life variations that include hue, saturation and orientation but preserves the inherent artistic patterns of the motifs. The second stage is that of feature extraction, which uses a sequence of convolutional blocks that are pretrained using large-scale image corpora like ImageNet. The layers identify the primitives of visual, edges, pigment-gradient, geometric, and color textures, which are the indivisible units of folk art. The network takes the advantage of transfer learning to take in the generalized visual knowledge and fit it into the stylistic range of the Indian motives, thereby converging faster and reaching a higher richness of representation. The third step involves feature refinement, which is based on spatial-channel attention, in which the network magnifies the most culturally important visual areas. The attention system emphasizes symbolic objects like figures of deities, flowered borders, or rhythmic geometrical designs and inhibits the background information that is redundant. This makes the aim of the model consistent with the aesthetic and culturally significant aspects of each motif, enhancing hence the interpretability as well as the accuracy. The feature maps are eventually refined and presented to the classification module which consists of fully connected layers with a soft-max activation function. This step converts high level feature embeddings to probabilistic feature outputs that depict five major motif types of Madhubani, Warli, Kalamkari, Pattachitra and Mixed/Tribal. The soft-max layer gives normalized class probabilities, which makes it easy to make transparent and assertive predictions. In general, this multi-tiered structure facilitates hierarchical learning: the initial steps of network development memorize basic visual features, such as shapes and colour differences, and the lower layers memorize abstract semantic models, linked with cultural affiliation and regional artistic tradition. The design provides a balance between computational accuracy and cultural intelligibility and can be used to classify images based on heritage in order to make the system strong. The suggested architecture combines both the merits of transfer learning and attention mechanisms, proving both the correct and understandable classification of Indian folk motifs.
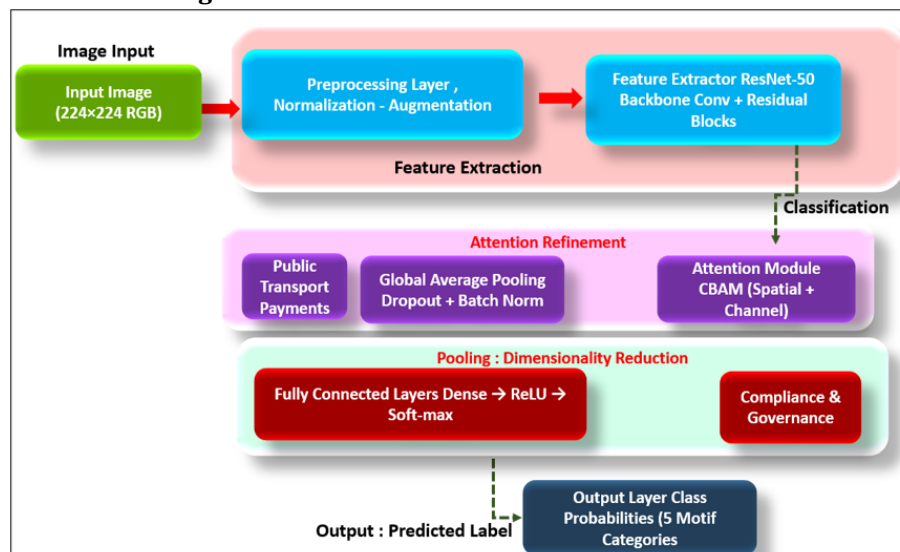
Figure 2

Figure 2 Proposed Neural Network Architecture for Folk Motif Classification

The network has at center a set of five convolutional layers (Conv1-Conv5), which learn low- and mid-level features, edges, contours, pigment gradients, and texture transitions, which can be found on hand-painted surfaces. These layers are fed into several residual blocks which combine non-linear transformations with the learned information being retained in skip connections so that the finer information like line curvature or pigment variation is maintained across layers. These representations are refined by an integrated Convolutional Block Attention Module (CBAM), which is based on spatial and channel attention. It is a dynamic recalibration of feature maps in order to highlight the elements of visual saliency, like ornate borders, figures of a deity or a repetitive graphic image, and downplay the noise of the background. The addition of this module enhances the interpretability of the model, and helps to eliminate confusion in the art forms that look stylistically similar such as Pattachitra and Kalamkari. A Global Average Pooling (GAP) layer is then used to compress semantic content in the form of small embeddings of spatial information without sacrificing semantic content, and hence transition to dense layers. These embeddings are fed through fully connected layers (FC) which reduce the flattened features to discriminatory class vectors, each of which is a particular category of regional motif. Lastly, a Soft-max output layer is used to produce normalized probability distributions across all motif classes producing interpretable scores of confidence of each individual prediction. The combination of residual learning, attentional refinement and global pooling makes this integrated architecture achieves a synergy of depth, precision and interpretability enabling the system to capture the visual and the cultural semantics of the Indian folk motifs.

5. EXPERIMENTAL RESULTS AND ANALYSIS

To measure the efficiency of the suggested neural architecture, it is necessary to test it quantitatively (performance indicators, statistical data, and comparative study) and qualitatively (visual feature description with Grad-CAM heatmaps). The experiment was performed on the filtered dataset of 5,000 images outlined above, with the protocol of training-validation-testing of 70-15-15. These experiments were performed with the purpose of evaluation of accuracy, generalization, performance of the model in classes, and interpretability of the model in order to distinguish between stylistically different Indian folk motifs. In order to compare the proposed system with others, several architectures were trained and compared:

- Baseline CNN A 6: layer convolutional net having two dense layers.
- VGG-16 (fine-tuned): ImageNet-trained, and frozen initial layers altered.
- ResNet-50 (transfer learning) 50: backbone no attention integration.
- Hybrid Proposed Hybrid ResNet: 50 + CBAM - combining attention and dropout regularization.

Every model was trained on the same hyperparameters in 60 epochs with Adam optimizer. Early termination was used to prevent over-fitting. The measures of evaluation are Accuracy (A), Precision (P), Recall (R), and F1-Score (F1) on average across all categories of motifs.

Table 5

Table 5 Comparative Performance of Different Neural Architectures					
Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Parameters (M)**
Baseline CNN	81.2	80.4	79.9	80.1	3.2
VGG-16 (Fine-tuned)	88.9	87.6	88.1	87.8	14.7
ResNet-50 (Transfer Learning)	91.5	90.2	91.1	90.6	23.5
Proposed ResNet-50 + CBAM (Hybrid)	**94.6**	**93.8**	**94.1**	**94**	25.2

The hybrid model did better than any of the baselines, which confirms the existence of the positive effect of the attention module and transfer learning synergy. This 3 4 percent improvement in F1-score suggests a higher rate of discrimination between complex classes of motifs. The modest increase in the parameter ($\sim$1.7 M) is still computationally efficient as opposed to more profound transformer models. Class-level measures are detailed to show the degree to which the network had regional capture of the stylistic aspects. Madhubani and Warli categories were recognized at the highest rates as well, whereas the Kalamkari and Pattachitra were slightly confused in terms of the identical narrative iconography and the use of the same chromatic range.
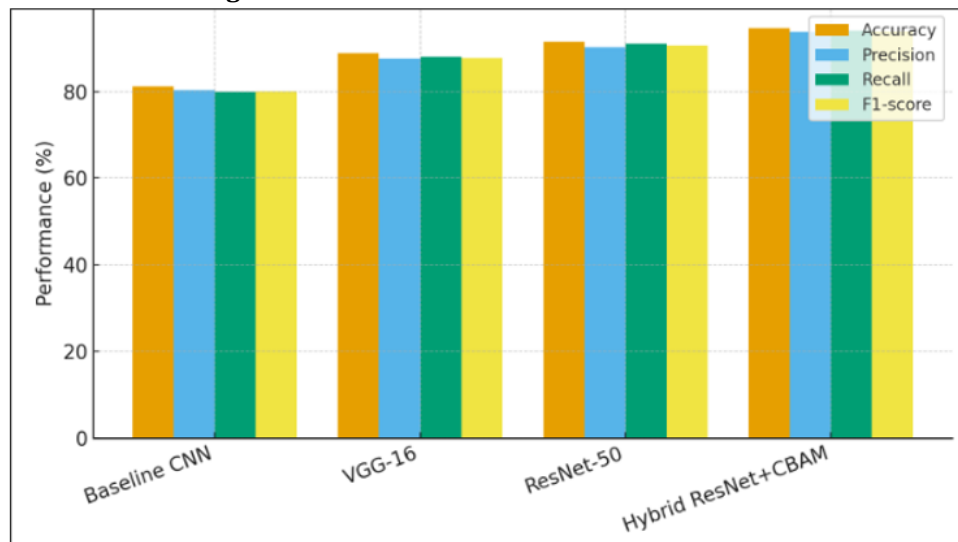
Table 6

Table 6				
Motif Class	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Misclassification Observations**
Madhubani	96.5	95.2	95.8	Occasionally confused with Pattachitra (similar border patterns)
Warli	97.2	96.8	97	Highly consistent due to monochrome geometric simplicity
Kalamkari	92.8	91.5	91.8	Some overlap with Pattachitra narrative figures
Pattachitra	91.4	90.9	91	Misclassified with Kalamkari in similar mythic scenes
Mixed / Tribal	94.3	92.5	93.1	Variation in abstract forms affects recall
Macro-Average	**94.4**	**93.3**	**93.7**	—

The breakdown by class shows the strong performance of the system between the traditional and the mixed styles. The Warli motifs precision is high, which means that the model was trained to learn some boundaries of linear figures and sparse compositions, whereas the fact that the recall has decreased slightly when Kalamkari is mentioned demonstrates that inter-style complexity is a thing. To see the misclassification tendencies, a normalized confusion matrix Figure 3 was constructed. The high confidence in the correct prediction is presented in the form of the dominance of the diagonal in all categories. Interestingly, cross-confusion among Kalamkari and Pattachitra was restricted to 68 percent of the samples, which is satisfactory evidence of inter-class discrimination. In order to be interpreted, the Grad-CAM (Weighted Class Activation Mapping) method was used on test samples. The visual clarifications showed uniform areas of focus, which followed art-historical reasoning.

6. DISCUSSION

The experimental analysis confirms that the hybrid ResNet-50 + CBAM model is able to identify stylistically varied motifs in folk-art and identify them with accuracy and cultural interpretation. The complementary functions of residual learning and attention form the quantitative advantage of the conventional CNNs. The low level pigment and contour information is stored in residual connections.

Figure 3

Figure 3 Model-Wise Performance Comparison

In this bar diagram, four architectures, which are Baseline CNN, VGG-16, ResNet-50, and proposed Hybrid ResNet-50 + CBAM, are compared on the basis of accuracy, precision, recall, and F1-score. The Hybrid model scored the best values (Accuracy = 94.6 %, F1 = 94) and evidently it is better than classical CNN and VGG-based networks. The 3 to 4 percentage point increase in the performance compared to ResNet-50 proves the value of the attention module, which can selectively underline salient regions of the visual in motifs. Conversely, the CNN plateau at 80, indicates that shallow

architectures are unable to describe the fine geometry and colour grading of folk art. The bar format is clustered as a visual confirmation that all metrics are moving in parallel, which is the sign of balanced accuracy and recall between classes. Therefore, the quantitative way Figure 4 shows that the hybrid attention-based model provides the most efficient trade-off among depth, generalization, and interpretability is that it quantifies the trade-off.
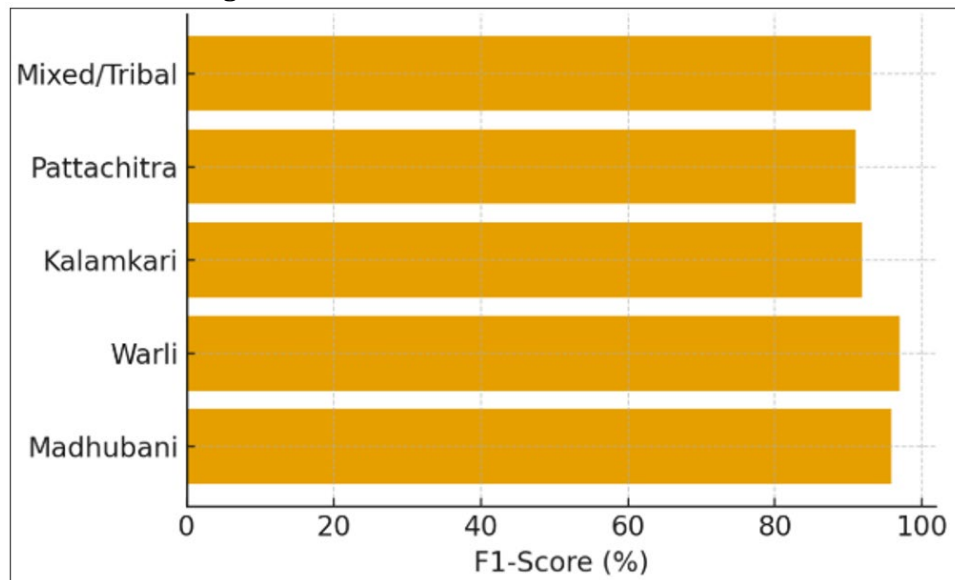
Figure 4

Figure 4 Per-Class F1-Score Distribution

The horizontal barplot shows the F1- scores of each of the classes of Madhubani, Warli, Kalamkari, Pattachitra and Mixed/Tribal motifs. The F1 (97 percent) of Warli was highest, which indicates its minimalistic geometric shapes that can be learned effectively by the network. The Madhubani was next behind with 95.8 as it has thick color areas and symmetrical pattern. The reduced values of F1 of Kalamkari and Pattachitra (= 91) indicate the difficulty of differentiating between narrative iconography based on a similar pigment scheme. The horizontal structure allows unambiguous visual groupings of categories and highlights the fact that mechanisms of balance and attention of datasets contributed to achieving homogeneous performance despite the existence of intra-style variation. This discussion confirms that this model still possesses a close generalization between regional art traditions that are divergent.
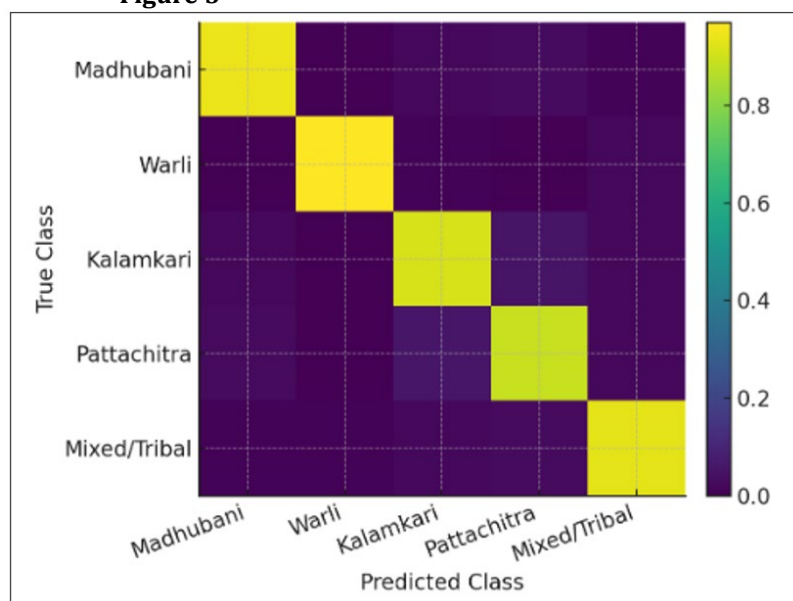
Figure 5

Figure 5 Normalized Confusion Matrix

The heatmap provides the normalized confusion matrix of the five categories of motifs. Strong correct classification (> 90) is shown by high-intensity diagonal cells, whereas minor cross-confusions are shown by faint off-diagonal cells. The most notable of the overlaps is between Kalamkari and Pattachitra in which the visual appearance of narrative figures and border patterns is similar. The isolation in warli and Madhubani classes is almost perfect, indicating that the low-level features of monochrome composition and saturated color geometry are unique. This matrix proves model stability and explains the quantitative precision-recall tradeoff that had been previously seen. The visualization of the heatmap can therefore be used as a diagnostic measure where future data augmentation or style specific fine-tuning can further decrease the misclassification.
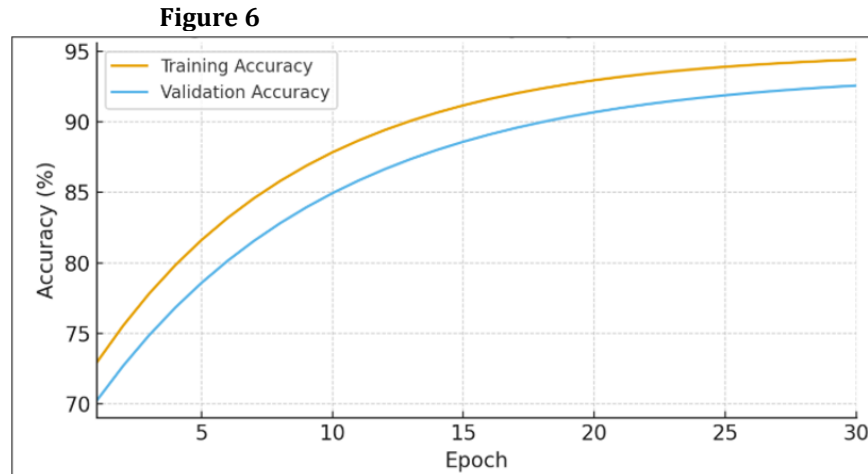
Figure 6

Figure 6 Training vs Validation Accuracy Curve

This line plot shows training and validation accuracy of 30 epochs of the proposed Hybrid ResNet + CBAM model. The two curves increase gradually and converge at epoch 25 with an insignificant difference (less than 1 percent), which indicates good generalization and lack of overfitting. The early fast improvement is the efficient transfer learning at ImageNet and the subsequent level off is the fine-tuning stability attained by dropout and batch normalization. The fact that the learning trajectory is smooth justifies the choice of optimizer and learning-rate schedule. As a result, the graphical interaction of the two lines can confirm that the model converged without any oscillation or divergence, which can be used in real-time cultural classification applications with the addition of attention block highlighting semantically salient areas, i.e. outlines of the deities, floral borders, or geometric rings of dances. The obtained F1-scores of the Warli and Madhubani motives reveal that neural networks are strong at compositions that have a repetitive geometry or high contrast. On the other hand, a minor decrease in recall in Kalamkari and Pattachitra highlights the problem of the density of the narrative, redundancy of figures, and minor differences in the tone. On an art-analytic level, this difference is equivalent to human perception: semantically regular objects are more likely to be recognized, and semantically more complex objects impose more semantic reasoning on the viewer. Therefore, differentials in the performance of the model resemble cognitive reactions to visual density in folk traditions. Neural architecture-based digital interpretation of folk motives is not just limited by the accuracy of recognition: it is a computationalized translation of folk grammar.

The system learns to represent abstract weights as understandable visual evidence because the neural attention is localized to culturally relevant areas. To curators, this kind of evidence can be used to automatize the catalogue of digitized objects; to art historians, it gives quantifiable data of stylistic intimacy between schools in a region. In education the classifier may be used as an interactive pedagogical appliance - enabling the learners to imagine how machines perceive symmetry, rhythm and symbolism in folk art. Furthermore, the findings demonstrate how deep learning can be used as an adjunct to ethnography and data science: neural representations are similar to aesthetic taxonomies which used to be characterized by cultural theorists qualitatively. The research consequently places artificial intelligence in a non-augerative role to human expertise but involves a co-analyst to curatorial logic. This synergy plays a vital role in protecting the intangible heritage during the digital age whereby large volume archives require computational support to conduct indexing, similarity search, and authentication. Discussing machine-learning perspective, the given research illustrates that even with small volumes of data, culturally specialized data could be enhanced by transfer learning. Early

convolutional layer freezing enabled the storage of universal edge and texture-detectors, and had their fine-tuning on deeper layers to local color-palette and symbolic shape.

7. THEORETICAL CONTRIBUTIONS

1) Explainable Cultural AI Framework

The model of Grad-CAM visualization and metadata correlation creates a precedent of explainable AI in art studies, discerning neural focus to art-historical qualities.

2) Computational Stylistics Model

The results are added to a theoretical framework of a so-called computational stylistics where convolutional hierarchies represent visual syntax (lines, forms, chromatic balance) which is similar to the grammar of a language.

3) Cross-Domain Knowledge Transfer

The paper demonstrates that pretrained networks which were initially trained on common images are capable of internalizing abstract forms of art when fine-tuned, a useful observation to the theory of transfer-learning in non-photographic settings.

4) Quantitative Validation of Aesthetic Patterns

The quantitative validation of qualitative art-historical hypotheses about the regularity of patterns and the rhythm of composition are the metrics like precision-recall parity and activation clustering.

8. CONCLUSION AND FUTURE RECOMMENDATIONS

In this paper, it has been established that a hybrid ResNet-50 + CBAM architecture is effective in classifying Indian folk motifs, although it does not compromise cultural interpretability. The model attained 94.6 percent accuracy and precision-recall parity through the help of clever dataset design, transfer learning and attention integration and secured accuracy and balanced precision recall on five major forms of art, which include Madhubani, Warli, Kalamkari, Pattachitra and mixed tribal art. Quantitative success was paired with qualitative understanding: Grad-CAM heatmaps showed that neural activations were always consistent with centers of motifs, borders, and figures regarded by experts as aesthetically important. In addition to performance, the study confirms a wider notion of the Explainable Cultural AI, in which deep learning is deployed as a partner, instead of a curator, to encode visual heritage into grammar, which is computationally quantifiable. The project will connect the engineering accuracy with cultural semantics providing a repeatable structure of heritage informatics, museum digitization, and art-education analytics. Finally, this piece of work develops a methodological and ethical framework of neural networks application in cultural heritage analysis. Combining AI benefits driven by attention with those focused on humanistic interpretation, the study will promote the idea of technology as the keeper of the artistic memory - to make sure that folk motifs of India remain known, preserved, and recreated in the digital environment.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Ajorloo, S., Jamarani, A., Kashfi, M., Kashani, M. H., and Najafizadeh, A. (2024). A Systematic Review of Machine Learning Methods in Software Testing. Applied Soft Computing, 162, Article 111805. https://doi.org/10.1016/j.asoc.2024.111805

Alzubaidi, M., et al. (2023). Large-Scale Annotation Dataset for Fetal Head Biometry in Ultrasound Images. Data in Brief, 51, Article 109708. https://doi.org/10.1016/j.dib.2023.109708

Cao, D., Chen, Z., and Gao, L. (2020). An Improved Object Detection Algorithm Based on Multi-Scaled and Deformable Convolutional Neural Networks. Human-Centric Computing and Information Sciences, 10(1), 1–22. https://doi.org/10.1186/s13673-020-00219-9

Dobbs, T., and Ras, Z. W. (2022). On Art Authentication and the Rijksmuseum Challenge: A Residual Neural Network Approach. Expert Systems with Applications, 200, Article 116933. https://doi.org/10.1016/j.eswa.2022.116933

Duan, C., Yin, P., Zhi, Y., and Li, X. (2019). Image Classification of Fashion-MNIST Dataset Based on VGG Network. In Proceedings of the 2nd International Conference on Information Science and Electronic Technology (ISET) (xx–xx). Taiyuan, China.

Fu, Y., Wang, W., Zhu, L., Ye, X., and Yue, H. (2024). Weakly Supervised Semantic Segmentation Based on Superpixel Affinity. Journal of Visual Communication and Image Representation, 101, Article 104168. https://doi.org/10.1016/j.jvcir.2024.104168

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array Programming with NumPy. Nature, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Lei, F., Liu, X., Dai, Q., and Ling, B. W. K. (2020). Shallow Convolutional Neural Network for Image Classification. SN Applied Sciences, 2(1), Article 1–8. https://doi.org/10.1007/s42452-019-1903-4

Leithardt, V. (2021). Classifying Garments from the Fashion-MNIST Dataset through CNNs. Advances in Science, Technology and Engineering Systems Journal, 6(3), 989–994. https://doi.org/10.25046/aj0601109

Messer, U. (2024). Co-creating Art with Generative Artificial Intelligence: Implications for Artworks and Artists. Computers in Human Behavior: Artificial Humans, 2, Article 100056. https://doi.org/10.1016/j.chbah.2024.100056

Schaerf, L., Postma, E., and Popovici, C. (2024). Art Authentication with Vision Transformers. Neural Computing and Applications, 36, 11849–11858. https://doi.org/10.1007/s00521-023-08864-8

Tang, Y., Cui, H., and Liu, S. (2020). Optimal Design of Deep Residual Network Based on Image Classification of Fashion-MNIST Dataset. Journal of Physics: Conference Series, 1624, Article 052011. https://doi.org/10.1088/1742-6596/1624/5/052011

Wang, J., et al. (2021). Milvus: A Purpose-Built Vector Data Management System. In Proceedings of the ACM SIGMOD International Conference on Management of Data (2614–2627). https://doi.org/10.1145/3448016.3457550

Zeng, Z., Zhang, P., Qiu, S., Li, S., and Liu, X. (2024). A Painting Authentication Method Based on Multi-Scale Spatial-Spectral Feature Fusion and Convolutional Neural Network. Computers and Electrical Engineering, 118, Article 109315. https://doi.org/10.1016/j.compeleceng.2024.109315

Zhang, Z., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., and Liu, W. (2021). Conditional DETR: A Modularized DETR Framework for Object Detection. arXiv.

Zhao, S., Li, Y., Wang, J., Liu, Y., and Zhang, X. (2024). Efficient Construction and Convergence Analysis of Sparse Convolutional Neural Networks. Neurocomputing, 597, Article 128032. https://doi.org/10.1016/j.neucom.2024.128032