Journal of Applied Data Sciences
Vol. 7, No. 2, May 2026, pp. 1020-1035

ISSN 2723-6471
1020

Pattern Recognition of Puta Dino Fabric Using Web-Based Convolutional Neural Network Method

Luther Alexander Latumakulita[1,*], Silviani Esther Rumagit[2], Hence Beedwel Lumentut[3], Frangky Jessy Paat[4], Jaidun Ramadhan Kaplale[5], Enny Itje Sela[6]

[1,5]*Information Systems Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sam Ratulangi Jl. Kampus Unsrat , Manado 95115, Indonesia*

[2,3]*Universitas Baliem Papua, Jln. SD Percobaan No. 50 Wamena, Papua, Indonesia*

[4]*Faculty of Agriculture, Universitas Sam Ratulangi, Jl. Kampus Unsrat , Manado 95115, Indonesia*

[6]*Information Technology Master's Program, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia*

(Received: October 15, 2025; Revised: December 10, 2025; Accepted: March 8, 2026; Available online: April 4, 2026)

Abstract

This study aims to develop an intelligent system capable of recognizing traditional woven motifs of Puta Dino, a culturally significant textile from Tidore Island. These motifs are visually complex, poorly documented, and hard for the public to distinguish, highlighting the need for a digital tool to support cultural preservation and accurate identification. This research is the first to build a structured Puta Dino motif database and provide an integrated model designed for real-world use. The approach captured primary images of eight validated motifs and applied systematic preprocessing, including normalization and data augmentation, to enhance variability and strengthen the dataset. A lightweight deep learning model predicated on a convolutional neural network was designed to achieve a compromise between accuracy and computational efficiency. The system was evaluated through cross-validation and independent test data, as well as multiple real-world trials utilizing a web interface. These trials involved different image capture scenarios, including from a distance, moderate distance, close and angled views, and when the fabric surface was folded. The model architecture and system interface with the system are illustrated in the relevant figures, and the tables provide performance data on the system's training, accuracy in motif classification, and achieved results in real-world conditions. The system demonstrated excellent classification accuracy in controlled test conditions. It showed real-world competency, accurately classifying most motifs in various conditions. The data also point to specific issues with motif recognition in extreme distortion cases, which reflect the typical issues of laboratory-to-field model deployment. The outcomes clearly demonstrate both the possibilities and the limitations of the currently available recognition of culturally significant textiles. The study concludes by exploring the possibilities of expanding the dataset and increasing the depth of learning through more sophisticated techniques, as well as enhancing accessibility to promote sustained community and cultural engagement.

Keywords: Puta Dino, Motif Recognition, deep learning, Cross Validation, Tidore Island.

1. Introduction

Puta Dino is a traditional woven textile from Tidore Island, North Maluku, Indonesia, representing an important aspect of local cultural identity and heritage. The terminology associated with Puta Dino reflects indigenous linguistic expressions rooted in traditional weaving practices, which, as noted in textile studies, are often shaped by diverse cultural and historical contexts [1], [2] . Beyond its functional role, traditional textiles serve as cultural symbols that embody social values and collective memory within communities [3], [4]. However, like many traditional textiles, Puta Dino experienced a period of decline due to limited documentation and modernization, leading to the adoption of fabrics from other regions in ceremonial practices. This phenomenon highlights the need for revitalization efforts to preserve and sustain traditional textile heritage [5]. In preserving Tidore culture and promoting Puta Dino the textile has been included in several exhibitions both abroad and nationally including G20 Future SMEs Village 2022 in Bali,

***Corresponding author: Luther Alexander Latumakulita (latumakulitala@unsrat.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i2.1103

ISSN 2723-6471

Indonesian Cultural Awards (AKI) 2023 in Jakarta, Indonesia Africa Forum (IAF) 2023 in Bali, exhibitions in Cape Town, Indonesian Wastra and Culture celebrations in Paris, and most recently in 2025 at the Nusantara Fine Arts Exhibition organized by the Nusantara Arts and Culture Committee (KSBN). These exhibitions introduced Puta Dino to many people, preserving the textile culture of Tidore and opening new economic opportunities for the Tidore weavers.

Despite its growing recognition, Puta Dino possesses a wealth of motifs that are not easily recognizable or memorable to everyone. There are approximately 14 motifs in traditional Tidore Puta Dino, including Jodati, Marasante, Barakati, Malila, Tuan Guru, Kalajengking, Amo, Cengkeh, Gomode Mabunga, Laha-laha, Dodara, Marimoi, Nyili Mou, and Tobaru, each with distinct meanings. Although traditional accounts recognize 14 Puta Dino motifs, formal scientific documentation of these motifs is currently unavailable. Field verification conducted through consultations with cultural elders and surveys at traditional markets revealed that only 8 of the 14 motifs are still produced and physically obtainable today. The remaining six motifs are considered rare or no longer woven, making it impossible to collect sufficient image samples for inclusion in the dataset. Consequently, the system developed in this study focuses on the eight motifs that are culturally validated and still exist in contemporary production. While this limitation may affect cultural completeness, it does not diminish the system's relevance; instead, it highlights the urgency of documenting endangered motifs and provides a foundation for future expansion once rare motifs become available for digitization.

Motif recognition remains a significant challenge. Intricate and highly detailed patterns often look similar to individuals without prior knowledge, while accessible information regarding motif names and their meanings is still limited in available media [6], [7]. At present, motifs are primarily introduced through printed leaflets that include an image along with a short description. However, these images are often too small and lack sufficient clarity to effectively convey the finer details of the designs [8] . Furthermore, the leaflet format—typically comparable to a small booklet—is not easily portable, making it an impractical medium for disseminating information [9], [10].

In this regard, it is evident that other, more versatile means of media are warranted to aid communities in recognizing Puta Dino fabric motifs and their meanings. Considering the advancements in digital technology, the use of smart systems is suitable for identifying objects, particularly the Puta Dino motif. A Convolutional Neural Network (CNN) is a type of artificial neural network that is particularly popular in the field of image data processing. CNN captures patterns with various forms of deviations and is highly resistant to distortions and basic shape alterations [11]. Due to these factors, this approach is suitable for recognizing Puta Dino fabrics and deciphering the meanings of the motifs. To facilitate the effective application of the system, a web-based platform was chosen for disseminating information, accessible from any mobile device, providing access to information about the motifs at any time and from any place.

Previously, studies have demonstrated that CNN can detect and identify specific patterns or objects within a field of view, making it an appropriate technique in this instance. This research aims to develop an intelligent system, specifically in the application of a deep learning CNN using the MobileNetV2 framework, to identify motifs or patterns in Puta Dino, a traditional woven fabric of the Tidore people. This is expected to make the motifs in these traditional fabrics more well-known and appreciated within the context of Indonesian culture in the city of Tidore Islands.

Currently, no scholarly work has been undertaken on the automated recognition of Puta Dino motifs; thus, there is an obvious research and technological gap concerning the application of computer vision to Tidore's cultural heritage. The motifs exhibit high intra-class variability, primarily due to variations in lighting, camera angle, fabric folds, and weaving inconsistencies. However, the motifs have strong inter-class similarities, which makes pattern recognition more difficult and requires a more sophisticated approach to image classification.

To address this gap, this study presents a lightweight deep learning approach using MobileNetV2, designed for practical and online settings. The uniqueness of this research consists of the following: (1) the first research project to approach Puta Dino motifs recognition as a challenge in computer vision, (2) the first researcher to build a validated primary dataset consisting of eight different motifs, and (3) the first to employ an efficient MobileNetV2 model geared for practical community service.

The primary objectives of this study are to develop a CNN-based classifier for the identification of eight distinct Puta Dino motifs and to assess the model's performance through 5-fold cross-validation and testing on an independent dataset. Furthermore, the study aims to integrate the trained model into a functional web-based motif recognition application and evaluate its performance under various challenging real-world image capture conditions.

All the elements arise from the same basis that justifies setting the research problem, delimiting the technical gap, and determining the study's contribution to image recognition in lightweight and real-time applications for the preservation of cultural heritage. This research aims to mitigate the lack of real-time automated recognition of the Puta Dino motif by deploying lightweight deep learning. This study is the first, and it primarily contributes to formalizing the recognition of the Puta Dino motif as a problem in computer vision. Additionally, it is the first to create a primary database with eight culturally authenticated motifs and to utilize MobileNetV2 for this purpose, enabling the deployment of the model with the efficiency required for real-time web applications. This novelty is aligned with the objectives of this research, which are to implement a model based on CNN to classify Puta Dino motifs and to validate its efficiency using a controlled cross-validation approach. The model is also intended to validate its performance on a standalone test of the web application, which is designed for community use, as well as to evaluate its performance under various conditions when photographs are taken in the real world.

2. Literature Review

Applying cutting-edge deep learning techniques to traditional fabric pattern recognition involves academic research methodologies for cultural heritage preservation. Deep learning neural network models are deployed to automate the detection and preservation of cultural artifacts, as different textile patterns are recognized and classified across various cultures.

Using CNN, Rasyidi et al. [12] studied and recognized batik patterns, applying multiple models to focus on six distinct patterns: Batik Banji, Batik Ceplok, Batik Kawung, Batik Mega Mendung, Batik Parang, and Batik Sekar Jagad. Of the total 994 images, an 80-20 split was used at an initial stage to create a training and test set for Convolutional Neural Networks and DenseNet algorithms. Rasyidi had the best test score. His studies were then accepted as the first pioneering research in Batik, utilizing deep learning applications to traditional textiles in Indonesia. Likewise, study [13] applied a CNN method to classify Lombok songket and Sasambo batik motifs, achieving an accuracy of around 90–95% using parameters such as 50–100 epochs, a learning rate of 0.001, batch size of 32, and multiple convolutional layers.

Despite these studies demonstrating the utility of CNN-based approaches for recognizing traditional textiles, such as Batik and songket, they rely on established datasets with consistent visual patterns and well-documented motifs. In contrast, Puta Dino motifs are culturally sensitive, underdocumented, and in some cases, rare and difficult to access. No previous studies have attempted to formalize the recognition of Puta Dino motifs as a computer vision task, which represents a cultural and methodological void. This research aims to fill that void by creating the first computer vision dataset for eight Puta Dino motifs that are culturally verified, and by deploying recognition models based on MobileNetV2, optimized for real-world settings. This research improves upon previous studies on Batik and ulos by focusing on a complex and visually diverse textile tradition and by providing a refined model for web-based applications. This work thus helps the textile pattern recognition research both culturally and technically.

This study's contribution lies in its focus on a more culturally unique and technically challenging tradition of textiles, whose motifs are underdocumented and facing thematic extinction. Building the first organized dataset, complete with an automated recognition model for Puta Dino documents, is a contribution not only to the field of computer vision but also to the efforts of cultural preservation, as the automated recognition model is developed within the dataset. Therefore, this study finds itself in a unique position, situated at the intersection of cultural preservation and cutting-edge deep learning technology.

3. Method

The dataset for this study consists of data first collected from Puta Dino Kayangan weaving house, the traditional Puta Dino fabric production center in Tidore. A total of 810 images spanning eight motifs, or designs, namely, Amo, Barakati, Dodara, Jodati, Kalajengking, Malila, Marasante, and Tobaru, were collected. The distribution of the dataset comprises 105 images for each of the motif classes, except for Barakati, which has 75 images due to the lesser occurrence of this motif. The images were captured in JPEG format under natural lighting using a 64MP smartphone camera and stored on the collection site. To maximize the dataset's diversity, enabling the model to learn the differences in the images, photos were taken from various angles, distances, and viewpoints of the fabric motifs. Substandard images with inadequate lighting conditions, motion blur, or incomplete motifs were also removed to maintain dataset quality and ensure that the final dataset consisted of images suitable for pattern recognition training.

3.1. Data Preprocessing and Augmentation

In this part of the preprocessing phase, the acquired primary data were separated into three sets: web testing data, testing data, and data for training and validation of the model. The web testing data consisted of 5 images per class, chosen to represent five diverse instances in web testing. This resulted in leaving aside 40 images for the web assessment. For the other 770 images, an 8:2 split was performed, where 80% (616 images) were used for preprocessing and in the training and validation phases. The remaining 20% (154 images) were used as testing data. The training and validation data were also divided using 5-fold cross-validation for a more robust model evaluation.

The images were resized to 224 pixels by 224 pixels to match the MobileNetV2 architecture's input requirements. Through Pixel Normalization, each pixel value was divided by 255 to change the value's range from 0-255 to 0-1. This normalization was performed during model training using ImageDataGenerator to improve stability and accelerate the rate of convergence. The Albumentations library was used to perform transformations to the original images to conduct Data Augmentation where in this case, the available transformations used in this research includes the following: 1. Manipulation of lighting and colors where Random Brightness Contrast was used to change the images brightness and contrast, Hue Saturation Value which changes the hue digitally by value of -10 to +10, alters saturation by -20 to +20, adjusts brightness by -10 to +10 to enhance the model's resiliency against changes in lighting, and RGB Shift which shifts the colors of the image by range of -15 to +15 in all the RGB channels to introduce color variations and 2. The addition of Gaussian noise, which introduces noise of varying intensity from 10.0 to 50.0, helps the model become more robust to images of varying quality.

In line with data balancing, classes with fewer than 300 images were augmented until the target of 300 images was reached. This made it possible to keep the images as data of the original samples. This also helped provide variation in the data to improve the training results. Two thousand four hundred images were obtained through augmentation, with 300 per class across eight classes. The data obtained was then subjected to 5-fold cross-validation, with each fold maintaining an 8:2 split of training and validation data, resulting in 1920 images for training and 480 images for validation in each fold.

3.2. Model Architecture

In this research, using a Convolutional Neural Network (CNN) with a MobileNetV2 architecture will aid in recognizing the motifs in Puta Dino. MobileNetV2 has numerous lightweight candidate models, such as EfficientNet-B0 and SqueezeNet, which are commonly deployed in mobile and web applications. Therefore, a comparative analysis of such models was conducted, focusing on the characteristics of the Puta Dino dataset in table 1 to select the most appropriate architecture.

Table 1. Comparison of Lightweight CNN Architectures

Model	Strengths	Limitations	Suitability for This Study
MobileNetV2	Very lightweight; fast inference; strong texture-pattern extraction; optimized for mobile/web applications	Slightly lower peak accuracy than EfficientNet on large datasets	Most balanced choice for small datasets, fine-grained motif patterns, and real-time web deployment
EfficientNet-B0	Higher accuracy on large and diverse datasets; advanced compound scaling	Heavier, slower inference; requires more computational resources	Overcapacity for this dataset; unnecessary latency for browser-based implementation
SqueezeNet	Extremely small model size (<5 MB); ultra-fast inference	Lower accuracy for fine-grained classification; weaker sensitivity to subtle geometric variations	Not suitable due to high inter-class similarity between Puta Dino motifs

MobileNetV2, as shown in table 1, demonstrates the most favorable balance of accuracy, efficiency, and practicality in deployment. It contains sufficient feature extraction capabilities to delineate the geometric and textural details of the Puta Dino motifs while retaining a sufficiently lightweight structure for quick inferences in a web system. EfficientNet-B0 has a higher accuracy level, but the Puta Dino dataset isn't large enough to warrant the computational inefficiencies brought on by the dataset's overhead. SqueezeNet does feature a lightweight architecture, but unfortunately, it lacks sufficient representational power to distinguish motifs with subtle visual differences. Therefore, for accurate and deployable motif classification at the required efficiency, MobileNetV2 should be the most appropriate architecture.

The positive aspects of MobileNetV2's structure justify its choice due to its relative compromise of performance and accuracy in terms of computation, as well as suitability for deployment in Web Applications [14], [15]. There are feature processing and classification sections as reflected in figure 1.

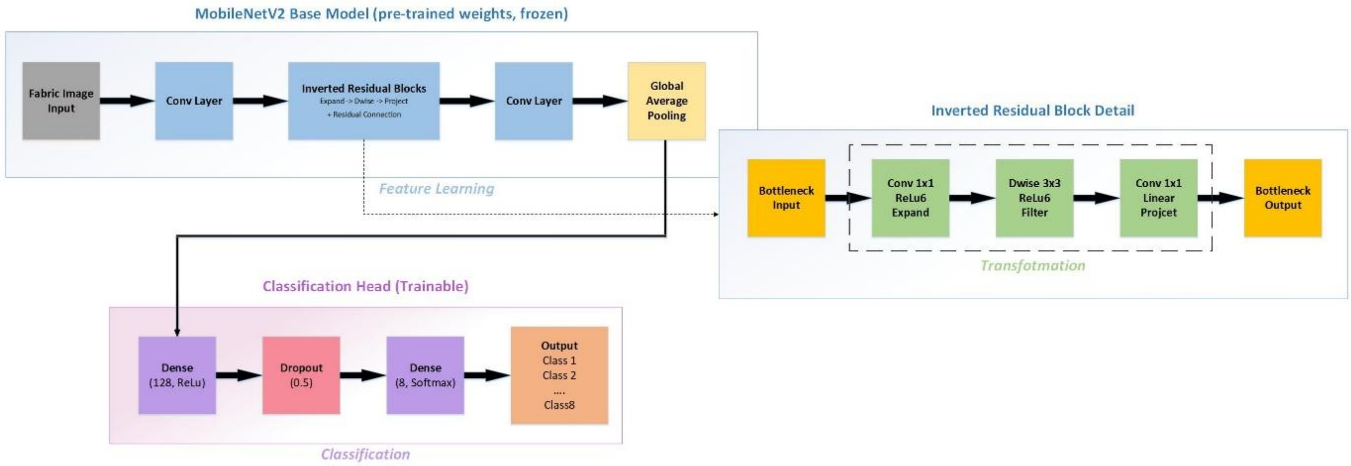


Figure 1. MobileNetV2 Architecture for Puta Dino Motif Recognition

The learning functionality of the feature processing section utilizes a MobileNetV2 pre-trained base model, as shown in figure 1, which comprises image data of 224×224×3 pixels. The base model is also characterised by the presence of an initial layer and an ensemble of inverted residuals. The blocks feature depth-wise convolutions arranged in 1x1-conv residuals, with ReLU6 activations attached, and include residual connections where applicable. To capture the idea of transfer learning, the base model's all layers are considered 'frozen' with pre-trained weights on the ImageNet dataset for feature extraction.

Specific to motif classification, trainable layers are included in the classification head. In the base model, Global Average Pooling is applied, where the spatial dimensions are scaled down, but significant features are retained. The classification layers include: (1) Dropout layer with 0.5 probability to prevent overfitting, (2) Dense layer with 128 neurons and ReLU activation for feature mapping, and (3) final Dense layer with eight neurons and softmax activation corresponding to the eight Puta Dino motif classes. The model is compiled using the Adam optimizer with a categorical cross-entropy loss function for multi-class classification.

In this study, all convolutional layers of the MobileNetV2 backbone were frozen to ensure stable feature extraction during training. This decision was based on the relatively limited size of the original dataset (810 images before augmentation; see Section 3.1), which poses a substantial risk of overfitting if deep layers are fine-tuned on small, domain-specific data. Prior literature suggests that fine-tuning high-capacity convolutional networks with insufficient data often results in rapid divergence between training and validation accuracy [16]. Moreover, the early and middle layers of MobileNetV2 encode low-level and mid-level features, such as edges, contours, textures, and repetitive patterns, which are highly transferable to woven fabric motif recognition tasks [17]. Keeping the ImageNet pretrained weights on these layers saves the layers from feature distortion, which they would likely suffer if fine-tuned with insufficient samples and decreased representation levels. Keeping the backbone frozen also saves on computational costs. It allows the training to be completed in a much shorter time, which in turn enables the model to be implemented in a web-based system that requires real-time inferences over an extended period. The performance, achieving a high level of 80% validation accuracy across all folds and on a separate test dataset with an accuracy of 98.05% (table 4), validates the frozen-backbone system as being computationally efficient and generalizable for Puta Dino motif classification. Other research may seek to optimize adjustments to deeper layers, while additional and larger sets of datasets include motifs.

3.3. Training Configuration

The training procedure employed was designed to apply the Adam optimizer and loss function set to categorical cross-entropy. To this effect, the model was trained to the 50th epoch while maintaining a batch size of 32. To prevent overfitting, early stopping with a patience of 5, and validation accuracy model checkpointing, were executed.

All the images served as input were resized to 224x224 pixels and normalized to the range of values [0, 1]. For the training, a 5-fold cross-validation methodology was employed, with the base model being MobilenetV2, which was pre-trained on ImageNet, and where the convolutional layers were frozen. The personal laptop experiments were performed using a dedicated GPU and TensorFlow and Keras, running within the Visual Studio Code environment.

3.4. Evaluation Metrics

To determine how well our model performed, the study compared the predicted labels to the actual ones, using a multi-class confusion matrix. A confusion matrix generates several metrics such as accuracy, precision, recall, and F1-score [18], [19], [20]. Here, accuracy refers to the total number of correct predictions. Precision refers to the proportion of predicted positives that actually belong to the class and are true positives. This translates into determining how many false positives the model predicts. Recall, also known as sensitivity, refers to the number of true positives among the actual positives. Measuring this proportion demonstrates the model's completeness in detecting positive cases. The F1 score is the most widely used metric that combines precision and recall, as it is the average of the two, and it is especially useful when the dataset is highly imbalanced.

These metrics are from the best-performing model, based on the training results of the 5-fold cross-validation, which achieved the maximum validation accuracy. For this evaluation, a different test set not used for training or validation was used. The confusion matrix enabled a qualitative assessment of the similarity between classes, allowing for the detection of patterns of repeated misclassification. The confusion of two motifs could indicate that the model was learning and generalizing the same visual features for those classes [21].

4. Results and Discussion

This section summarizes the outcomes of the training, validation, and testing of the proposed model, as well as model performance analytics across several folds, out-of-sample testing, web deployment, and prediction confidence and class proximity. This also helps in determining the predictive capabilities and functionality of the proposed system in both ideal and practical conditions. To ensure a thorough evaluation of performance, the CNN model was trained through 5-fold cross-validation on augmented datasets. In figure 2, the training and validation performance over the folds are shown, indicating the progression of accuracy (left) and the reduction of loss (right) during the training epochs.

From figure 2, the learning curves reflect the performance of the folds, and all folds show significant improvement over the training epochs. Notably, the validation accuracy is nearly equal to the training accuracy, with both decreasing as the training and validation losses decrease. This indicates no overfitting with effective advancement of learning in all folds. This summary of the training and validation performance is illustrated in table 2.

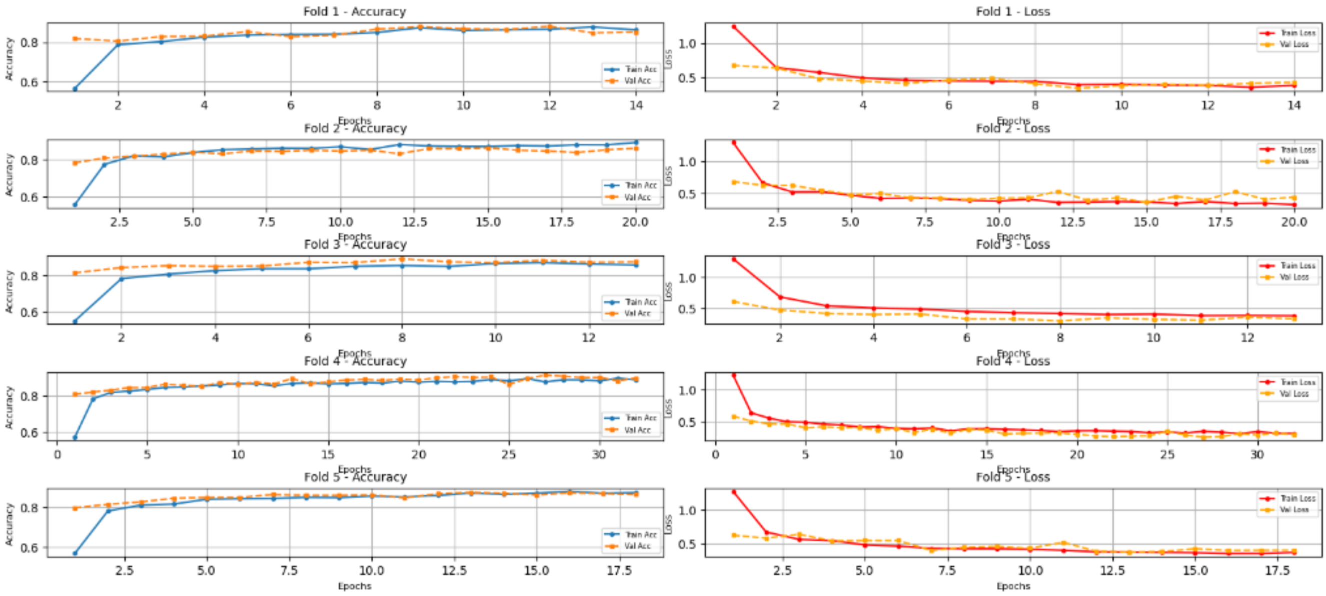


Figure 2. Training and Validation Performance Across 5-Fold Cross-Validation

The results depicted in table 1 show Fold 4 as the best in terms of accuracy, with 91.45%. The model's consistent performance across the folds (validation accuracy: 86.04%-91.45%) indicates the reliability of the proposed CNN architecture. The model will be the best-performing one and will be stored for later testing, with the aim of simulating real-life usage to analyze its performance beyond the training scope. This method conducts an in-depth analysis of the model in accordance with the suggestion of Ying [22], as deep learning model evaluations should extend beyond common validation attempts to assess performance across multiple scenarios indicative of production use. The model's ability to distinguish among the eight motif classes in the independent test is illustrated in detail in the class-wise prediction distribution in figure 3.

Table 2. Training and Validation Performance Summary

Fold	Max Training Accuracy	Max Validation Accuracy	Min Training Loss	Min Validation Loss
1	87.76%	88.12%	0.3524	0.3371
2	88.85%	86.04%	0.3275	0.3632
3	86.87%	88.95%	0.3889	0.3075
4	89.73%	91.45%	0.3065	0.2485
5	88.02%	87.50%	0.3528	0.3722

Figure 3 presents the confusion matrix for Fold 4, which illustrates the model's classification performance across the eight Puta Dino motif classes. The model's classification results are indicated by the heavily diagonal configuration of the matrix, reflecting the model's correct predictions for the majority of samples in every class. A significant number of classes exhibit 100% correct classification, where all 20 samples in the class are predicted correctly. The only two courses with a small amount of misclassification include one Dodara sample, which was expected as Marasante, and two Tobaru samples, which were also predicted as Marasante. The sparse off-diagonal in the confusion matrix reflects the low degree of visual confusion within these motifs, which aligns with the principle of fine-grained classification of woven patterns.

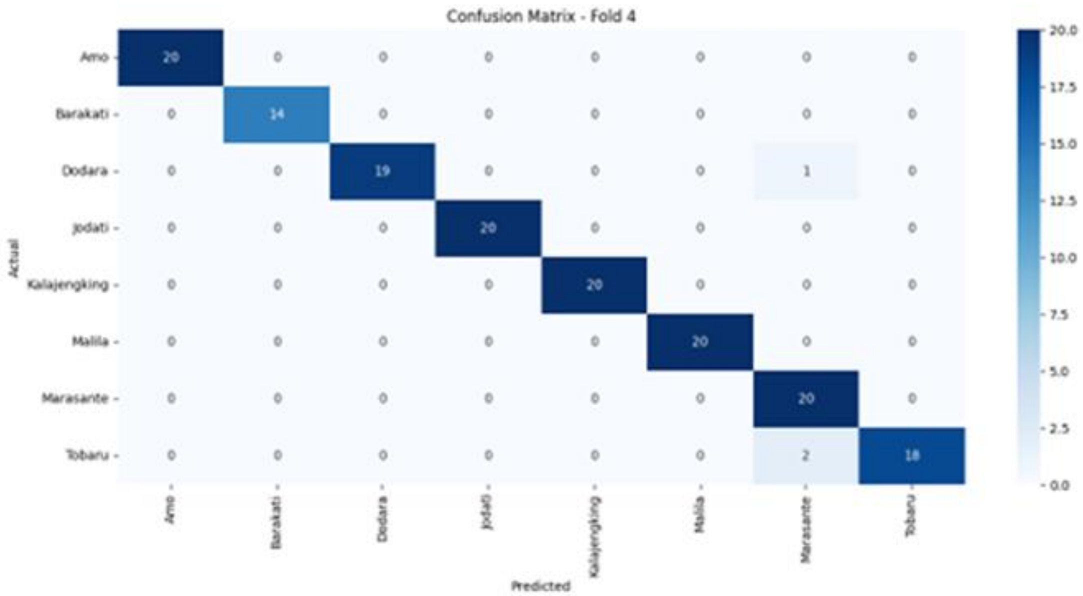


Figure 3. Confusion Matrix for Test Set Evaluation

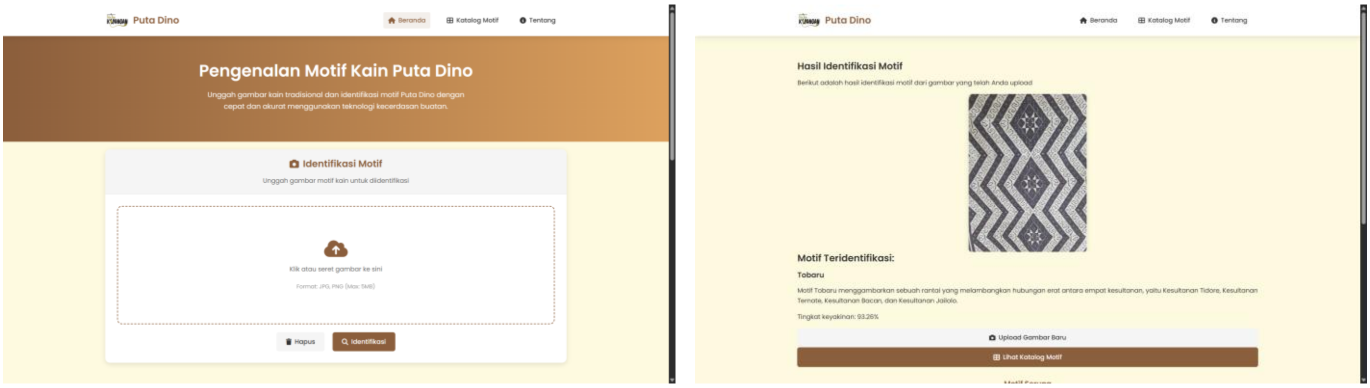
The model's predictive ability is further evidenced by the large number of samples in the diagonal of the confusion matrix, which demonstrates the model's ability to capture both mid- and high-level hierarchical features of the patterns. This validates the use of MobileNetV2 as a suitable representation for classifying the motifs. The motifs with clearly defined geometrical structures, such as Jodati, Kalajengking, and Malila, also show 100% classification accuracy. In contrast, motifs with high confusion due to subtle differences in structural patterns and textural details still show an acceptable degree of classification accuracy.

In addition, the consistent performance demonstrated in this fold corroborated the results of the other folds during the five-fold cross-validation, where all folds attained over 80% accuracy. This pattern suggests that the data splitting procedure effectively preserved the proportion of classes, acting as a successful stratified cross-validation. With every class being sufficiently represented, the cross-validation procedure effectively limited bias and provided a valid measure of the model's actual performance. Table 3 presents the overall evaluation metrics.

Table 3. Test Set Performance Metrics

Metric	Value
Accuracy	98.05%
Average Precision	98.37%
Average Recall	98.12%
Average F1-Score	98.12%

From table 3, while all other classes had F1-scores over 0.90, 5 motif classes (Amo, Barakati, Jodati, Kalajengking, and Malila) achieved perfect classification (precision and recall 1.0). With a test accuracy of 98.05%, it demonstrates a strong generalization ability. Based on the experiments conducted, the best-performing model achieved 98% accuracy. Hence, to streamline the process of Puta Dino woven fabric motif identification, a model implementation into a web-based system was developed. The purpose of the site is to assist with identifying Puta Dino woven fabric motifs. Remarkable was the implementation of the model into the framework, culminating in the functional web-based system depicted in figure 4. Users can upload images of fabric on the home page interface (figure 4(a)), and the next page, which displays the identified motif name, cultural significance, and level of confidence, provides prediction results (figure 4(b)).



(a) (b)

Figure 4. Web-Based System Interface: (a) Homepage Image Upload, (b) Prediction Results Display

In assessing the internet-based system, complete shots of five real-world scenarios involving image capture were painstakingly recorded and analyzed. In the first of the five conditions, the distant situation allowed complete observation of the entire fabric encompassing small, detailed motifs. In the moderate close position, the image capture location was relatively speaking an intermediate distance away from the fabric, allowing the motif to be interpreted in detail close enough to be distinguished. Close-up conditions involve shooting an image of the fabric from a position of extremely close range with very significantly dominated patterns. Angled conditions employed a non-perpendicular position in relation to the camera. Lastly, the Folded condition assessed the fabric for unevenness, introducing certain occlusions in the abstract patterns on the fabric. The system's performance in each of the five conditions was assessed, and the results are tabulated in table 4. Recognition results and confidence levels for all eight motifs across five real-world image capture conditions were analyzed, and the results were summarized.

Table 4. Web-Based System Performance Results

No.	Class	Recognition Performance (Confidence Level)					Max Confidence
		Distant	Mod-Close	Close-Up	Angled	Folded	
1	Amo	R (95.48%)	R (84.22%)	NR (60.22%)	R (75.41%)	R (64.86%)	95.48%
2	Barakati	R (99.26%)	R (66.45%)	NR (55.28%)	R (99.62%)	R (95.20%)	99.62%
3	Dodara	R (99.57%)	R (99.22%)	R (97.69%)	R (99.66%)	R (98.43%)	99.66%
4	Jodati	R (75.16%)	R (98.46%)	R (69.90%)	R (83.40%)	R (97.21%)	98.46%
5	Kalajengking	R (82.24%)	R (98.86%)	R (93.72%)	R (100.0%)	NR (95.07%)	100.0%
6	Malila	R (98.35%)	R (99.91%)	R (99.92%)	R (100.0%)	R (99.56%)	100.0%
7	Marasante	R (86.35%)	R (93.72%)	R (99.29%)	R (99.89%)	R (94.15%)	99.89%
8	Tobaru	R (86.17%)	R (93.26%)	R (98.44%)	R (97.70%)	R (84.22%)	98.44%

Note: R = Recognized, NR = Not Recognized

Based on table 4, the system exhibited considerable robustness, with the angled condition yielding the best results (100% recognition rate, 94.46% average confidence). That is, the system can efficiently manage non-perpendicular orientations. Both far and midrange were equally strong in recognition, as all motifs in these ranges were correctly recognized. On the contrary, the closed and folded conditions were the most problematic, as the closed condition had two motifs returned as 'Not Recognized' (NR), and one motif in the folded condition also returned NR. The difficulties stem primarily from incomplete motif capture, loss of global structure, and the geometric alterations typical of fabric manipulations.

The NR results may also be explained by the motif's visual properties in combination with the image capture conditions. The Kalajengking (Scorpion) motif, for instance, relies on the complete head-tail structure, as it becomes

ISSN 2723-6471

unrecognizable when folded due to occlusion and geometric deformation, which obscure the necessary features for coherent global and mid-level patterns to be extracted within MobileNetV2. Thus, the model is unable to associate incomplete representations with the Kalajengking class.

The same explanation applies to the NR examples for the Amo and Barakati motifs for the close-up condition. The repeated linear pattern of the Amo motif collapses into visually similar pieces when photographed extremely closely, and the model loses the cross-sectional global structural context necessary for classifying the pattern. The same phenomenon occurs with the Barakati motif, whose identifiable, repetitive square pattern is difficult to discern until a close-up photo is taken, when it is reduced to a pattern of fragmented, small patches that do not conform to the square pattern learned with the Barakati motif during the model's training. The square pattern of Barakati, together with the Amo linear pattern, explains the NR classification.

Aside from the NR cases, table 3 also provides details of motifs that have specific features that enhance the model's performance within its limitations. The case of the Dodara motif is particularly notable, as the model performed well on it across the board (97.69-99.66% confidence). The Dodara motif's distinctive combination of geometric arrangements and radial symmetry appears to provide strong, stable features recognizable by the model, even with variations in distance, angle, or minor textural changes, indicating that Dodara's symmetrical, even structure is closely aligned with the model's learned features to a high degree.

On the other hand, the confidence levels of the Tobaru motif are more variable, at 84.22% when folded under and 98.44% when observed closely at an angle. This shows mid-range features of Tobaru are discernible when clearly imaged, but recognizability suffers under fabric deformation. Geometric strokes of moderate repetition in Tobaru's motif elements can also be found in other motifs and are especially noticeable when folded or distorted, which accounts for the slight decrease in confidence. These findings demonstrate the extent to which the complexity of the motif and the consistency of its geometry influence recognizability in the different states of real-world capture, which includes fabric deformation.

The conclusions drawn from the data presented in table 4 are consistent with the well-known 'lab-to-real gap' in the literature, where training images become controlled and fail to represent the diversity found in more practical application situations. MobileNetV2 exhibits impressive robustness in many situations; however, motifs with extreme occlusion, severe over-cropping, or geometric disruptions consistently demonstrate lower classification reliability. Regardless, the software demonstrates strong overall performance, with solid capabilities on well-defined motifs (Dodara) and minor performance weaknesses on motifs with global shape distortion (Amo, Barakati, Scorpion, Tobaru in some circumstances). To help clarify these points and demonstrate how the model's behavior can extend beyond controlled assessments, the next section will review the system's performance in greater detail in the field.

The purpose of this section is to analyze and interpret the model's performance attributes observed during the comprehensive evaluation of the web-based system. The analysis addresses specific performance trends across varying capture conditions, confidence and similarity calculations, the mechanisms involved, and the practical implications of the observed phenomena in terms of the analysis target, namely, real system deployment. From this analysis, the conclusions that follow aim to assess system performance under various conditions and serve as indicators of model robustness and boundaries, completing this overview.

The evaluation results confirmed the angled condition as the best performing condition, with the removal of the non-liable watching angle as the most effective condition, achieving a 100% recognition rate with an average confidence of 94.46%. All motifs were recognized in both distant and moderately close conditions. In contrast, the close-up and folded conditions were more challenging.

As Liu et al.'s methods described in [23], the recognised versus total test case ratio, 37 correct recognitions out of 40 test cases in total for all conditions, was recognised as the overall system accuracy, with the overall system accuracy yielding 91.5% errors outperforming the 90% threshold as acceptable for the practical recognisation of the pattern. The accuracy rate is sufficient to conclude that the system is suitable for deployment in the preservation of cultural heritage, targeting heritage design scenarios [24], [25]. The purpose of performance was justified by the recognition system's practical application, which was the focus of a separate ablation study to assess the impact of data augmentation and classifier architecture configuration on the model's overall generalization.

Ablation studies were conducted to investigate the impact of various data augmentation configurations and classifiers on the models' performance. The comparisons considered (1) training models on the original (810 images) and on the augmented dataset (2400 images), and (2) the varying numbers of neurons present in the dense layer of the classifier head. Among the comparisons, models trained on the non-augmented dataset achieved an accuracy of 83.45% in validation as opposed to an accuracy of 89.75% in validation on the augmented dataset. This represents a statistically significant 6% increase in accuracy, which can be attributed to augmentation strategies—specifically, brightness level randomization, hue saturation modifications, and Gaussian noise—enhancing the models' generalization and robustness on varying illumination and textured surfaces.

The second comparison examined the effect of model capacity on performance by varying the dense layer configuration to 64, 128, and 256 neurons. These choices were ½ and 2 times the baseline values, which provided the minimal and maximum learning capacities on 128 neurons. The results in table 4 show that the configuration of 128 neurons exceeds the observable performance of all other configurations, as indicated by a validation accuracy of 91.45%. Moreover, the configurations of 64 neurons and 256 neurons, respectively, exhibit evident signs of underfitting and overfitting, resulting in increased computational time. Ablation study results for data augmentation and classifier configuration are presented in table 5.

Table 5. Ablation Study Results for Data Augmentation and Classifier Configuration

Configuration	Training Accuracy	Validation Accuracy	Observation
Without Augmentation	89.12%	83.45%	Overfitting observed
With Augmentation	93.45%	89.75%	Better generalization
Dense 64 neurons	88.50%	87.80%	Underfitting
Dense 128 neurons	90.95%	91.45%	Best performance
Dense 256 neurons	92.70%	90.62%	Slight overfitting

The data in table 5 also suggests that both data augmentation and classifier design influence model generalization. As determined, the configuration with augmented data and a 128-neuron dense layer achieves the best trade-offs in terms of accuracy, computation, and suitability for web applications. To delve deeper and explain the choice of model and data augmentation for our final analysis, we examine the impact of data augmentation on model robustness and validation performance in table 6.

Table 6. Analytical Effect of Each Augmentation Technique on Model Robustness and Performance

Augmentation Technique	Transformation Description	Intended Purpose	Expected Contribution to Robustness	Analytical Effect on Model Performance
Random Brightness Contrast	Randomly modifies brightness and contrast levels within a controlled range	Reduces sensitivity to uneven illumination and contrast differences	High improves resilience to underexposed/overexposed field conditions	Helps stabilize accuracy across folds; reduces misclassification caused by lighting variations
Hue Saturation Value (HSV) Shift	Adjusts hue (±10), saturation (±20), and brightness (±10)	Increases variation in color distribution to mimic environmental color shifts	Medium to high, improves generalization under diverse lighting/color profiles	Improves recall for classes whose colors are similar; reduces overfitting to specific color tones present in training data
RGB Shift	Shifts R, G, B channels (±15 each channel)	Creates color perturbations to prevent	Medium, encourages more robust feature extraction by reducing color-channel bias.	Reduces FP for classes differentiated by texture rather than color;

		dependence on specific color channels		stabilizes validation accuracy
Gaussian Noise	Adds random noise with intensity 10.0–50.0	Forces the model to learn discriminative features despite image quality degradation	Medium increases noise tolerance and resilience to low-quality image capture	Reduces overfitting; slightly increases FN at high noise intensity but improves generalization overall
Resizing (224×224)	Scales all images to MobileNetV2-compatible resolution	Standardizes input size for feature consistency	Indirect but essential, ensures uniform feature extraction	Improves training convergence; prevents performance drops due to inconsistent resolutions
Pixel Normalization (0–1 scaling)	Divides pixel values by 255	Stabilizes gradient flow and accelerates convergence	Indirect but essential for stable training	Reduces training loss oscillation; improves accuracy and generalization consistency
Dataset Balancing via Augmentation to 300 Images per Class	Increases samples for each class until reaching an equal class size	Eliminates imbalance-driven bias during training	Very high, prevents majority-class domination	Significantly improves the F1-score and inter-class fairness, boosting overall accuracy to above 80% consistency across five folds.
5-Fold Stratified Split on Augmented Data	Maintains class proportion across folds (8:2 split)	Ensures fair evaluation and avoids fold-specific imbalance	Very high, stabilizes TP–FP–FN behavior	Produces stable accuracy (>80%) across all folds; confirms evaluation reliability

While augmentation was implemented as a unitary pipeline, table 6 gives a detailed explanatory framework for the intention and anticipated value of every augmentation strategy adopted. It is evident from this assessment that the specific transformation and augmentation provided the model with greater performance improvement on previously seen data by focusing on particular variability issues within the data, such as variations in light, color, and noise, thereby validating the consistent accuracy of over 80% across all five partitions. The application of a stratified 5-fold technique also guarantees that these gains are properly allocated among the different classes and are not merely the result of a data imbalance.

The specific softmax layer of the CNN used here provides the confidence values as a measure of the model's confidence in a given case [26]. For example, in the case of the Barakati motif classification, where a confidence of 99.26% was achieved, the model displayed a high confidence level. Heuristic calculations of similarity values exist that utilize the confusion matrix [27], as shown in (1).

$$Similarity = \min\,(60 + (count\ \times 15), 95) \tag{1}$$

On the other hand, the assistive functionality allows the system to increase its usefulness in real-world use case scenarios by providing users with alternative suggestions of motifs based on the misclassification of the motif, thereby assisting users.

The geometrically structured motifs (Dodara, Kalajengking, Malila) achieved superior, consistently above 95% recognition rates, as they have distinct patterns which allow for creating strong representations on features, as opposed to the Organic-inspired motifs (Amo, Barakati), which have high sensitivity to capture conditions, especially if the scenario is close-up, where patterns are fragmented. The variance between the accuracy in testing environments (98%) and actual web execution (92.5%) epitomizes the classic “lab-to-real gap.” This gap is prevalent in various fields, including computer vision. As some studies indicate, models lose between 12% and 15% of their accuracy when transitioning from curated laboratory datasets to real-world datasets [28]. This is one of the fundamental flaws in computer vision compared to real-world applications, which aligns exactly with the results we obtained [29], [30], [31] [32], [33], [34].

Given the current system's performance, the gap between 92.5% and 98% accuracy in laboratory testing highlights the need for designing more robust strategies. Several tactical solutions can be incorporated into the forthcoming work to reduce the current lab-to-real gap and enhance the system's consistency under varying conditions. Table 7 outlines an array of strategies alongside the Puta Dino motif preservation.

Table 7. Strategies to Reduce the Lab-to-Real Gap and Enhance Motif Preservation

Strategy	Description	Expected Benefit	Applicability for Puta Dino Motifs
Domain Adaptation (Unsupervised / Semi-Supervised)	Aligns feature distributions between training (lab) and deployment (real-world) domains using techniques such as adversarial learning or feature alignment.	Reduces sensitivity to lighting changes, camera differences, and fabric distortions.	Valuable for motifs photographed from varied distances and environments.
Synthetic Data Generation (GAN-based / Diffusion Models)	Generates additional training images simulating extreme lighting, folds, angles, or low resolution using GANs or diffusion models.	Expands training diversity and reduces overfitting to controlled environments.	Helpful in capturing rare motifs and real-world conditions that are difficult to replicate.
Style Transfer Augmentation	Applies texture-preserving transformations that mimic different camera sensors, environments, or cloth textures.	Helps model generalize to unseen visual styles while retaining motif structure.	Addresses variations in smartphones used by visitors or community members.
Curriculum Learning	Gradually trains the model, starting from clean data and progressing to heavily augmented or degraded data.	Stabilizes learning and improves robustness against extreme distortions.	Effective for motifs with subtle differences that become harder to recognize when degraded.
Test-Time Augmentation (TTA)	Performs prediction on several augmented versions of the same input image and averages the results.	Improves stability during inference and reduces misclassification under noisy conditions.	Beneficial for web uploads with inconsistent lighting or partial occlusion.
Active Learning from User Feedback	The system periodically re-trains using misclassified examples submitted by users.	Enables continuous improvement and reduces real-world error rates.	Ensures long-term preservation as new motif variations or weaving techniques emerge.
Metadata-Aware Training	Incorporates metadata (distance, angle, capture condition) during training.	Allows the model to specialize or adapt to specific acquisition contexts.	Supports prediction under conditions like folded cloth or very close-up views.

The strategies highlighted in table 7 offer tangible routes for narrowing the lab-to-real gap and strengthening system defense in the upcoming development iterations. Unsupervised and semi-supervised domain adaptation strategies can help models improve their internal feature representation alignment more effectively with real-world conditions by mitigating the effects of lighting and camera discrepancies. The generation of synthetic data through GANs and diffusion models can expand the variety and richness of training samples, as well as the generalizability of models for rare or difficult-to-photograph motifs. Style transfer data augmentation and curriculum learning further enhance stability by training on a variety of visual transformations per model. Test-time augmentation enhances the reliability of a model's predictive performance at inference, while active learning supplements a system with user feedback, allowing it to adapt over time. Finally, metadata-aware training can help adapt predictions to contextual variables such as distance or angle. These measures collectively have a real-world impact on vision systems and conserve Puta Dino motifs through computer vision.

5. Conclusion

This research successfully created a classification model based on MobileNetV2, achieving groundbreaking success with 98% accuracy on previously unseen test data and 92.5% accuracy in a practical web deployment scenario, enabling the identification of 8 motifs in authenticated Puta Dino woven textiles. This study highlighted the valuable and practical application of lightweight convolutional neural networks in textiles. Along with community engagement, the web system built to meet the challenges of dynamic real-world deployment proved culturally valuable to the community. Nonetheless, the difference noted in fully controlled testing and real-world testing arises from the complex ‘lab-to-real’ problem in computer vision. This challenge often arises in insufficiently professional conditions, such as extreme close-up images of folded textiles under poor lighting. The system can be strengthened through further development of domain adaptation, curriculum learning, and active learning, as well as rapid system synthesis, to provide more detailed coverage of the textile digitization initiative.

Without getting too technical, this study focuses on some of the core ethical and practical issues impacting the digitization of Macedonian Puka textile patterns. The digitization of Puta Dino patterns must be done with utmost care to avoid cultural insensitivity, respect local weaving practices, and uphold the principles of the concerned local community. Concurrently, the introduction to the actual world poses challenges concerning the scarcity of devices and limited internet access, especially in the remote and peripheral regions of Tidore. Employing offline-capable designs, lightweight edge deployment, and user-experience-centered design would be essential to address these obstacles and maximize the system’s net benefit. Ultimately, the study lays the groundwork for potential future techno-procedural endeavors in the area of pattern recognition, while emphasizing the importance of responsible engagement, contextual awareness, and practical implementability, which, from the outset, uphold the living tradition of weaving in Tidore.

6. Declarations

6.1. Author Contributions

Conceptualization: L.A.L., S.E.R., H.B.L., F.J.P., J.R.K., and E.I.S.; Methodology: E.I.S.; Software: L.A.L.; Validation: L.A.L., E.I.S., and S.E.R.; Formal Analysis: L.A.L., E.I.S., and S.E.R.; Investigation: L.A.L.; Resources: E.I.S.; Data Curation: E.I.S.; Writing Original Draft Preparation: L.A.L., E.I.S., and S.E.R.; Writing Review and Editing: E.I.S., L.A.L., and S.E.R.; Visualization: L.A.L. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E. Alba, M. Schreiber, and A. García, “Weaving Words for Textile Museums: The Development of the Linked SILKNOW Thesaurus,” *Heritage Science*, vol. 10, no. Jan., pp. 1–12, 2022, doi: 10.1186/s40494-022-00681-x.

[2] F. Zhao and L. Wang,"Glossary of Textile Terminology (Based on the Documents from Dunhuang and Turfan)," *Journal of the Royal Asiatic Society,* vol. 23, no. 2, pp. 349–387, 2013, doi: 10.1017/S1356186313000291.

[3] A. Melelli, M. Bianchini, and L. Frontoni, "Textile Heritage Motifs to Decorative Furniture Surfaces," *J. Cultural Heritage,* vol. 52, no. Nov., pp. 192–201, 2021, doi: 10.1016/j.culher.2021.10.006.

[4] M. A. Wijnhoven and A. Moskvin, "Digital Replication and Reconstruction of Cultural Heritage: A Case Study," *J. Cultural Heritage,* vol. 46, no. Apr., pp. 1–9, 2020, doi: 10.1016/j.culher.2020.04.010

[5] F. D. H. Er, E. Y. Bulgun, and E. Ö. Adanır, "Preservation of a Textile Culture Through a Digital Cultural Heritage," *International Journal of Science and Technology Society,* vol. 6, no. 2, pp. 25–32, 2018, doi: 10.11648/j.ijsts.20180602.11.

[6] T. A. P. Sidhi, B. Y. Dwiandiyanta, and F. K. S. Dewi, "Batik Motifs Detection Using Pattern Recognition Method," *Jurnal Buana Informatika*, vol. 11, no. 1, pp. 1–10, 2020.

[7] A. Oktarino, Y. D. Tasri, and A. Efendi, "Identification of Batik Motif Based Deep Learning-Convolutional Neural Network Approach," *Journal of Ocean, Mechanical and Aerospace – Science and Engineering,* vol. 68, no. 3, pp. 139–147, 2024.

[8] A. A. Kasim, M. Bakri, A. Hendra, and A. Septriani, "Spatial and Topology Feature Extraction on Batik Pattern Recognition: A Review," *J. Informatika,* vol. 2021, no. Jun., pp. 1–12, 2021.

[9] J. P. Anidya and H. D. Purnomo, "Design of Batik Motif Detection System Using Deep Learning Method," *International Journal of Information Technology and Business*, vol. 7, no. 2, pp. 9–14, 2025.

[10] E. Y. Puspaningrum, E. P. Mandyartha, and F. A. Akbar, "Batik Motif Recognition Using Machine Learning Method as Educational Media (Case Study: CV. Titik Batik)," *Nusantara Sci. Technol. Proc.,* vol. 2025, no. Jul., pp. 612–617, 2025.

[11] V. Ayumi, I. Nurhaida, and H. Noprisson, "Implementation of Convolutional Neural Networks for Batik Image Dataset," *International Journal of Computing Science and Applied Mathematics,* vol. 8, no. 1, pp. 5–8, Feb. 2022, doi: 10.12962/j24775401.v8i1.5053.

[12] M. A. Rasyidi and T. Bariyah, "Batik pattern recognition using convolutional neural network," *Bulletin of Electrical Engineering and Informatics,* vol. 9, no. 4, pp. 1430–1437, Aug. 2020, doi: 10.11591/eei.v9i4.2385.

[13] A. Ariessaputra, et al., "Classification of Lombok Songket and Sasambo Batik Motifs Using Convolutional Neural Network (CNN)," *Journal of Information and Visualization (JOIV),* vol. 8, no. 1, pp. 1–8, 2024, doi: 10.30630/joiv.8.1.2101.

[14] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *IEEE Conf. Comput. Vis. Pattern Recognit.,* vol. 2018, no. Jun., pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[16] C. Wang, Z. Liu, J. Fei, Z. Long, P. Wang, and T. Yu, "Evaluating the generalizability and transferability of acoustic leak detection models for water distribution networks," *Water Res.,* vol. 286, no. Nov., pp. 1–12, 2025, doi: 10.1016/j.watres.2025.124273.

[17] A. Makandar et al., "Deep Learning–Based Dual-Pipeline Framework for Acute Lymphoblastic Leukemia Classification With Explainable AI Integration," *Applied Computational Intelligence and Soft Computing,* vol. 2025, no. Jan., pp. 1–12, 2025, doi: 10.1155/acis/8886728.

[18] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol, CA, USA: O'Reilly Media, Inc., 2019.

[19] N. E. Ramli, Z. R. Yahya, and N. A. Said, "Confusion Matrix as Performance Measure for Corner Detectors," *Journal of Advanced Research in Applied Sciences and Engineering Technology,* vol. 29, no. 1, pp. 256–265, 2022, doi: 10.37934/araset.29.1.256265.

[20] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.

[21] K. Malik and C. Robertson, "Landscape similarity analysis using texture encoded deep-learning features on unclassified remote sensing imagery," *Remote Sens (Basel),* vol. 13, no. 3, pp. 1–26, Jan. 2021, doi: 10.3390/rs13030492.

[22] X. Ying, "An Overview of Overfitting and its Solutions," *J. Phys. Conf. Ser.,* vol. 2019, no. Mar., pp. 1–10, 2019. doi: 10.1088/1742-6596/1168/2/022022.

[23] Z. Liu, C. Zhang, C. Li, S. Ding, Y. Dong, and Y. Huang, “Fabric defect recognition using optimized neural networks,” *J Eng Fiber Fabr,* vol. 14, no. Jan., pp. 1–12, 2019, doi: 10.1177/1558925019897396.

[24] S. S. Sarwar, A. Ankit, and K. Roy, “Incremental Learning in Deep Convolutional Neural Networks Using Partial Network Sharing,” *IEEE Trans. Emerg. Topics Comput. Intell.,* vol. 5, no. 3, pp. 420–432, Jun. 2021, doi: 10.1109/TETCI.2020.2992846.

[25] T. Boyadzhiev, G. Lagani, L. Ciampi, G. Amato, and K. Ivanova, "Comparison of different deep neural network models in the cultural heritage domain," *arXiv,* vol. 2025, no. Jan., pp. 1–12, 2025. [Online]. Available: https://falchi.isti.cnr.it/pisaDataset/

[26] T. Pearce, A. Brintrup, and J. Zhu, "Understanding softmax confidence and uncertainty," *arXiv,* vol. 2021, no. Jun., pp. 1–12, 2021. doi: 10.48550/arXiv.2106.04972. [Online]. Available: http://arxiv.org/abs/2106.04972

[27] R. Taori, A. Dave, V. Shankar, N. Carlini, B. Recht, and L. Schmidt, "Measuring robustness to natural distribution shifts in image classification," *NeurIPS,* vol. 2020, no. Dec., pp. 13922–13933, 2020. Available: https://modestyachts.github.io/imagenet-testbed/

[28] M. A. I. Hussain, B. Khan, Z. Wang, and S. Ding, “Woven fabric pattern recognition and classification based on deep convolutional neural networks,” *Electronics (Switzerland),* vol. 9, no. 6, pp. 1–12, Jun. 2020, doi: 10.3390/electronics9061048.

[29] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *ICLR,* vol. 2019, no. May, pp. 1–12, 2019. [Online]. Available: https://github.com/hendrycks/robustness

[30] H. Tabkhi, "Real-world computer vision for real-world applications: Challenges and directions," in Lecture Notes in Networks and Systems, Cham, Switzerland: Springer, 2022, pp. 627–634. doi: 10.1007/978-3-031-16072-1_53.

[31] H. Schieber, K. C. Demir, C. Kleinbeck, S. H. Yang, and D. Roth, "Indoor synthetic data generation: A systematic review," *Comput. Vis. Image Underst.,* vol. 240, no. Mar., pp. 103907–103907, 2024. doi: 10.1016/j.cviu.2023.103907.

[32] W. Li, Y. Guo, J. Zheng, H. Lin, C. Ma, L. Fang, and X. Yang, "Bridging the gap between object detection in close-up and high-resolution wide shots," *Comput. Vis. Image Underst.,* vol. 249, no. Dec., pp. 104181–104181, 2024. doi: 10.1016/j.cviu.2024.104181.

[33] R. Liu, C. Yang, W. Sun, X. Wang, and H. Li, "StereoGAN: Bridging synthetic-to-real domain gap by joint optimization of domain translation and stereo matching," *IEEE Conf. Comput. Vis. Pattern Recognit.,* vol. 2020, no. Jun., pp. 12757–12766, 2020.

[34] Z. Wang, W. Tao, Z. Zhang, and H. Zhao, “Towards bridging the synthetic-to-real gap in quantitative photoacoustic tomography via unsupervised domain adaptation,” *Photoacoustics,* vol. 2025, no. Jan., pp. 100736–100736, 2025, doi:10.1016/j.pacs.2025.100736.