

Automated Video Recognition of Traditional Indonesian Dance Using Hyperparameter-Tuned Convolutional Neural Network

Santi Purwaningrum^{*1}, Agus Susanto², Hera Susanti³, Mohammed Ayad Alkhafaji⁴

^{1,2}Multimedia Engineering Technology, Politeknik Negeri Cilacap, Indonesia

³Electronics Engineering, Politeknik Negeri Cilacap, Indonesia

⁴College of Technical Engineering, National University of Science and Technology, Dhi Qar, Iraq

Email: ¹santi.purwaningrum@pnc.ac.id

Received : Jul 22, 2025; Revised : Aug 25, 2025; Accepted : Aug 27, 2025; Published : Oct 16, 2025

Abstract

Traditional Indonesian dances serve as a vital expression of cultural identity and regional heritage, yet their preservation through intelligent video recognition remains limited due to technical challenges in motion complexity, costume variation, and the lack of annotated datasets. Prior research commonly employed Convolutional Neural Networks (CNNs) with manually defined hyperparameters, which often resulted in overfitting and poor adaptability when applied to dynamic and real-world video inputs. To overcome these limitations, this study proposes a robust and adaptive classification framework utilizing a hyperparameter-tuned CNN model. The approach automatically optimizes key training parameters such as learning rate, batch size, optimizer type, and epoch count through iterative experimentation, thereby maximizing the model's ability to generalize across both static and temporal data domains. The model was trained using image datasets representing three traditional dances (Gambyong, Remo, and Topeng), and subsequently tested on segmented frames extracted from YouTube videos. Results indicate strong model performance, achieving 99.67% accuracy on the training set and 100% accuracy, precision, recall, and F1-score across all testing videos. The proposed method successfully bridges the gap between still-image learning and real-world motion recognition, making it suitable for practical applications in digital archiving and cultural documentation. This study's contribution lies not only in the model's technical effectiveness but also in its support for preserving intangible cultural assets through intelligent and automated video-based recognition. Future work may incorporate temporal modelling or multi-camera perspectives to further enrich motion understanding and extend the system to broader performance domains.

Keywords : *Automated Video Recognition, Convolutional Neural Network, Hyperparameter Tuning, Indonesian Traditional Dance, Video Recognition.*

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. INTRODUCTION

Traditional dances are an integral part of the vast cultural heritage of Indonesia, capturing historical chronicles, ethnic identity, and regional diversity [1], [2], [3]. Against the backdrop of the growing global interest in cultural preservation by using digital technologies, there is a strong need for intelligent systems to automatically identify and classify traditional dance movements using video data analysis [4], [5]. The use of automatic video recognition using deep learning approaches, specifically Convolutional Neural Networks (CNNs), has shown strong potential in the field of recognizing human actions and gestures [6], [7], [8]. However, applying similar techniques to traditional dances, especially those with origins in Indonesia, is a relatively unexplored field [9], [10]. It is partly because of the restricted availability of annotated datasets and the fact that dance movements have a high level of variation regarding regional style, costumes, and rhythmic patterns.

Previous works by [3], [11], [12] attempted to identify dance have often relied on manually tuned hyperparameters during model training. While such approaches have achieved preliminary success, they

tended to lead to overfitting problems, poor generalizability, or subpar classification performance when tested against actual video inputs, e.g., those on YouTube [3], [12]. In addition, most of these works focused solely on image-based training and ignored the temporal dynamics involved in video data [11]. To overcome such limitations, the current study presents an automatic dance recognition system for traditional Indonesian dance, which relies on a hyperparameter-optimized convolutional neural network (CNN) model. The model is trained on image datasets and then tested on chosen clips from online videos, with the hope of boosting model robustness, accuracy, and generalizability to unseen video content through the use of hyperparameter optimization, while at the same time promoting cultural conservation through smart visual recognition [13], [14].

Research by [11] focused on developing a hybrid architecture called TransCNN-DSSS, which combines Transformer-based dynamic streams and CNN-based static streams to evaluate the quality of dance movements. The model integrates an attention mechanism to enhance its ability to analyze motion features such as fluency, rhythm, and expressiveness, achieving high performance with 90% accuracy, 89% recall, and an F1-score of 0.90. This approach effectively captures both spatial and temporal aspects of dance sequences, making it suitable for evaluating professional performance. However, a notable limitation of this work lies in its reliance on fixed optimization settings, as it does not incorporate hyperparameter tuning techniques. This could restrict the model's adaptability to diverse video contexts and reduce its robustness when generalized to broader datasets or real-world footage.

Research by [3] proposed a CNN-based classification system using frame-level segmentation to recognize traditional Indonesian dances such as Gambyong, Remo, and Topeng. The model was designed to extract detailed spatial features, including body posture, foot movement, and facial expressions, leading to a reported accuracy of 97.5%. This work highlights the importance of culturally specific models for dance recognition and contributes significantly to the digital preservation of regional heritage. Nonetheless, the training process in this study employed a standard ADAM optimizer with manually defined hyperparameters, which may have limited the potential to reach higher performance stability or prevent overfitting. Without systematic hyperparameter optimization, the model's generalizability to different dance styles or video conditions remains uncertain.

Research by [12] presented a dance motion quality evaluation system based on a ST-CNN with 3D feature extraction and LSTM aggregation. The model was validated on NTU-RGB60 and NTU-RGB120 datasets and achieved its highest accuracy of 90.8%, demonstrating solid performance in capturing both the temporal flow and contextual dependencies within dance videos. In addition, the model's proposed integration with IoT-based platforms positions it for real-time intelligent dance assessment. Despite these advantages, the approach relied on static hyperparameter settings without any systematic tuning or optimization process. This lack of tuning potentially limits the model's adaptability to different types of dance videos, especially those with varying lighting conditions, background noise, or motion complexity. Consequently, the model may not generalize well outside its original test environment when compared to more flexible architectures that incorporate automated hyperparameter optimization.

Based on the three related research above, this study addresses several identified limitations by proposing a CNN-based recognition framework specifically tailored for traditional Indonesian dance, enhanced through systematic hyperparameter tuning. While previous models demonstrated promising accuracy, they relied on manually configured hyperparameters and often lacked evaluation on real-world, unconstrained video data. In contrast, this study introduces a more robust and generalizable approach by training the model on image datasets and testing it on temporal segments from publicly available YouTube videos. The integration of hyperparameter optimization techniques aims to improve model performance by reducing overfitting and enhancing adaptability to various video conditions. Through this, the proposed method not only strengthens the technical foundation of dance recognition

systems but also contributes to the broader goal of cultural preservation by enabling automated identification of traditional dance performances in online media.

Despite these promising results, prior studies still exhibit a critical limitation. Most models achieved relatively high accuracy but relied on manually configured hyperparameters, making them less adaptive to diverse real-world video inputs. In particular, they often failed to generalize well when applied to uncontrolled environments, such as YouTube videos, due to the absence of automated hyperparameter optimization. This research explicitly addresses that gap by proposing a hyperparameter-tuned CNN that enhances adaptability and robustness for video-based recognition of traditional Indonesian dances. The novelty of this approach lies in systematically optimizing training parameters to overcome overfitting and strengthen cultural preservation through intelligent video recognition.

2. METHOD

As seen in Figure 1, the proposed research framework is structured into three main stages: data preparation, initialization, and evaluation. In the data preparation stage, image and video sources are collected from publicly available datasets and YouTube videos featuring traditional Indonesian dances. The image-based data is designated for training the classification model, while the video-based data is segmented and converted into individual frames to simulate real-time testing conditions. This dual-source approach allows the model to be trained on well-defined visual features while being evaluated against dynamic, unconstrained sequences extracted from actual video performances.

The initialization phase involves the design and training of a CNN whose architecture and parameters are refined through hyperparameter tuning. The tuning process applies systematic optimization techniques to identify the most effective learning rate, batch size, number of epochs, and other critical parameters, resulting in a set of optimized hyperparameters that improve model performance and generalization. Once trained, the model outputs a classification database which is then assessed using a confusion matrix to quantify accuracy, precision, recall, and other performance metrics. The final stage applies the trained model to the video frames for automated pattern recognition, completing the evaluation process and demonstrating the system's ability to recognize traditional dance movements from real-world video content.

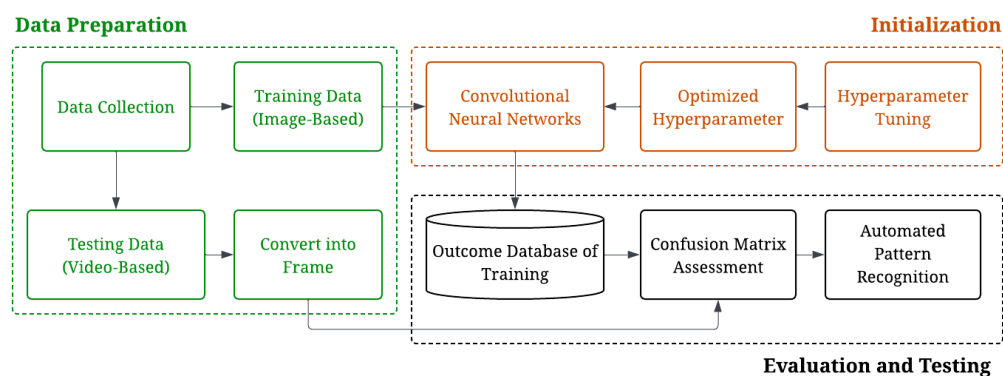


Figure 1. Proposed Scheme

2.1. Data Collections

The image dataset used in this study was sourced from a publicly available cultural repository and research datasets containing three classes of traditional Indonesian dances: Gambyong, Remo, and Topeng. Specifically, the dataset includes 4,293 Gambyong images, 4,502 Remo images, and 4,175 Topeng images. All images were resized to 227×227 pixels with three RGB channels and normalized to values between 0 and 1 to ensure input consistency. To improve model robustness and reduce

overfitting, standard data augmentation techniques were applied, including random horizontal flipping, small-angle rotations ($\pm 15^\circ$), brightness adjustments, and random zooming. These augmentations simulated natural variations in dancer orientation, lighting, and costume appearance. A visual example of the image dataset used for training is presented in Figure 2.

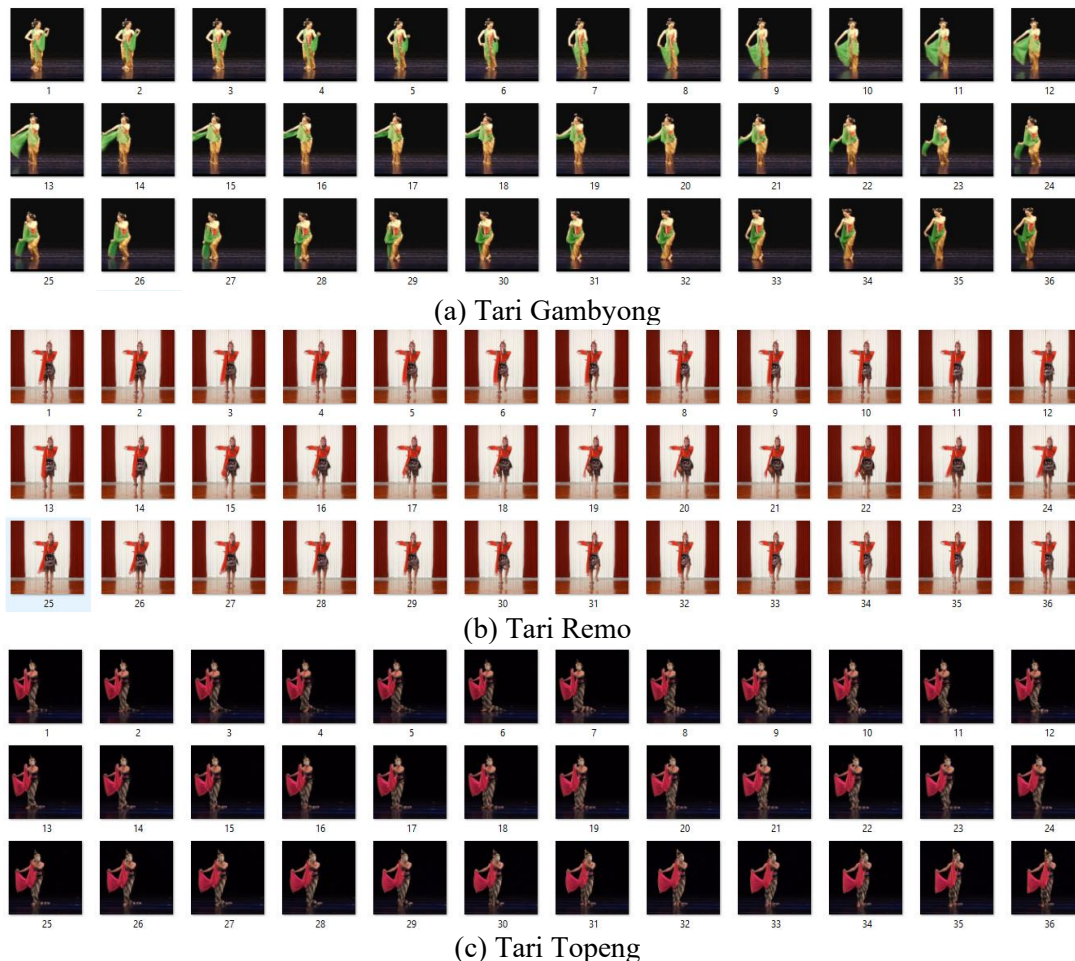


Figure 2. Sample of Image Datasets

For the testing dataset, traditional dance videos were collected from YouTube using the keywords “Tari Gambyong,” “Tari Remo,” and “Tari Topeng.” The selection criteria required videos to be in HD quality ($\geq 720p$), contain complete dance sequences, and have minimal background occlusion. From each selected video, a representative segment (40th–93rd second) was trimmed to capture the most expressive and consistent parts of the performance. The trimmed clips were then converted into frame sequences at a rate of 30 fps, producing several thousand samples per class. This systematic extraction ensured compatibility with the image-trained CNN while enabling evaluation on real-world, dynamic video inputs [15], [16], [17].

2.2. Hyperparameter Initialization

Hyperparameter configuration plays a critical role in the training efficiency and generalization capability of deep learning models. Previous studies on dance recognition, such as those by [11], [12], and [3], applied manually selected hyperparameters, typically using the Adam optimizer with fixed learning rates and predefined epoch settings. However, such manual approaches may not yield optimal performance, particularly when applied to dynamic and variable real-world data such as video frames. In contrast, this study adopts a hyperparameter tuning strategy to automatically search for the most

effective combinations of optimizer, learning rate, batch size, and epoch count [14], [18], [19]. In this study, the hyperparameter optimization was conducted using a Bayesian optimization approach, which adaptively balances exploration and exploitation during the search process. The optimizer types explored included Adam, SGD, and RMSprop. The learning rate was varied in the range of 0.00001 to 0.01, batch sizes were tested between 16 and 128, and the maximum epoch count was set between 10 and 100. Each candidate configuration was trained and evaluated using a 20% validation split from the training dataset. The final optimal configuration was determined based on the highest validation accuracy while also monitoring loss stability to avoid overfitting. This systematic tuning process ensured reproducibility and transparency while maximizing the robustness of the proposed CNN model. As seen in Table 1, prior studies consistently relied on fixed hyperparameter configurations without exploring alternative training settings, which may have limited model adaptability. In contrast, this research applies an automated tuning approach across multiple optimizer types, learning rates, batch sizes, and epoch configurations. This systematic strategy enables the proposed model to dynamically adapt its learning process based on the dataset characteristics, ultimately improving its robustness and predictive performance when evaluated on both static image data and frame-based video inputs.

Table 1. Hyperparameter Initialization

Study By	Optimizer	Max Epoch	Learning Rate	Batch Size	Tuning Status
[11]	Adam	40	0.0001	N/A	Manual
[12]	Adam	N/A	0.0001	N/A	Manual
[3]	Adam	16	0.0001	N/A	Manual
Our	Adam	Tuned	Tuned	Tuned	Automated (Tuned)

2.3. Convolutional Neural Networks

The architecture of the Convolutional Neural Network (CNN) used in this study is specifically designed to extract hierarchical visual features from traditional Indonesian dance frames and classify them into three categories: Gambyong, Remo, and Topeng. The input to the network consists of RGB images with a resolution of $227 \times 227 \times 3$, extracted from segmented video frames. These inputs are passed through a sequence of convolutional and pooling layers that progressively learn spatial patterns such as posture, costume structure, and movement contours relevant to each dance type. The initial layers of CNN, can be seen in Figure 3.

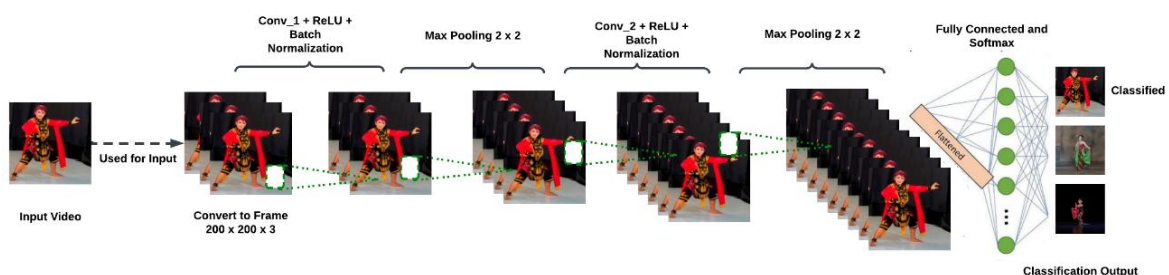


Figure 3. Initial Layers of CNN

The first convolutional block includes a 2D convolutional layer with a 3×3 filter and 8 output channels, followed by a batch normalization layer and a ReLU activation function to introduce non-linearity [20], [21]. This is followed by a 2×2 max pooling layer to reduce spatial dimensions. The second block expands the feature depth using 16 filters, again accompanied by batch normalization, ReLU, and a pooling layer. The third convolutional block consists of 32 filters, further deepening the model's capacity to learn complex visual features. After the feature extraction phase, the output is flattened and passed to a fully connected layer with three output neurons, corresponding to the three

dance classes. A softmax layer is then applied to generate class probabilities, followed by a classification layer that assigns the final label [22], [23].

2.4. Confusion Matrix Assessment

The performance of the CNN model was assessed based on the use of a confusion matrix that provides a class-wise analysis by recording the number of correct and incorrect predictions obtained for each of the classical dance classes: Gambyong, Remo, and Topeng. The matrix is a critical tool for computing various key performance measures such as precision, recall, F1-score, and overall accuracy [7], [24], [25]. These measures provide important information about the model's ability to distinguish between the three classes, especially in cases where visual similarities might lead to misclassification. By showing the distribution of predicted labels against actual labels, the confusion matrix helps in identifying whether the model is overconfident about some classes and underestimates other classes [26], [27]. With the complex and subtle nature of cultural dance data, such an analysis is essential in determining the reliability and generalization of the classification system for video-based usage. The calculation formulas for the classification performance metrics, including accuracy, precision, recall, and F1-score, are provided in Eq (1) - (4), where true positive (TP) is a situation where the model correctly classifies a dance frame into its respective class. For example, when a frame of Tari Remo is accurately predicted as Remo, it is counted as a true positive. On the other hand, a false positive (FP) is when a frame of another type of dance, Gambyong or Topeng, is mistakenly predicted as Remo. Additionally, a false negative (FN) comes about when a Tari Remo frame is incorrectly labeled as either Gambyong or Topeng. These situations show how the elements of the confusion matrix are used in multi-class dance recognition and how the evaluation metrics indicate the classification performance of the model in practical scenarios.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$F1 - score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (4)$$

3. RESULT

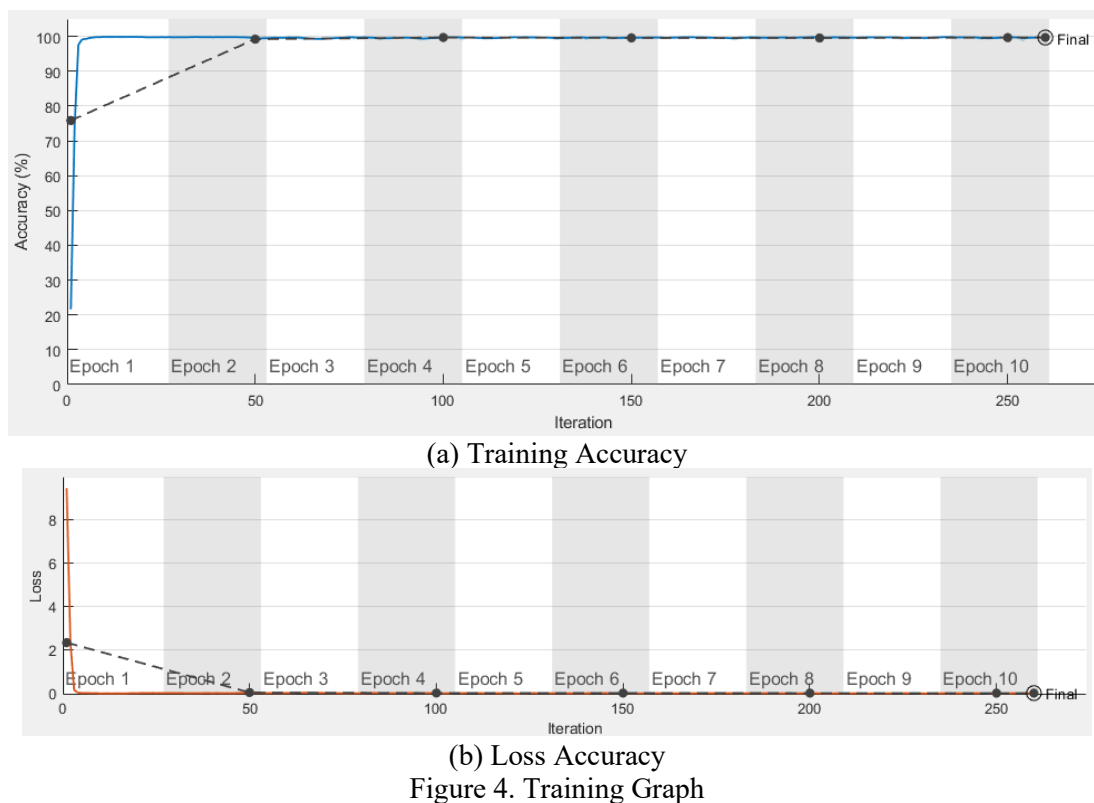
The training and evaluation processes in this study were conducted using MATLAB as the primary development environment, leveraging its Deep Learning Toolbox for model construction, training, and visualization. To support high-performance computation, the training was executed on a workstation equipped with an AMD Ryzen 7 7800X3D processor, an NVIDIA RTX 5070 GPU, 32 GB of RAM, and a 2 TB SSD. This hardware configuration was fully allocated to MATLAB, ensuring optimal GPU acceleration and memory bandwidth during convolutional training and inference tasks. The system was designed to handle large-scale image inputs and rapid iteration across multiple hyperparameter settings.

Before the training process, the prepared datasets were evaluated to ensure consistency and quality. All image data were successfully resized to a standardized resolution of 227×227 pixels with three RGB channels and normalized to values between 0 and 1, while augmentation such as horizontal flipping, small-angle rotation, and brightness adjustment effectively increased dataset diversity. The three classes of Gambyong, Remo, and Topeng were confirmed to remain balanced, reducing potential

bias during training. For the video dataset, the selected YouTube clips ($\geq 720p$ resolution) were trimmed between the 40th and 93rd second, and frame extraction at 30 fps produced several thousand frames per class with clear motion continuity and minimal background noise. The detailed outcomes of the training process, including accuracy progression and loss convergence, are presented in subsection 3.1. Meanwhile, the evaluation results using video-based frame classification are discussed in subsection 3.2.

3.1. Training Section

The results of the training phase in this study are visualized in Figure 4, which illustrates the model's performance over successive epochs. As shown in Figure 4(a), the training accuracy demonstrates a consistent upward trend, indicating that the convolutional neural network successfully learned the distinguishing features of each traditional dance class. Meanwhile, Figure 4(b) presents the training loss curve, which exhibits a gradual decline, reflecting the model's increasing ability to minimize classification errors during the learning process. These trends suggest that the training was stable and effective, with no signs of early stagnation or significant overfitting throughout the training iterations.



As shown in Figure 4, the training accuracy rapidly increased during the initial epochs, reaching above 95% by the end of the first epoch, and then stabilized at 99.67% across the remaining iterations. Meanwhile, the training loss sharply declined and approached near zero within only a few epochs, remaining stable afterward. This convergence behavior can be attributed to the use of automated hyperparameter optimization, which systematically selected the most effective learning rate, batch size, and optimizer configuration. These tuned parameters allowed the model to converge faster, avoid overfitting, and consistently capture discriminative features of Gambyong, Remo, and Topeng dances. Together, these results demonstrate that the proposed CNN architecture, when combined with optimized hyperparameters, is highly effective and well-generalized for the classification task. The detailed

performance metrics derived from the training phase, including precision, recall, and F1-score for each class, are presented in Table 2.

Table 2. Matrix Performance Each Input Video

Input Video	AVR Frame Accuracy	AVR Frame Precision	AVR Frame Recall	AVR Frame F1-Score
Tari Topeng [15]	100%	100%	100%	100%
Tari Gambyong [16]	100%	100%	100%	100%
Tari Remo [17]	100%	100%	100%	100%

As presented in Table 2, all evaluation metrics, including accuracy, precision, recall, and F1-score, reached 100% across the three dance classes. This perfect performance reflects the balanced dataset distribution, the effectiveness of CNN-based feature extraction, and the representativeness of the testing frames extracted from YouTube videos. The augmentation process during training also contributed to the model's ability to generalize effectively from static images to dynamic video frames. Nevertheless, it should be noted that the testing dataset consisted of a relatively small number of video clips, which may not fully capture the diversity of real-world performance conditions. Future evaluations with larger and more varied datasets will be necessary to further validate the robustness of the proposed approach.

3.2. Automated Pattern Recognition by Movement

After completing the training process using static image data, the trained model was evaluated on video datasets by converting each input video into a sequence of individual frames. Each frame was then classified independently to identify the corresponding traditional dance category. The model consistently recognized motion patterns based on body posture, costume flow, and dancer orientation, with a focus on key distinguishing elements such as arm gestures and cloth movement. The red-colored overlays indicate the dominant movement areas learned by the model for each frame, highlighting its capability to extract and respond to spatial cues relevant to dance identification. The classified patterns extracted from these video frames are visualized in Figure 5. The movement pattern for each frame is highlighted with a red bounding region, indicating the specific spatial areas that the model used to perform classification. These red frames reflect the dominant features captured across sequential motion, such as hand gestures and cloth dynamics, which are critical in distinguishing traditional dance types.

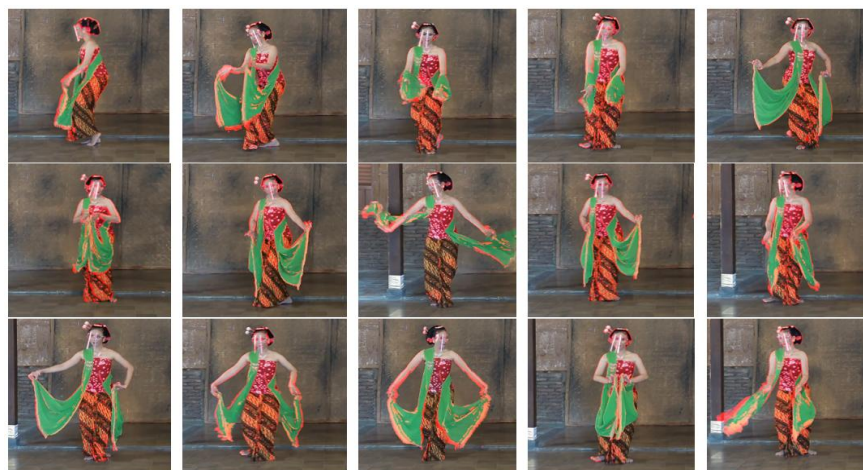


Figure 5. Classified Movement Patterns per Frame for Tari Gambyong

4. DISCUSSIONS

As seen in Table 3, the proposed model outperforms previous studies in terms of classification performance. While earlier works [11] and [12] achieved accuracy levels of 90% and 90.8% respectively without reporting detailed precision, recall, or F1-score metrics, and [3] reached an impressive 98.2% accuracy with perfect class-wise metrics, the model in this study achieved the highest overall accuracy of 99.6%. Furthermore, it obtained perfect scores of 100% for precision, recall, and F1-score, indicating the model's exceptional ability to correctly identify and distinguish all classes without misclassification. These results highlight the effectiveness of the hyperparameter tuning strategy and the robustness of the proposed CNN architecture in recognizing traditional Indonesian dances.

Table 3. Comparison with Other Study

Study By	Methods	Accuracy	Precision	Recall	F1-Score
[11]	CNN	90%	N/A	N/A	N/A
[12]	CNN	90.8%	N/A	N/A	N/A
[3]	CNN	98.2%	100%	100%	100%
Our	CNN	99.6%	100%	100%	100%

This performance gain can be attributed to multiple factors, one of the most significant being the use of automated hyperparameter tuning. While earlier studies typically used manually selected parameters such as fixed optimizers and learning rates, the proposed model dynamically explored combinations of optimizers (Adam, SGD, RMSprop), learning rates, batch sizes, and epoch settings. This allowed the network to achieve better convergence, avoid underfitting or overfitting, and adapt more effectively to the unique features present in traditional dance image and video data. The use of optimization not only improved classification precision, but also contributed to the model's robustness in handling real-world video frames with varying motion, lighting, and costume attributes.

The main contribution of this research lies in combining a tailored CNN architecture with hyperparameter optimization to improve recognition performance on cultural datasets, which are often complex and underrepresented in computer vision literature. By leveraging tuned training strategies and bridging the gap between static image-based learning and dynamic video testing, this study provides a scalable and intelligent solution for digital heritage preservation. This approach can serve as a baseline for future works involving traditional performance recognition, multimodal cultural datasets, or adaptive classification frameworks in resource-constrained domains.

5. CONCLUSION

This study presents a novel framework for the automated recognition of traditional Indonesian dances by employing a hyperparameter-tuned Convolutional Neural Network, trained on labeled image datasets and evaluated on frame-based segments extracted from real-world video recordings. The proposed model exhibited exceptional performance, achieving 99.67% training accuracy and maintaining 100% accuracy, precision, recall, and F1-score across all test videos, as shown in the classification of every extracted frame from Gambyong, Remo, and Topeng dances. These results highlight the model's ability to generalize from static image learning to dynamic movement recognition in uncontrolled environments. The incorporation of automated hyperparameter optimization marks a key innovation of this research, distinguishing it from earlier works that relied on fixed manual configurations. In real-world contexts, the system holds significant value for cultural preservation, digital documentation, and intelligent archiving of traditional performing arts. For future directions, the framework can be expanded using temporal sequence modeling such as 3D CNNs or LSTM networks,

and tested on more complex scenarios involving multi-view videos, group choreography, or live recognition systems.

ACKNOWLEDGEMENT

This research has been funded by Ministry of Higher Education, Science and Technology (KEMENDIKTISAINTEK) by Decree No. 029/C3/DT.05.00/PL/2025 in the Basic Research Scheme.

REFERENCES

- [1] J. Zhang, "Historical development of Chinese dance: from folklore to Hip Hop," *Research in Dance Education*, pp. 1–18, Apr. 2025, doi: 10.1080/14647893.2025.2486244.
- [2] K. Shinde, "Regional Diversity of Buddhist Heritage Tourism in South Asia and Southeast Asia," *Heritage*, vol. 8, no. 4, p. 121, Mar. 2025, doi: 10.3390/heritage8040121.
- [3] E. Mulyanto, E. M. Yuniarno, I. Hafidz, N. E. Budiyanata, A. Priyadi, and M. Hery Purnomo, "Modified Deep Pattern Classifier on Indonesian Traditional Dance Spatio-Temporal Data," *EMITTER International Journal of Engineering Technology*, vol. 11, no. 2, pp. 214–233, Dec. 2023, doi: 10.24003/emitter.v11i2.832.
- [4] N. Zhen and P. J. Keun, "Ethnic dance movement instruction guided by artificial intelligence and 3D convolutional neural networks," *Sci Rep*, vol. 15, no. 1, p. 16856, May 2025, doi: 10.1038/s41598-025-01879-2.
- [5] Z. Li, "Image analysis and teaching strategy optimization of folk dance training based on the deep neural network," *Sci Rep*, vol. 14, no. 1, p. 10909, May 2024, doi: 10.1038/s41598-024-61134-y.
- [6] N. R. D. Cahyo and M. M. I. Al-Ghiffary, "An Image Processing Study: Image Enhancement, Image Segmentation, and Image Classification using Milkfish Freshness Images," *IJECAIR International Journal of Engineering Computing Advanced Research*, vol. 1, no. 1, pp. 11–22, 2024.
- [7] I. P. Kamila, C. A. Sari, E. H. Rachmawanto, and N. R. D. Cahyo, "A Good Evaluation Based on Confusion Matrix for Lung Diseases Classification using Convolutional Neural Networks," *Advance Sustainable Science, Engineering and Technology*, vol. 6, no. 1, p. 0240102, Dec. 2023, doi: 10.26877/asset.v6i1.17330.
- [8] N. R. D. Cahyo, C. A. Sari, E. H. Rachmawanto, C. Jatmoko, R. R. A. Al-Jawry, and M. A. Alkhafaji, "A Comparison of Multi Class Support Vector Machine vs Deep Convolutional Neural Network for Brain Tumor Classification," in *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, Sep. 2023, pp. 358–363. doi: 10.1109/iSemantic59612.2023.10295336.
- [9] C. Junyou and O. Chantamala, "The Lead Male in the Puxian Opera: Maintaining Cultural Identity and Traditional Body Practice in the Context of Chinese Contemporary Dance," *The International Journal of Critical Cultural Studies*, vol. 22, no. 2, pp. 155–174, 2024, doi: 10.18848/2327-0055/CGP/v22i02/155-174.
- [10] L. Sunarti and T. R. Fadeli, "Preserving Javanese identity and cultural heritage in Malaysia," *Cogent Arts Humanit*, vol. 8, no. 1, Jan. 2021, doi: 10.1080/23311983.2021.1956068.
- [11] J. Qu, "A dance movement quality evaluation model using transformer encoder and convolutional neural network," *Sci Rep*, vol. 14, no. 1, p. 32058, Dec. 2024, doi: 10.1038/s41598-024-83608-9.
- [12] W. Qin and J. Meng, "The research on dance motion quality evaluation based on spatiotemporal convolutional neural networks," *Alexandria Engineering Journal*, vol. 114, pp. 46–54, Feb. 2025, doi: 10.1016/j.aej.2024.11.025.
- [13] K. Zhou, J. Yang, C. C. Loy, and Z. Liu, "Learning to Prompt for Vision-Language Models," *Int J Comput Vis*, vol. 130, no. 9, pp. 2337–2348, Sep. 2022, doi: 10.1007/s11263-022-01653-1.
- [14] L. Liao, H. Li, W. Shang, and L. Ma, "An Empirical Study of the Impact of Hyperparameter Tuning and Model Optimization on the Performance Properties of Deep Neural Networks," *ACM Transactions on Software Engineering and Methodology*, vol. 31, no. 3, pp. 1–40, Jul. 2022, doi: 10.1145/3506695.

-
- [15] Fathir Seniman Tari, "TARI TOPENG SEKARTAJI ISI SURAKARTA || INDONESIA TRADITIONAL MASK DANCE." Accessed: Jun. 17, 2023. [Online]. Available: <https://youtu.be/TioUqzaqrj8>
- [16] APC Kebumen, "Anglep Praba Candrasmurti_Tari Gambyong Pangkur." Accessed: Jun. 17, 2023. [Online]. Available: <https://youtu.be/NjobG1Ptd1o>
- [17] Valentin Vivilia, "Tari Remo Gagrak Anyar." Accessed: Jun. 17, 2023. [Online]. Available: <https://youtu.be/RbeyfjuA8ew>
- [18] A. H. Victoria and G. Maragatham, "Automatic tuning of hyperparameters using Bayesian optimization," *Evolving Systems*, vol. 12, no. 1, pp. 217–223, Mar. 2021, doi: 10.1007/s12530-020-09345-2.
- [19] S. Rahman, M. Ramli, F. Arnia, R. Muharar, and A. Sembiring, "Performance analysis of mAlexnet by training option and activation function tuning on parking images," *IOP Conf Ser Mater Sci Eng*, vol. 1087, no. 1, p. 012084, Feb. 2021, doi: 10.1088/1757-899x/1087/1/012084.
- [20] F. Farhan, C. A. Sari, E. H. Rachmawanto, and N. R. D. Cahyo, "Mangrove Tree Species Classification Based on Leaf, Stem, and Seed Characteristics Using Convolutional Neural Networks with K-Folds Cross Validation Optimalization," *Advance Sustainable Science Engineering and Technology*, vol. 5, no. 3, p. 02303011, Oct. 2023, doi: 10.26877/asset.v5i3.17188.
- [21] M. M. I. Al-Ghiffary, N. R. D. Cahyo, E. H. Rachmawanto, C. Irawan, and N. Hendriyanto, "Adaptive deep learning based on FaceNet convolutional neural network for facial expression recognition," *Journal of Soft Computing*, vol. 05, no. 03, pp. 271–280, 2024, doi: <https://doi.org/10.52465/joscex.v5i3.450>.
- [22] R. R. Ali *et al.*, "Learning Architecture for Brain Tumor Classification Based on Deep Convolutional Neural Network: Classic and ResNet50," *Diagnostics*, vol. 15, no. 5, p. 624, Mar. 2025, doi: 10.3390/diagnostics15050624.
- [23] M. M. Srikantamurthy, V. P. S. Rallabandi, D. B. Dudekula, S. Natarajan, and J. Park, "Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning," *BMC Med Imaging*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12880-023-00964-0.
- [24] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies (Basel)*, vol. 12, no. 7, p. 113, Jul. 2024, doi: 10.3390/technologies12070113.
- [25] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem," *Technologies (Basel)*, vol. 9, no. 4, Dec. 2021, doi: 10.3390/technologies9040081.
- [26] S. C. Kim and Y. S. Cho, "Predictive System Implementation to Improve the Accuracy of Urine Self-Diagnosis with Smartphones: Application of a Confusion Matrix-Based Learning Model through RGB Semiquantitative Analysis," *Sensors*, vol. 22, no. 14, Jul. 2022, doi: 10.3390/s22145445.
- [27] C.-L. Fan, "Evaluation Model for Crack Detection with Deep Learning: Improved Confusion Matrix Based on Linear Features," *J Constr Eng Manag*, vol. 151, no. 3, Mar. 2025, doi: 10.1061/JCEMD4.COENG-14976.
-