EMITTER International Journal of Engineering Technology
Vol. 11, No. 2, December 2023, pp. 214~233
DOI : 10.24003/emitter.v11i2.832

Modified Deep Pattern Classifier on Indonesian Traditional Dance Spatio-Temporal Data

Edy Mulyanto[1,4], Eko Mulyanto Yuniarno[1,2], Isa Hafidz[1,3], Nova Eka Budiyanta[1,5], Ardyono Priyadi[1], Mauridhi Hery Purnomo[1,2]

[1]Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[2]Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[3]Department of Electrical Engineering, Institut Teknologi Telkom Surabaya, Surabaya 60231, Indonesia
[4]Department of Informatics Engineering, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[5]Department of Electrical Engineering, Universitas Katolik Indonesia Atma Jaya, Jakarta, 12930, Indonesia
Corresponding author: hery@ee.its.ac.id

Received September 12, 2023; Revised October 15, 2023; Accepted December 1, 2023

Abstract

Traditional dances, like those of Indonesia, have complex and unique patterns requiring accurate cultural preservation and documentation classification. However, traditional dance classification methods often rely on manual analysis and subjective judgment, which leads to inconsistencies and limitations. This research explores a modified deep pattern classifier of traditional dance movements in videos, including Gambyong, Remo, and Topeng, using a Convolutional Neural Network (CNN). Evaluation model's performance using a testing spatio-temporal dataset in Indonesian traditional dance videos is performed. The videos are processed through frame-level segmentation, enabling the CNN to capture nuances in posture, footwork, and facial expressions exhibited by dancers. Then, the obtained confusion matrix enables the calculation of performance metrics such as accuracy, precision, sensitivity, and F1-score. The results showcase a high accuracy of 97.5%, indicating the reliable classification of the dataset. Furthermore, future research directions are suggested, including investigating advanced CNN architectures, incorporating temporal information through recurrent neural networks, exploring transfer learning techniques, and integrating user feedback for iterative refinement of the model. The proposed method has the potential to advance dance analysis and find applications in dance education, choreography, and cultural preservation.

Keywords: Convolutional neural network, Indonesian traditional dance, Modified deep pattern classifier, Spatio-temporal data.

1. INTRODUCTION

Indonesian traditional dance is an artistic expression that portrays the cultural heritage and identity of a society [1]. It involves performances characterized by distinctive movements, traditional costumes, and indigenous music [2]. Despite the high cultural and historical values associated with traditional dance, several issues, without adequate efforts to promote and preserve this cultural heritage, there is a risk that traditional dances will continue to erode and eventually fade away from public consciousness.

Although traditional dances have significant cultural and historical importance, some concerns about their preservation exist. One of them concerns the waning interest of the younger generation in learning and preserving this form of dance. With the advent of popular culture and modernization, young individuals tend to be more drawn to contemporary entertainment and the latest trends, consequently neglecting traditional dance. Additionally, insufficient funding and government support pose challenges in maintaining traditional dance [3]. Without adequate efforts to promote and preserve this cultural heritage, there is a risk that traditional dances will continue to erode and eventually fade away from public consciousness.

One of the challenges faced is the difficulty for modern individuals in distinguishing different types of traditional dances in Indonesia [4]. This is due to several factors. Firstly, the influence of globalization and popular culture, which are increasingly dominant, has exposed modern individuals more to modern and international forms of entertainment. Consequently, they may be less familiar with Indonesian traditional dances and lack knowledge about the distinctive characteristics, costumes, and movements that differentiate each dance form [5]. Secondly, changes in modern lifestyles and priorities also affect the interest and understanding of traditional dances. This lack of knowledge and understanding results in difficulties in distinguishing which dance is being observed [5],[6],[7],[8]. Therefore, enhancing awareness, education, and appreciation of Indonesian traditional dances in modern society is important to preserve and uphold this cultural heritage.

The study utilizes Convolutional Neural Network (CNN) architecture, a renowned approach in the field of image processing and pattern recognition is performed [9]. By employing convolutional layers, CNN can identify essential features such as movement patterns, body shapes, costume textures, and stage layouts that differentiate each type of traditional dance. Furthermore, pooling layers are used to reduce the dimensionality of these features while retaining crucial spatial distribution information [10]. Additionally, CNN can recognize complex movement patterns and coordination among dancers, including group formations, interactions between dancers, and tempo changes. Through an iterative training process, the CNN model continuously improves its identification capabilities by

optimizing parameters and minimizing errors [11]. Furthermore, the researchers investigated the application of Deep Convolutional Neural Networks (DCNN) in achieving high accuracy in classifying traditional Indian dance styles [9]. The ResNet50 architecture, in combination with Deep Convolutional Neural Networks (DCNN), was employed. The researcher offers valuable insights into the potential of CNN in dance classification, making significant contributions to researchers, choreographers, and dance enthusiasts alike. Furthermore, the findings of this study have substantial implications for automating the analysis and documentation processes associated with cultural dance forms [9]. In other work, CNN structure was optimized for the classification of eight distinct Indian Classical Dances [12]. The findings highlight the potential of this method in improving accuracy and efficiency in various classification applications involving CNN. A set of CNN models was employed to analyze, categorize, and create representations of traditional African dances using video data [13]. Additionally, the researchers applied a specialized human pose estimation algorithm to one of the dance datasets, developing a transferable model that can be utilized across various environments. CNN can be used effectively in classifying dance forms to provide insights and contributions from researchers, choreographers, and dance enthusiasts to automate the analysis and documentation of cultural dances.

In this research, our primary focus was on classifying Indonesian traditional dances using the CNN algorithm and its various layers. We use CNNs because they are designed to handle spatial data such as images and videos. The convolution layer of CNN can extract important spatial features from each video frame, which helps the model understand the spatial patterns in each time step. The concept of shared weights in CNN allows the model to use and update the same parameters for each small area of input. This is particularly appropriate for visual data such as conventional dance videos as it allows the model to learn useful local patterns across the input space. CNNs have a structure that allows learning features hierarchically. Early layers learn simple feature-detectors, and deep layers learn more complex and abstract features. This invariance is especially important for dance video classification, where motion can occur in various positions or orientations. However, the CNN's invariance to spatial shifts and rotations allows the model to recognize important movements or patterns at multiple levels of the video. We aimed to explore the effectiveness of CNN in accurately categorizing and distinguishing different styles of Indonesian traditional dances. Our proposed contribution to this research can be described as follows.

1. Modified deep pattern classifier of dance movements in videos using a CNN to recognize sequential dance movements accurately.
2. Evaluation model's performance using a testing spatio-temporal dataset in Indonesian traditional dance videos is performed.

3. The model's performance was evaluated on a separate subset of the dataset with video data for testing.
4. Visual analysis was performed with color-coded overlays to verify the identified dance elements.

This research contributes to society through the preservation and documentation of cultural heritage as well as preserving traditional dances. This research also provides educational resources about traditional dance. The results of this research can make significant contributions to education and cultural conservation. Knowing and understanding traditional dances can raise public awareness of their cultural wealth. Traditional dance can increase the attractiveness of cultural tourism. This research introduces traditional dance to foreigners, so it can lead to new economic opportunities and help sustainable development.

This article is organized into several sections: Section 2 discusses related research. Section 3 discusses the originality of this research. Section 4 presents the design of the system used in the study and Section 5 presents the results and discussion, Section 6 concludes the article.

2. RELATED WORKS

Several studies have also utilized dance video datasets [1],[4]. These research efforts employed various deep learning methods to train on the datasets, aiming to generate optimal network models. Among these studies, some employed CNN architecture methods. A CNN architecture is a type of neural network that exhibits a grid-like topology and utilizes convolutional operations for feature extraction on data. CNN has emerged as the dominant and powerful network architecture in image processing and computer vision [14],[15].

There are several methods for classifying video data, one of which is classifying one frame at a time. This method involves using CNN to examine all frames in each clip individually [16]. Another approach involves extracting features using Convolutional Neural Networks (CNN) and passing the sequences to an RNN. This model first processes each video frame through Inception, saving the output from the network's final layer. It then converts this into an extracted feature set for training in the RNN model, which utilizes LSTM layers [17],[18],[19],[20],[21]. Other studies classify videos by selecting the best frames for the training process [21],[22],[23],[24]. In the paper [9], the frame extraction method identifies areas that provide information for each frame and selects frames based on the similarities between those areas. Frames used for subsequent processes are identified using optical flow transfer between frames [21],[22]. The selection of the best frames in [23],[24] is based on frame scores generated from the proposed calculation method. Score calculations take into account the representation and features of the frames.

There are several studies that discuss traditional Indonesian dance. Paper [25] on the documentation of traditional Indonesian dance motion capture, a

study that discusses the analysis of Balinese dance silhouette sequence patterns using Bag of Visual Movement with HoG and SIFT features. Furthermore, paper [26] about detecting the same pattern in Balinese dance choreography using Convolutional Neural Network (CNN) and Analysis Suffix Tree. This research aims to classify and analyze patterns in Balinese dance using deep learning techniques. Paper [27] is research that proposes the classification of Pakarena dance images, which is a traditional dance from South Sulawesi, Indonesia, using the Convolutional Neural Network (CNN) algorithm. This research achieved an accuracy rate of 92.5% for image classification.

3. ORIGINALITY

Several studies address the classification of traditional dances using deep learning. Paper [28] proposed classifying Indian classical dance actions using CNN to identify and classify various movements in Indian classical dance. The paper modified the CNN with 3 Dropout functions and 2 Stochastic pooling and continued to flatten and fully connected. Another study focused on classifying Turkish folk dances using deep learning techniques, including CNN [29]. This study used the pre-trained VGG16 architecture, making no modifications to the VGG16 network. Another study [30] proposed a method to understand dance semantics using spatio-temporal features coupled with an RNN network. Although not specific to traditional dances, this approach demonstrates the use of spatio-temporal features, which can be relevant to classifying traditional and folk dances. This research requires considerable learning time as it processes two architectures.

The originality in this research lies in modifying the CNN architecture by combining different layers with the aim to explore the effectiveness of CNN in accurately categorizing and differentiating various Indonesian traditional dance styles. Our proposed contribution is to modify the layers by adding a convolution layer combined with normalization to recognize features in the video clip frames, as shown in Figure 2. The system framework of the research begins with the input in the form of a video dataset, and extracted into frames that will be trained into the proposed network to obtain a model, the model is used to classify the dance video. The modified Deep Pattern Classifier model by setting Parameters and Hyperparameters, which are shown in table 1. The performance of the model is evaluated by using a spatio-temporal test dataset of Indonesian traditional dance videos. We also analyzed the resulting model under overfitting conditions and good training conditions. Visual analysis was performed using color-coded overlays to verify the identified dance elements, as shown in Figure 5 and Figure 6.

4. SYSTEM DESIGN

The Indonesian Traditional Dance dataset consists of a comprehensive collection of information regarding three types of traditional dances in Indonesia: Gambyong dance, Topeng dance, and Remo dance. This dataset comprises a total of 13,354 samples, with each sample represented as a 200×200 pixel image containing three color channels (RGB). The dataset is divided into two main components: an image dataset and a video dataset.

The Image dataset was divided into a 20% testing subset and an 80% training subset. The testing subset, comprising approximately 2,671 samples, was utilized to evaluate the model's performance on unseen data. Conversely, the training subset, consisting of around 10,683 samples, was employed to train the model in recognizing and differentiating dance movements. This division ensures a balanced approach between training and evaluation, enabling effective development and assessment of the model's capabilities. Figure 1. shows sample images of Indonesian traditional dance with models (a) topeng, (b) gambyong, and (c) remo.

The testing dataset was expanded by incorporating traditional dance videos from YouTube as additional resources [31],[32],[33]. These videos were processed by segmenting them into individual frames, effectively converting them into image format. Each frame independently, capturing the intricate details of posture, footwork, and facial expressions demonstrated by the dancers.

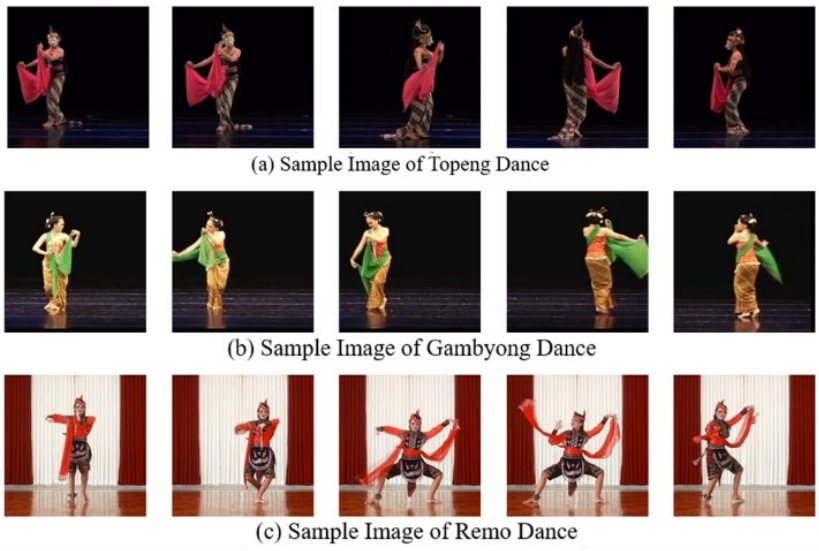

Figure 1. Sample Image of Indonesian Traditional Dance with Models (a) Topeng, (b) Gambyong, and (c) Remo.

The flow diagram of the system is shown in Figure 2. Divide the dataset at random into K folds. Verify that each fold's class distribution is balanced, particularly if the dataset contains classes that aren't balanced. Establish a set of optimisation parameters, such as the learning rate, batch size, and number of epochs. Use the training set to train the model for a predetermined

number of epochs. To gauge the performance of the model, use test data. Use suitable assessment metrics, such as accuracy, precision, recall, or F1-score, to assess the model on test data. Restart the process from the beginning by substituting new values for the parameters. Based on assessment metrics, identify the settings that yield the greatest results. To train the model on the complete dataset, use optimum parameters.

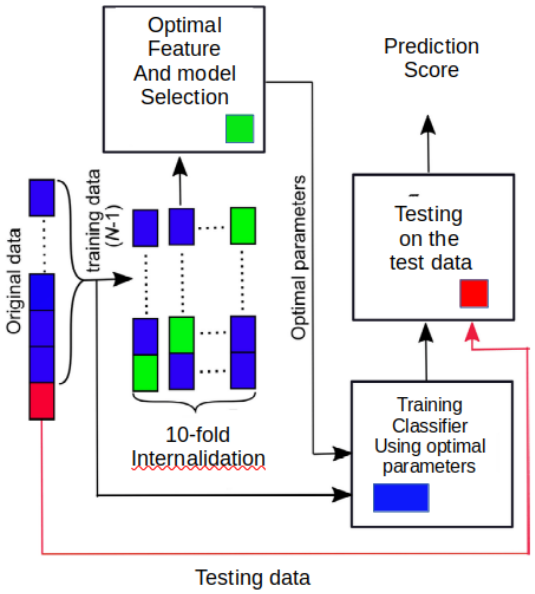

Figure 2. Flow System Diagram Cross Validation

4.1 Convolutional Neural Network (CNN)

CNN is based on the fundamental concept of convolution, which involves applying small filters (kernels) to input data with a specified stride (34). These filters extract important features from the input by engaging in mathematical convolution operations [35]. This convolution process aids the network in identifying visual patterns such as edges, corners, textures, and other relevant features within the data. The architecture of a CNN consists of several distinct layers, including convolutional layers, pooling layers, and fully connected layers [36]. The final fully connected layer connects the relevant features to the classes that need to be identified [37]. The convolution operation is applied to input data using learnable filters or kernels. The equation for the convolution operation in CNN can be explained in equation (1).

$$(f \times w)[i,j] = \sum_{m} \sum_{n} f[m,n] \times w[i-m, j-n] \tag{1}$$

where $(f \times w)[i,j]$ represents the output value at the position (i,j) after applying convolution to the input data f using the convolutional kernel w.
The summation is performed over the spatial dimensions of the input data, and the kernel is applied to the corresponding receptive field. Figure 3

illustrates the architecture of layer in modified deep pattern classifier using CNN, and Figure 4 presents a flowchart of the modified deep pattern classifier to improved pattern recognition and classification of Indonesian traditional dances through spatio-temporal dataset.

The procedure for modified deep pattern classifier in Figure 4 can be explained as follows.

1. Take input spatio-temporal dataset images/videos with a size of 200×200 pixels and 3 color channels (RGB).
2. Perform convolution on the input using multiple convolutional layers with different filters. Each convolutional layer is followed by a batch normalization layer to improve network stability. Apply the ReLU activation function after the batch normalization layer.
3. Use max pooling layers after each convolutional layer to reduce spatial dimensions and computational complexity. In this process, the maximum value within each pooling region is selected as the most significant representation of the features.
4. Repeat the convolution and max pooling steps to generate deeper and more abstract features.
5. Finally, add fully connected layers that connect all neurons from the previous layers to the output layer. The number of neurons in the output layer corresponds to the number of categories or classes to be classified.
6. Add a softmax layer at the output layer to generate probability distributions for each class. This allows the network to produce probabilities for each class, indicating the likelihood of the input image/video belonging to each class.
7. Use a classification layer to determine the class with the highest probability for the input image/video.
8. Train the CNN model using the Adam optimizer with a mini-batch size of 128 and a maximum of 4 epochs. During training, monitor the model's progress using a validation dataset. Once trained, apply the CNN model to classify spatio-temporal dataset frames extracted from videos. Perform additional preprocessing steps if necessary.

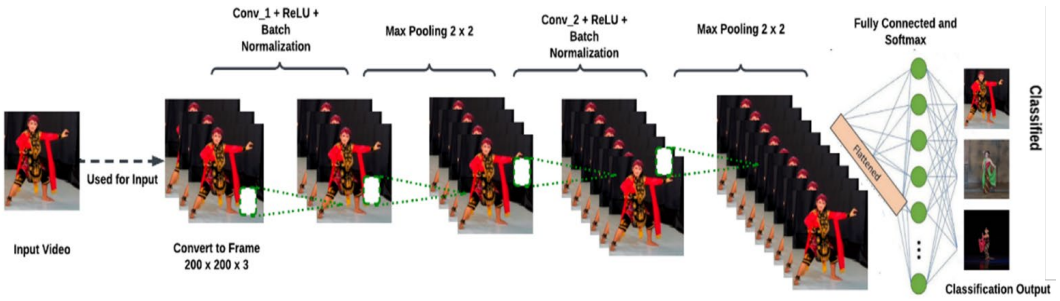

Figure 3. Structure of layer in modified deep pattern classifier using CNN from Indonesian traditional dance video

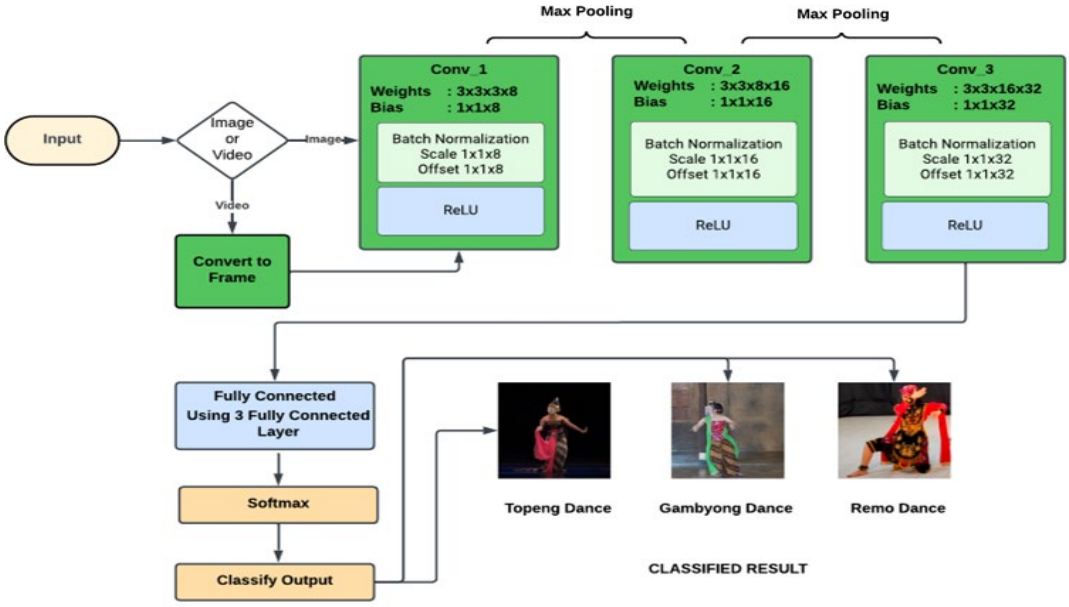

Figure 4. Flowchart of the modified deep pattern classifier to improved pattern recognition and classification of Indonesian traditional dances through spatio-temporal dataset

9. Obtain predicted labels for each frame using the trained model and display the frames with their corresponding predicted labels for visual inspection.
10. Additionally, apply frame differencing techniques to detect and highlight moving objects in consecutive frames. By comparing the grayscale representations of the current and previous frames, calculate the absolute difference. Regions with pixel differences above a threshold are marked with a yellow color overlay on the frames.

4.2 Confusion Matrix

In classification analysis, the confusion matrix is a vital tool used to evaluate the performance of a classification model [38]. The confusion matrix depicts the comparison between the actual labels and the predicted labels of samples within a dataset. Due to evaluate results for each CNN layer, the equations can be written in (2) until (5).

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{2}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{3}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{4}$$

$$F1-score = \frac{2*(Precision*Recall)}{(Precision+Recall)} \tag{5}$$

where True Positive (TP) is the number of positive samples correctly classified, False Positive (FP) is the number of negative samples incorrectly classified as positive, False Negative (FN) is the number of positive samples incorrectly classified as negative, and True Negative (TN) is the number of negative samples correctly classified. By utilizing TP,TN,FP, and FN, several evaluation metrics can be computed. Accuracy portrays the extent to which the model successfully classifies samples overall.

Precision measures the model's ability to correctly classify positive samples. Recall (Sensitivity) describes the model's capability to detect all positive samples present. F1-score provides the harmonic mean between precision and recall, offering an overall assessment of the model's performance [39]. The formulas for these evaluation metrics can be calculated based on the values within the confusion matrix [40]. In the conducted classification analysis, the confusion matrix and related evaluation metrics are employed to comprehensively evaluate the performance of the classification model [41].

4.3 Cross-Validation

Cross-validation is one of the effective techniques to address overfitting in machine learning models. Cross-validation helps measure the extent to which the model can generalize to unseen data, thus helping to identify whether overfitting is occurring. In this study to address overfitting, we separate the dataset into two parts: one for training and one for testing. This split was randomized, ensuring that both datasets represented a uniform distribution. Furthermore, we also perform cross-validation when performing hyperparameter tuning to avoid choosing parameters that are optimized for only one particular data partition.

Overfitting occurs when a machine learning model gets too used to training data, causing performance degradation on new data. Overfit models learn data training patterns, including noise and errors, to the point of losing the ability to generalize invisible data accurately. Overfitting happens due to the model's complexity having too many parameters compared to the available training data. Over-training can also lead to overfitting as the model adapts to uniqueness and irrelevant variations in the training data. A good model and an overfitted model represent two different scenarios in machine learning. A good model balances generalizing and making accurate predictions on unseen data. It captures underlying patterns in data training without too much noise pressure or irrelevant variation. On the other hand, overfitted models fail to generalize to new data and focus more on details and variations in data training, resulting in poor performance on unseen data. Overfitted models tend to be complex and sensitive to small fluctuations in the training data. Conversely, a good model strikes a harmonious balance between capturing important patterns and avoiding unnecessary complexity, resulting in more robust and reliable predictions of new data.

5. EXPERIMENT AND ANALYSIS

This study initially employed a classical CNN architecture to develop a deep learning model without utilizing specialized architectures. However, upon analyzing the training results, evidence of overfitting in the model became apparent. The overfitting led to an accuracy level of 99.89%, indicating that the model essentially memorized the training data perfectly but struggled to generalize effectively to new data. This observation is reflected in the accuracy and loss graphs during training. The loss graph demonstrated a downward trend for the training data, whereas the validation loss remained stagnant or even increased. To address this issue, the researchers decided to modified deep pattern classifier of dance movements in videos using CNN by incorporate the GoogleNet architecture into the model. By incorporating proposed method, it is anticipated that overfitting can be reduced, and the model's capability to generalize to new data can be enhanced.

Based on Figure 5, the utilization of the classic architecture model resulted in overfitting, despite the inclusion of regularization techniques such as dropout and others. Despite these efforts, the model still exhibited a tendency to perfectly memorize the training data and struggled to generalize well to new data. However, in Figure 6, the incorporation of the proposed architecture and the addition of several features, including dropout, successfully reduced the occurrence of overfitting. Consequently, it can be concluded that the utilization of the GoogleNet architecture, along with its inherent features such as inception modules, dimensionality reduction, auxiliary classifiers, and regularization techniques, resulted in significant improvements in addressing the overfitting issue within the CNN model.

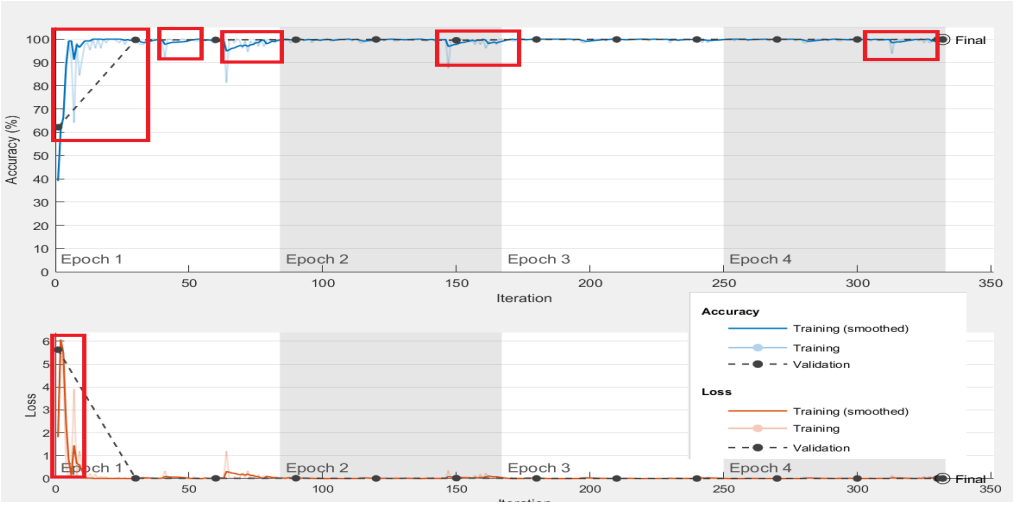

Figure 5. Overfiting Training Progress with Classic Architecture (Red represents overfitting)

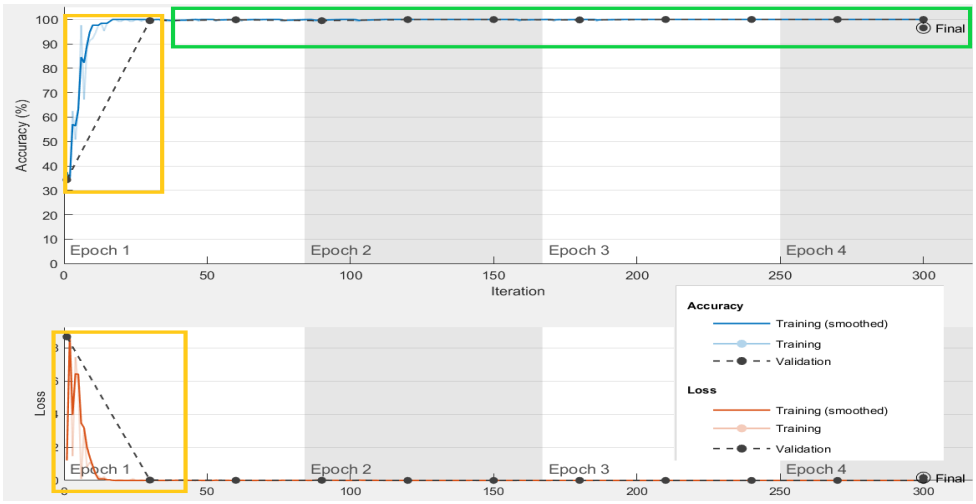

Figure 6. Good Training Progress with Proposed Architecture (Green represents Normal and yellow represents Reduce the overfit

Overfit reduction based on GoogleNet contributes to reducing overfitting, including utilising inception modules and dimensionality reduction techniques. The inception modules in GoogleNet are designed to capture features at multiple scales and resolutions, allowing the model to learn diverse and discriminative features. This helps prevent overfitting by promoting generalization to unseen data. Additionally, the architecture incorporates 1×1convolutions as a means of dimensionality reduction, effectively reducing the model's complexity and preventing it from memorizing the training data excessively. By controlling the model's capacity, GoogleNet mitigates the risk of overfitting and encourages better generalization. Table 1 presents the parameters and hyperparameters of the CNN model, while Table 2 compares the layers between the classic architecture model and the GoogleNet model.

Based on Equation 1, the algorithm processes the input data using the parameters and hyperparameters specified in Table 1 and Table 2 can be explain as follow:

1. Input Tensor:
   $Input = W \times H \times C = 200 \times 200 \times 3$
2. Convolutional Layer:
   - Filter Size: 7×7
   - Number of Filters: 64
   - Stride: $S \leftarrow 2$
   - Padding: $P \leftarrow W \times H = 200 \times 200 \times 3 'same'$
   - Activation Function: ReLU
   - Output Size: $\left[\frac{H}{2}, \frac{W}{2}, 64\right] \leftarrow 100,100,64$
3. Max Pooling Layer:
   - Pooling Size: $3x3$
   - Stride: $S \leftarrow 2$

- Output Size: $\left[\frac{H}{4}, \frac{W}{4}, 64\right] \leftarrow 50,50,64$

4. Inception Module 1:
   - Convolutional Layer 1:
     Filter Size: 1×1
     Number of Filters: 32
     Activation Function: ReLU
     Output Size: $50 \times 50 \times 64$
   - Convolutional Layer 2:
     Filter Size: 3×3
     Number of Filters: 96
     Padding: $P \leftarrow W \times H = 200 \times 200 \times 3 'same'$
     Activation Function: ReLU
     Output Size: $50 \times 50 \times 96$
   - Convolutional Layer 3:
     Filter Size: 5×5
     Number of Filters: 16
     Padding: $P \leftarrow W \times H = 200 \times 200 \times 3 'same'$
     Activation Function: ReLU
     Output Size: $50 \times 50 \times 16$
   - Max Pooling Layer:
     Pooling Size: 3×3
     Stride: $S \leftarrow 1$
     Padding: $P \leftarrow W \times H = 200 \times 200 \times 3 'same'$
     Output Size: $50 \times 50 \times 64$
   - Concatenation Layer:
     Input: [Output of Conv Layer 1, Output of Conv Layer 2, Output of Conv Layer 3, Output of Max Pooling Layer]
     Output Size: $50 \times 50 \times 240$
5. Average Pooling Layer:
   - Pooling Size: 7×7
   - Output Size: 1x1x240
6. Dropout Layer:
   - Dropout Rate: 0.25
   - Output Size: $1 \times 1 \times 240$
7. Fully Connected Layer:
   - Number of Neurons: 3
   - Activation Function: Softmax
   - Output: Probability Distribution over Classes

where H represents Height, W represents Width, C represents Channel, S represents Stride, and P represents Padding. The output of the concatenated tensor is the result of the Inception layer. Based on the training progress

mentioned above, the video was processed and divided into 1591 frames. Each frame represents an individual image. This processing was completed within 93 seconds. To evaluate the model's performance, the dataset used for training was split into three subsets, ensuring a balanced distribution of labels across each subset. The model's predictions were then compared against the ground truth labels, resulting in the confusion matrix for the 1591 predicted frames.

Table 1. Parameter and Hyperparameter of Modified Deep Pattern Classifier Model

Hyperparameter	**Parameter**
Adam Optimazation	L1 and L2 Regulation
128 Mini Batch Size	25% Dropout Rate
4 MaxEpoch	-
30 Validation Frequency	-

Table 2. A Comparison Layers of Classic Architecture Model and GoogleNet Model

Classic Layers (Overfitting)	**Proposed Layers (Reduced)**
Input image with dimensions 200x200x3	Input image with dimensions 200x200x3
Convolution with 3x3 filter size and 8 filters	Convolution with 3x3 filter size and 64 filters
Batch normalization	Batch normalization
ReLU activation function	-
Max pooling with 2x2 filter size and stride 2	ReLU activation function
Max pooling with 3x3 filter size and stride 2	-
Classic Layers (Overfitting)	**Proposed Layers (Reduced)**
Convolution with 2x2 filter size and 16 filters	Convolution with 3x3 filter size and 64 filters
Convolution with 3x3 filter size and 32 filters	InceptionLayer
25% Dropout layer	Inception module with various parallel branches
Fully connected layer with 3 output classes	Average pooling with 7x7 filter size
Softmax for probabilistic classification	25% Dropout layer
Classification layer	Fully connected layer with number of output classes
-	Softmax for probabilistic classification
-	Classification layer

From the obtained confusion matrix, various performance metrics such as accuracy, precision, sensitivity (recall), and F1-score can be derived.

Accuracy measures the extent to which the model correctly classifies all classes. Figure 7 shows the comparison of the confusion matrix in gambyong with (a) Overfited Model and (b) Good Model. Precision quantifies the correctness of positive predictions made by the model. Sensitivity assesses the model's ability to identify positive samples accurately. The sensitivity value directly affects the accuracy metric; higher sensitivity indicates better identification of positive cases. The F1-score is the harmonic mean of precision and sensitivity, providing a more objective evaluation in imbalanced datasets. These metrics offer a comprehensive understanding of the model's performance and are commonly employed to compare classification models. Table 3 describes a comparative analysis of model identification when testing various dance movements using captured frames.

Table 3. A comparison of model identification for testing different dance movements using captured frames.

Testing	Testing Dataset	Precision	Sensitivity / Recall	F1-Score
Overfitted Testing	Gambyong.mp4	98.7%	100%	99.3%
Good Testing	Gambyong.mp4	96.9%	100%	98.4%
Overfitted Testing	Remo.mp4	100%	93%	96.9%
Good Testing	Remo.mp4	100%	96.9%	98.4%
Overfitted Testing	Topeng.mp4	97.3%	97%	98.1%
Good Testing	Topeng.mp4	96.7%	97.5%	98.6%

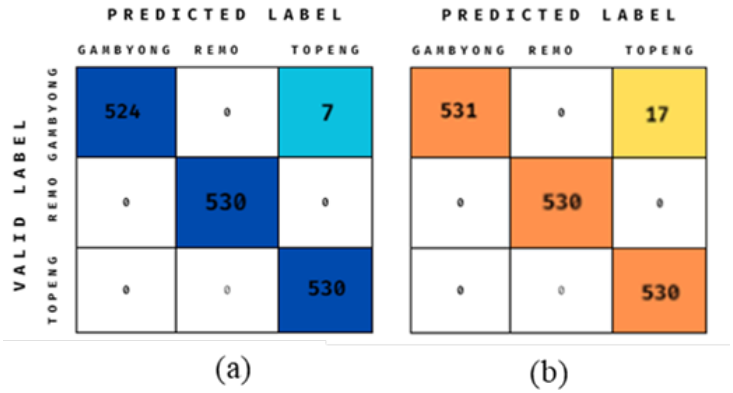

Figure 7. A comparison of confusion matrix in gambyong with (a) Overfited Model and (b) Good Model

Based on Table 3, classification of the spatio-temporal frame dataset based on motion identification can be successfully performed with impressive accuracy. In Good Testing, the average precision is 97.86%, the average recall is 98.13% and the average F1-Score is 98.46%. This high level of accuracy attests to the effectiveness of the employed classification model,

particularly in discerning and categorizing different dance movements from the captured frames. The successful classification of these images provides valuable insights into the potential applications of motion-based image analysis and highlights the model's robustness in handling complex visual data. By accurately classifying each frame, valuable information about the dynamics and patterns of the dance movements can be extracted, contributing to a deeper understanding and analysis of the performance.

6. CONCLUSION

Traditional dances in Indonesia have complex classification, but existing methods still rely on manual and subjective analysis. This study explores the modified deep pattern classifier of dance movements in videos using a CNN, which were processed by segmenting them into spatio-temporal dataset. By analyzing each frame independently, the model captured intricate details of posture, footwork, and facial expressions demonstrated by the dancers. The results demonstrate a high accuracy and reliable classification of the dataset. These outcomes underscore the potential of proposed method in addressing complex image classification tasks and inspire confidence in the model's ability to make dependable predictions in analogous datasets or real-world scenarios. Furthermore, the classification of individual frames based on motion identification yielded an impressive accuracy, In Good Testing, the average precision is 97.86%, the average recall is 98.13% and the average F1-Score is 98.46%. and in Overfitting Testing, the average precision is 98.6%, the average recall is 96.6% and the average F1-Score is 98.1%. This demonstrates the effectiveness of the classification model in discerning and categorizing different dance movements from the captured frames. The successful classification of these images highlights the robustness of the model in handling complex visual data and provides valuable insights into the potential applications of motion-based image analysis. Accurate classification of each frame also contributes to a deeper understanding and analysis of dance performance by extracting valuable information about the dynamics and patterns of the movements.

For future research, several areas can be explored for applying modified deep pattern classifier to dance classification. Delving into advanced CNN architectures such as ResNet or DenseNet can improve model performance by capturing complex features and patterns in the video, combining temporal information through RNN or attention mechanisms to improve the model's understanding of the temporal dynamics of dance movements. Learning transfer with trained models on large-scale video data sets can improve generalizability and efficiency. In addition, it involves feedback through active learning strategies to create a dance classification system that is more interactive and user-centered in education. This study opens up an avenue for dance analysis and find applications in dance education, choreography, and cultural preservation.

Acknowledgments

We would like to express our gratitude to Ministry of Research, Technology, and Higher Education of the Republic of Indonesia for providing research funding through the Penelitian Kompetitif Nasional-Penelitian Fundamental, SBK Riset Dasar 2023.

REFERENCES

[1] Emmanuel C. Maraña, Ramiella Anne A. Arpon, Irvin Lance L. Capuchino, Romiele Anne F. Casiño, Kate Rossleth T. Casuga, Jenifer G. Aguilar. **Strengthening of best practices in the preservation of cultural diversities**: *A phenomenological research. GSC Adv Res Rev* 2023; Vol. 15, pp.46–62, 2023

[2] Shoji M, Takafuji Y, Harada T. **Behavioral impact of disaster education: Evidence from a dance-based program in Indonesia**. *Int J Disaster Risk Reduct* ,Vol. 45, 2020

[3] Tresnawaty B, Risdayah E**. Religion and Media: Anthropological study of religious behavior in the film "Little House on the Prairie"**. *ETNOSIA J Etnografi Indones*, Vol. 8, pp.116–125, 2023

[4] Cruz AGB, Seo Y, Scaraboto D. **Between Cultural Appreciation and Cultural Appropriation: Self-Authorizing the Consumption of Cultural Difference**. *J Consum Res*, 2023

[5] Domingues AR, Mazhar MU, Bull R. **Environmental performance measurement in arts and cultural organisations: Exploring factors influencing organisational changes**. *J Environ Manage*, Vol 326, 2023

[6] Natar M, Age MYC. CACI: **The Contradiction Between the Nature and Practice of Modern Manggarai Society with Its Relevance to the Character Formation of the Millennial Generation**. *Int J Soc Serv Res*, Vol. 3, pp. 1166–1172, 2023

[7] Handayani R, Narimo S, Fuadi D, Minsih M, Widyasari C. **Preserving Local Cultural Values in Forming the Character of Patriotism in Elementary School Students in Wonogiri Regency**. *J Innov Educ Cult* Res Vol. 4, pp. 56–64, 2023

[8] Jain N, Bansal V, Virmani D, Gupta V, Salas-Morera L, Garcia-Hernandez L. **An Enhanced Deep Convolutional Neural Network for Classifying Indian Classical Dance Forms**. *Appl Sci* Vol.11, pp. 6253, 2021

[9] Mu J. **Pose Estimation-Assisted Dance Tracking System Based on Convolutional Neural Network**. *Comput Intell Neurosci*, pp.1–10. 2022

[10] Victoria AH, Maragatham G. **Automatic tuning of hyperparameters using Bayesian optimization**. *Evol Syst* 2021, Vol.12, pp.217–223, 2021

[11] Challapalli JR, Devarakonda N. **A novel approach for optimization of convolution neural network with hybrid particle swarm and grey wolf algorithm for classification of Indian classical dances**. *Knowl Inf Syst* 2022, Vol.64, pp. 2411–2434, 2022

[12] Odefunso AE, Bravo EG, Chen YV. **Traditional African Dances Preservation Using Deep Learning Techniques**. *Proc ACM Comput Graph Interact Tech 2022*, Vpl.5, pp.1–11, 2022

[13] Liu M, Jervis M, Li W, Nivlet P. **Seismic facies classification using supervised convolutional neural networks and semisupervised generative adversarial networks**. *GEOPHYSICS, Vol.85*, p. 47–58, 2020

[14] Ranjbarzadeh R, Dorosti S, Jafarzadeh Ghoushchi S, Safavi S, Razmjooy N, Tataei Sarshar N, et al. **Nerve optic segmentation in CT images using a deep learning model and a texture descriptor**. *Complex Intell Syst, Vol.*8, pp. 43–57, 2022

[15] Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and F. F. Li, **Large-scale video classification with convolutional neural networks**, *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1725-1732, 2014

[16] J. Y. H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, **Beyond short snippets: Deep networks for video classification**, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 07, pp. 4694-4702, 2015

[17] A. Yenter and A. Verma, **Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis** *in 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2017*, pp. 540-546, 2017

[18] E. Ergün, F. Gürkan, O. Kaplan and B. Günsel, **Video action classification by deep learning**, *2017 25th Signal Processing and Communications Applications Conference (SIU)*, Antalya, Turkey, pp. 1-4, 2017

[19] M. Abdullah, M. Ahmad and D. Han, **Facial Expression Recognition in Videos: An CNN-LSTM based Model for Video Classification**, 2020 International Conference on Electronics, Information, and Communication (ICEIC), *Barcelona, Spain*, pp. 1-3, 2020

[20] M. A. Russo, A. Filonenko and K. -H. Jo, **Sports Classification in Sequential Frames Using CNN and RNN**, *2018 International Conference on Information and Communication Technology Robotics (ICT-ROBOT)*, Busan, Korea (South), pp. 1-3, 2018

[21] Savran Kızıltepe, R., Gan, J.Q. & Escobar, J.J. **A novel keyframe extraction method for video classification using deep neural networks**. *Neural Comput & Applic,* 2021

[22] S. Kulhare, S. Sah, S. Pillai and R. Ptucha, **Key frame extraction for salient activity recognition**, *2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, pp. 835-840, 2016

[23] S. Jadon and M. Jasim, **Unsupervised video summarization framework using keyframe extraction and video skimming**, *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India, pp. 140-145, 2020

[24] H. Wang, F. Nie, H. Huang, and Y. Yang**, Learning frame relevance for video classification**, *Proceedings of the 19th ACM international conference on Multimedia*, pp. 1345–1348, 2011.
[25] Hegarini, E., Dharmayanti, Syakur, A.. **Indonesian traditional dance motion capture documentation**. *2016 2nd International Conference on Science and Technology-Computer (ICST)*, pp.108-111. 2016.
[26] Jaya, I.K.H.T, Kesiman, M.W.A, Sunarya, I.M.G., **Detecting the Same Pattern in Choreography Balinese Dance Using Convolutional Neural Network and Analysis Suffix Tree**, *Scientific Journal of Electrical Engineering Computers and Informatics*, Vol 8, No 3, 2022.
[27] Ibrahim; Abdul, Rachmat, **Pakarena dance image classification using convolutional neural network algorithm**, *ILKOM Scientific Journal*, Vol. 13, No. 2, pp. 134-139, 2021
[28] Kishore, P.V.V., Kumar, K V K , Eepuri, Kiran & Sastry, A , Maddala, Teja , Anil Kumar, D. & Prasad, M.V.D.. (2018). **Indian Classical Dance Action Identification and Classification with Convolutional Neural Networks***. Advances in Multimedia*, pp. 1-10, 2018.
[29] Nazari, H., Kaynak, S, **Classification of Turkish Folk Dances using Deep Learning**. *International Journal of Intelligent Systems and Applications in Engineering*, pp. 226–232, 2022.
[30] Shailesh, S, Judy, M.V., **Understanding dance semantics using spatio-temporal features coupled GRU networks**. *Entertain Comput.*, pp. 484, 2022.
[31] Tari Topeng Sekartaji Isi Surakarta, Indonesia Traditional Mask Dance - YouTube n.d. https://www.youtube.com/watch?v=TioUqzaqrj8 (accessed August 3, 2023).
[32] Anglep Praba Candrasmurti_Tari Gambyong Pangkur - YouTube n.d. https://www.youtube.com/watch?v=NjobG1Ptd1o (accessed August 3, 2023)
[33] Tari Remo Gagrak Anyar - YouTube n.d. https://www.youtube.com/watch?v=RbeyfjuA8ew (accessed August 3, 2023)
[34] Ghosh A, Sufian A, Sultana F, Chakrabarti A, De D. **Fundamental Concepts of Convolutional Neural Network***. In: Balas VE, Kumar R, Srivastava R, editors. Recent Trends Adv. Artif. Intell. Internet Things, vol. 172, Cham: Springer International Publishing*, pp. 519–67, 2020
[35] Khan A, Sohail A, Zahoora U, Qureshi AS. **A survey of the recent architectures of deep convolutional neural networks**. *Artif Intell Rev*, Vol. 53, pp. 455–516, 2020
[36] Michael Onyema E, Balasubaramanian S, Suguna S K, Iwendi C, Prasad BVVS, Edeh CD. **Remote monitoring system using slow-fast deep convolution neural network model for identifying anti-social activities in surveillance applications**. *Meas Sens*, Vol. 27, 2023

[37] Diwan T, Anirudh G, Tembhurne JV. **Object detection using YOLO: challenges, architectural successors**, **datasets and applications**. *Multimed Tools Appl* 2023, Vol. 82, pp. 243–275, 2023

[38] Daviran M, Shamekhi M, Ghezelbash R, Maghsoudi A. **Landslide susceptibility prediction using artificial neural networks, SVMs and random forest: hyperparameters tuning by genetic optimization algorithm**. *Int J Environ Sci Technol*, Vol.20, pp.259–276, 2023

[39] Ajayi OG, Ashi J. **Effect of varying training epochs of a Faster Region-Based Convolutional Neural Network on the Accuracy of an Automatic Weed Classification Scheme**. *Smart Agric Technol*, Vol.3, 2023

[40] Kim S-C, Cho Y-S. **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, 2022

[41] Kara OC, Venkatayogi N, Ikoma N, Alambeigi F. **A Reliable and Sensitive Framework for Simultaneous Type and Stage Detection of Colorectal Cancer Polyps**. *Ann Biomed Eng, Vol*.51, pp.1499–1512., 2022