Jurnal Krisnadana (Jurnal Komputer, Sistem Kendali & Jaringan)
Vol 5, Issue 3, May 2026, pp. 713-725
https://doi.org/10.58982/tt9bdq83

E-ISSN: 2808-3563
P-ISSN: 2808-3520

Image Classification of Traditional Musical Instruments from East Nusa Tenggara Using CNN and VGG19

Marvin Gilbrand Adu[1], I Gede Totok Suryawan[2*], Ida Bagus Ary Indra Iswara[3]

[1,2*,3] Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia
[*] totok.suryawan@gmail.com

ARTICLE INFO

Article history:
Received 22 June 2026
Revised 22 July 2026
Accepted 24July 2026

***Keywords**:*
Deep Learning;
Convolutional Neural;
Network;
VGG19;
Traditional Musical Instruments;
Image Classification

ABSTRACT

This research is motivated by the declining ability of the younger generation to recognize the physical forms of traditional musical instruments from East Nusa Tenggara (NTT). This study aims to apply a Convolutional Neural Network (CNN)-based Deep Learning algorithm to classify seven types of musical instruments: Sasando, Moko, Gong, Gendang, Foydoa, Jungga, and Knobe Oh. A dataset of 610 images was used, split 80% for training and 20% for testing, and subjected to preprocessing steps, including resizing and data augmentation. The performance of a Custom CNN architecture was compared against VGG19 (fine-tuning) across variations of learning rates (0.001, 0.0001, and 0.00001). The results showed that the Custom CNN achieved optimal performance at a learning rate of 0.0001, with a training accuracy of 91.63% and a validation accuracy of 82.50%. Meanwhile, the fine-tuned VGG19 model achieved 100% training accuracy and 93.33% validation accuracy. Confusion Matrix evaluation on the test data demonstrated that the best model achieved 100% accuracy, indicating that the system successfully extracts and recognizes the visual features of each instrument.

1. Introduction

East Nusa Tenggara Province (NTT) is an island region in Indonesia that possesses a wealth of highly distinctive material cultural heritage. One of the most significant representations of the ethnic and spiritual identity of the people of NTT is reflected in the diversity of its traditional musical instruments. There are seven (7) main instruments that play a crucial role in customary rites, namely Sasando, Moko, Gong, Gendang, Foydoa, Jungga, and Knobe Oh. However, in the current era of globalization, there has been a massive degradation of visual knowledge among the younger generation. Limited access to education and dependence on conventional direct observation have led to high inconsistency in recognizing the visual characteristics, details, and physical morphology of each instrument. This gap in visual information creates major obstacles in the process of digitalization and preservation of local cultural heritage. Therefore, a computational solution is needed that can serve as a reliable digital visual validation infrastructure.

Artificial Intelligence technology in the field of Computer Vision, through Deep Learning approaches, offers a highly prospective methodology to address this identification crisis. The Convolutional Neural Network (CNN) algorithm is an ideal architecture because it has superior advantages in automatically extracting high-level morphological image features (such as metal surface texture, string density details, and instrument geometry), much like the human visual system. The integration of CNN has proven capable of producing highly precise object classification models for cultural artifacts and is adaptive to variations in visual environments.

The reliability of neural network-based models in the field of cultural preservation has been supported by various previous studies. Study by [1] successfully classified Javanese musical instruments using the MobileNetV2 architecture, achieving a training accuracy of 99.30%. In mycology, [2] pplied a three-layer convolutional CNN for genus-level mushroom classification, obtaining a validation accuracy of 82%. Furthermore, [3] ilized the VGG-16 model with a 20% dropout configuration to classify ten types of traditional Balinese cuisine, achieving an optimal accuracy of 97.5%. In textile visualization, [4] designed a nine-layer feature extraction CNN model that proved robust in recognizing Yogyakarta batik motifs with an accuracy of 87.83%. Finally, [5] demonstrated that the implementation of fine-tuning techniques on the NasNetMobile architecture could boost flower image classification accuracy to 99.15%.

Although previous studies have demonstrated the robustness of CNN architectures, gap analysis indicates that there has not yet been any specific research evaluating the performance of computational models against the visual complexity of various traditional musical instruments from NTT. The novelty (state of the art) of this study lies in conducting an in-depth comparative analysis between the robustness of a self-built Custom CNN architecture and the effectiveness of the VGG19 architecture based on Transfer Learning with fine-tuning applied to the upper convolutional blocks. Both models are rigorously tested across learning rate variations and image augmentation manipulations to analyze the mathematical adaptation limits and the stability of error convergence curves in recognizing the complex morphology of traditional NTT instruments.

2. Method

Related Work

Numerous studies have investigated the application of Convolutional Neural Networks (CNNs) in the fields of computer vision and image processing across various objects of study. As an operational foundation, understanding the fundamental stages of digital image processing is crucial to improve image quality prior to model analysis—a basic concept widely applied in vocational academic environments [6].

The reliability of these Deep Learning models has been proven effective in solving identification problems in agriculture, botany, and biology. In the agricultural sector, baseline CNN models are capable of accurately identifying and distinguishing types of diseases on potato leaves [7]. For flora and organic materials, CNNs have been implemented to classify various flower images [8], as well as to differentiate traditional spices with high similarity in texture and shape [9]. In the field of animal taxonomy, Deep Learning approaches based on CNNs have also been used for specific animal image classification, such as species recognition within the *Panthera* genus [10].

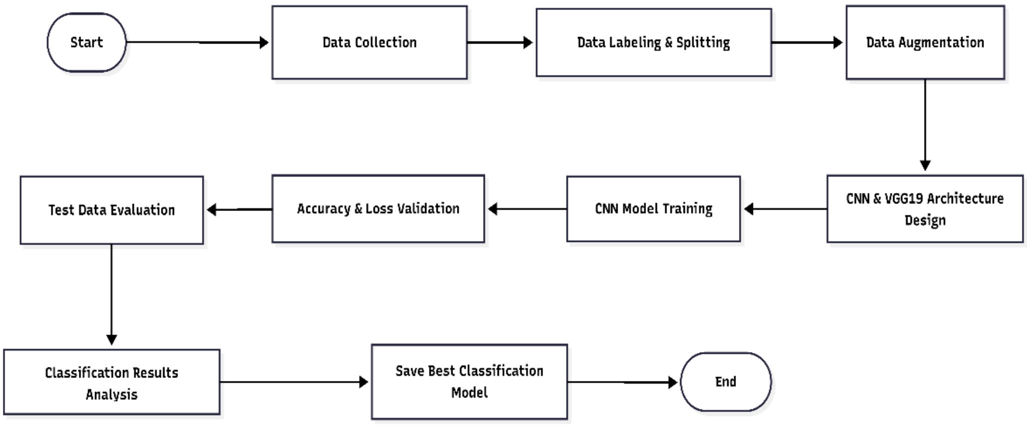
The ability of CNNs to extract complex textural features also includes the identification of physical materials and cultural heritage ornaments. In material analysis, the application of CNN integrated with the K-Means method enables detection and clustering of sand grain images based on their physical characteristics [11]. In the field of cultural arts preservation, CNN implementation is applied as a framework for batik pattern recognition systems, which require high precision regarding geometric details [12]. These previous studies collectively affirm that CNN architectures are effective in handling the classification of complex visual pattern images across various domains. Therefore, the application of CNN methods to classify physical features in images of traditional musical instruments from East Nusa Tenggara is a relevant and appropriate computational approach.

In addition to optimization on X-ray data, the dropout technique in CNN architectures has also proven effective in overcoming overfitting issues for classifying traditional cultural objects with high visual complexity. For example, the implementation of CNN with dropout layers can consistently classify the motifs of Balinese gringsing woven fabrics, achieving a test accuracy of 82% [13]. The characteristics of handling complex shape variations with CNN have also been successfully applied in the realm of ancient manuscript character recognition, such as recognizing handwritten Balinese script, which features unique geometric forms based on individual writing styles where the appropriate learning rate configuration achieves an accuracy of 94.4% [14]. These explorations demonstrate that CNNs are highly adaptive in recognizing detailed visual features in art and cultural representations.

In addition to the domain of cultural preservation, variations of CNN models enhanced with overfitting reduction techniques have also demonstrated solid performance when tested using primary datasets in the field of ophthalmology. For cataract disease classification, CNN models equipped with dropout layers can automatically extract eye image features to address the limitations of conventional medical diagnosis resources. Testing of these models shows high performance stability, with accuracy rates reaching 94% on primary datasets and 92% on public datasets [15]. Another comparative characteristic in chronic disease recognition was evaluated through model depth, such as ResNet-50, which demonstrated superior accuracy consistency compared to DenseNet-121 in classifying diabetic retinopathy severity levels [16]. This range of literature collectively strengthens the theoretical foundation that CNN-based approaches are highly appropriate for identifying and classifying complex physical features in images of traditional musical instruments from East Nusa Tenggara.

Research Method

The research steps are systematically designed in Figure 1 to develop a comprehensive image classification system for local cultural heritage objects of East Nusa Tenggara (NTT). The process begins with dataset collection, labeling and data splitting, digital image preprocessing, data augmentation, the design of two modeling architectures, and continues through static training and model testing analysis for seven classes of traditional musical instruments: Sasando, Moko, Gong, Gendang, Foydoa, Jungga, and Knobe Oh. The data collection technique combines a primary approach—direct visual documentation using an Oppo A74 smartphone camera with 1080p resolution at the NTT Provincial Museum, yielding 14 original photos—and a secondary approach of 596 images sourced from digital repositories and search engines. This results in a total dataset of 610 images, rich in viewpoint, distance, lighting, and background variations.



.

Figure 1. Flowchart of the Proposed Research Stages.

All digital images then undergo supervised learning labeling using the flow_from_directory function with the class_mode='categorical' parameter, converting text labels into binary matrices via One-Hot Encoding to prevent inter-class bias. The dataset is proportionally partitioned at a ratio of 80% (490 images) for training data and 20% (120 images) for validation data to isolate test data from the risk of data leakage. Next, preprocessing involves resizing images to a 224 × 224 pixel square and applying pixel intensity normalization on a 1./255 scale to stabilize gradient calculations during backpropagation. To address sample imbalance, data augmentation techniques are dynamically applied using the ImageDataGenerator function, with parameters for tilt distortion, scale enlargement, horizontal flipping, angle rotation, and variations in lighting intensity.

Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 222, 222, 32)	896
batch_normalization_5 (BatchNormalization)	(None, 222, 222, 32)	128
max_pooling2d_4 (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_5 (Conv2D)	(None, 109, 109, 64)	18,496
batch_normalization_6 (BatchNormalization)	(None, 109, 109, 64)	256
max_pooling2d_5 (MaxPooling2D)	(None, 54, 54, 64)	0
conv2d_6 (Conv2D)	(None, 52, 52, 128)	73,856
batch_normalization_7 (BatchNormalization)	(None, 52, 52, 128)	512
max_pooling2d_6 (MaxPooling2D)	(None, 26, 26, 128)	0
conv2d_7 (Conv2D)	(None, 24, 24, 256)	295,168
batch_normalization_8 (BatchNormalization)	(None, 24, 24, 256)	1,024
max_pooling2d_7 (MaxPooling2D)	(None, 12, 12, 256)	0
flatten_1 (Flatten)	(None, 36864)	0
dense_2 (Dense)	(None, 512)	18,874,880
batch_normalization_9 (BatchNormalization)	(None, 512)	2,048
dropout_1 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 7)	3,591

Total params: 19,270,855 (73.51 MB)
Trainable params: 19,268,871 (73.50 MB)

(a). CNN

Layer (type)	Output Shape	Param #
input_layer_3 (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1,792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36,928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73,856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147,584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295,168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590,080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv4 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv4 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten_3 (Flatten)	(None, 25088)	0
dense_6 (Dense)	(None, 256)	6,422,784
dropout_3 (Dropout)	(None, 256)	0
dense_7 (Dense)	(None, 7)	1,799

Total params: 26,448,967 (100.89 MB)
Trainable params: 15,863,815 (60.52 MB)
Non-trainable params: 10,585,152 (40.38 MB)

(b). VGG-19

Figure 2. CNN & VGG-19 Architecture

The model architecture design stage adopts a comparative study approach between two artificial neural network models CNN and VGG19 models are shown in Figure 2. The first model is a Custom CNN architecture built sequentially from scratch using four main convolutional blocks. These blocks use 3×3 kernel filters with 32, 64, 128, and 256 filters, configured with Rectified Linear Unit (ReLU) activation functions and reinforced by Batch Normalization and 2×2 Max Pooling. In the classification layers of the Custom CNN, features are flattened via a Flatten layer, processed by a Dense layer with 512 ReLU-activated neurons, Batch Normalization, a 0.5 Dropout layer, and concluded with a 7-neuron Softmax Output layer. The second model, for comparison, utilizes the VGG19 architecture via Transfer Learning, freezing the weights of blocks 1 through 4 while unfreezing block 5 to adapt to the morphological patterns of NTT instruments. The original VGG19 classification layers are replaced with a custom top structure consisting of a Flatten layer (25,088 elements), a 256-neuron ReLU-based Dense layer, a 0.3 Dropout layer, and a 7-neuron Softmax Output layer.

3. Result and Discussions

A. Result

The research stages were systematically designed to develop a comprehensive image classification system for local cultural heritage objects from East Nusa Tenggara (NTT). The process began with dataset collection, data labeling and splitting, digital image preprocessing, data augmentation, the design of two modeling architectures, and culminated in the model training phase. The data collection technique combined a primary approach—direct visual documentation at the NTT Provincial Museum using an Oppo A74 smartphone camera with 1080p resolution, yielding 14 original photos—and a secondary source of 596 images from digital repositories, resulting in a total dataset of 610 images rich in angle, distance, and lighting variations. After collection, the data underwent supervised learning-based labeling using the flow_from_directory function with class_mode='categorical', which converted directory labels into binary matrices via One-Hot Encoding. The dataset was then proportionally partitioned at a ratio of 80% (490 images) for training data and 20% (120 images) for validation data to isolate test data from the risk of leakage.

To address sample imbalance and broaden visual variation, data augmentation techniques were dynamically implemented (on-the-fly) during training using the ImageDataGenerator function. The augmentation configurations applied to the model included parameters for shear distortion (shear_range=0.1), scale enlargement (zoom_range=0.1), horizontal flipping (horizontal_flip=True), image rotation (rotation_range=15), and variation in lighting intensity limits (brightness_range=[0.8, 1.2]). By applying these parameters, the system effectively explores morphological variations of objects without increasing physical storage demands, thereby enhancing the generalization ability of the neural network in recognizing features of traditional NTT musical instruments from various viewpoints and diverse lighting conditions. The augmentation technique is illustrated in Figure 3.

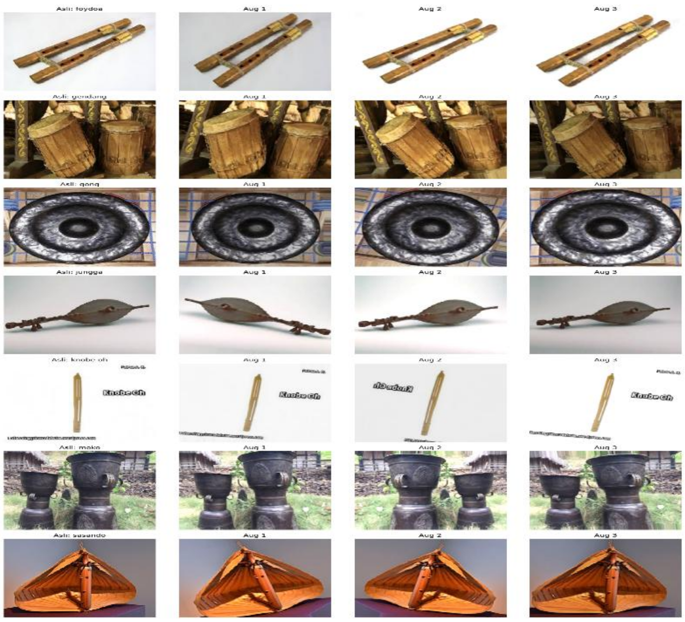


Figure 3. Flowchart of the Proposed Research Stages.

This study implements a comparative approach between a Custom CNN architecture—built sequentially from scratch using four main convolutional blocks with 3×3 kernel filters (32, 64, 128, and 256 filters), ReLU activation functions, as well as Batch Normalization and 2D Max Pooling (2×2). In the classification layers of the Custom CNN, features are flattened via a Flatten layer to be processed by a Dense Layer (512 neurons, ReLU), Dropout (0.5), and an Output Layer with 7 neurons (Softmax). The second model is VGG19 via Transfer Learning, where the weights of blocks 1 through 4 are frozen, while block 5 is unfrozen to better adapt to the morphological patterns of NTT instruments. The original VGG19 classification layers are replaced with a custom top structure consisting of a Flatten layer (25,088 elements), Dense Layer (256 neurons, ReLU), Dropout (0.3), and an Output Layer with 7 neurons (Softmax). All models are trained for 100 epochs to ensure optimal accuracy convergence.

To objectively assess classification performance, experiments were conducted using a static testing method by comparing the performance of the Custom CNN and VGG19 Fine-Tuning architectures across three learning rate variations: 0.001, 0.0001, and 0.00001. These variations were applied to map model sensitivity to weight update speeds, considering the differences between a model built from scratch and a transfer learning model with pre-existing weights. All tests were conducted using the Adam optimizer configuration and a training duration of 100 epochs to ensure consistent evaluation results in every validation scenario. The model training performance results are presented in Table 1.

Table 1. Validation Accuracy and Loss Value.

No.	Architecture Model	Configuration Hyperparameter	Validation Accuracy	Validation Loss
1	Custom CNN	Learning Rate = 0.001	75.83%	2.7289
2	Custom CNN	Learning Rate = 0.0001	82.50%	1.1905
3	Custom CNN	Learning Rate = 0.00001	82.50%	1.2700
4	VGG19 Fine-Tuning	Learning Rate = 0.001	16.67%	1.9420
5	VGG19 Fine-Tuning	Learning Rate = 0.0001	90.83%	0.5854
6	VGG19 Fine-Tuning	Learning Rate = 0.00001	93.33%	0.4211

Based on the data in Table 1, the selection of the learning rate value is proven to play a crucial role in determining the success of convergence for each model architecture. In the Custom CNN model built from scratch, the most optimal configuration is achieved at a learning rate of 0.0001, resulting in an accuracy of

82.50% and the lowest loss at 1.1905. Using a larger learning rate (0.001) yields less optimal performance with an accuracy of 75.83% and a high loss of 2.7289, while further decreasing the value to 0.00001 does not improve accuracy and instead slightly worsens the loss to 1.2700. This indicates that 0.0001 is the best balance point for training stability in the Custom CNN architecture. Conversely, the VGG19 Fine-Tuning architecture shows much more extreme sensitivity to the learning rate value due to the use of transfer learning. When configured with an aggressive learning rate of 0.001, the model experiences total convergence failure, with accuracy dropping to 16.67% due to the corruption of the pre-trained ImageNet weights. However, when the learning rate is conservatively reduced to 0.00001, VGG19 Fine-Tuning achieves the highest peak performance in this study, reaching an accuracy of 93.33% and a minimum loss index of 0.4211. This trend demonstrates that models with complex initial weights require extra caution and smaller learning rates to maintain the integrity of pre-trained features, enabling much better recognition of NTT instruments. The convergence trends of accuracy and model error over 100 epochs are visualized in Figure 4.

Further analysis of the classification performance of both model architectures was conducted using confusion matrices to examine the prediction accuracy for each class of traditional NTT musical instruments. Based on the visualizations, the VGG19 Fine-Tuning model demonstrates significant dominance with a test accuracy of 93.33%, far surpassing the Custom CNN model, which achieved an accuracy of 82.50%. The superiority of the VGG19 model is reflected in the solid concentration of numbers along the main diagonal of the matrix, indicating that the use of pre-trained weights through transfer learning enables far more discriminative extraction of instrument visual characteristics compared to the Custom CNN. In the VGG19 Fine-Tuning confusion matrix, model effectiveness is especially prominent for the Sasando and Knobe Oh instruments, both of which were perfectly classified with 20 and 10 correct predictions, respectively. High performance is also shown in the Gendang and Gong classes, with 19 correct predictions each. Minor error fluctuations were observed only in the Foydoa instrument, with 9 correct predictions and misclassifications to the Jungga (2 samples) and Gong (1 sample) classes, as well as in the Moko class, which received 17 correct predictions with minor errors scattered across other classes. The minimal spread of data outside the main diagonal further affirms the reliability of VGG19 in minimizing visual confusion among instruments.

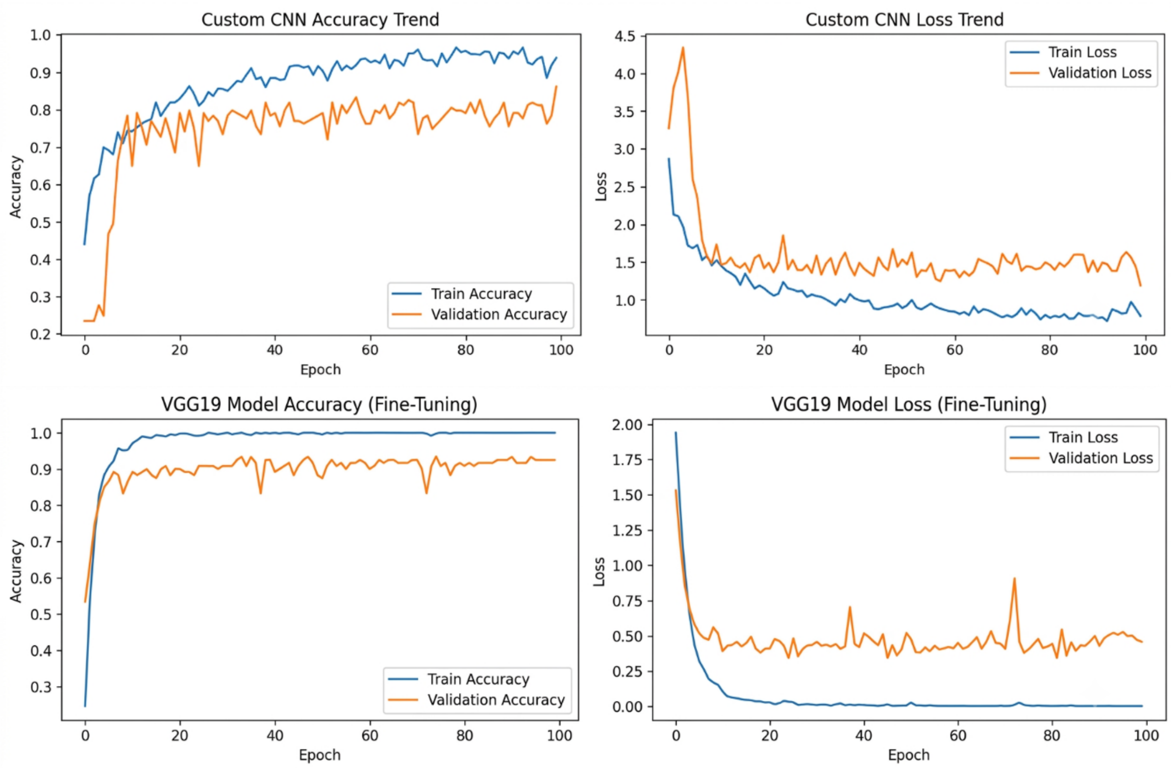


Figure 4. CNN and VGG19 Training Results.

In contrast, the confusion matrix for the Custom CNN shows a higher pattern of misclassification spread outside the main diagonal, indicating overlapping feature extraction. The most significant performance drop occurs in the Foydoa instrument, which was correctly predicted only 6 times, while the remaining samples were misclassified into nearly all other instrument classes. Notable ambiguity is also seen in the Moko class, where the model incorrectly predicted four samples as Gong. The Custom CNN's inability to distinguish specific geometric or textural characteristics of these instruments demonstrates that a from-scratch architecture has limitations in recognizing complex visual feature details, thus requiring deeper representational capacity or a much more diverse training dataset.

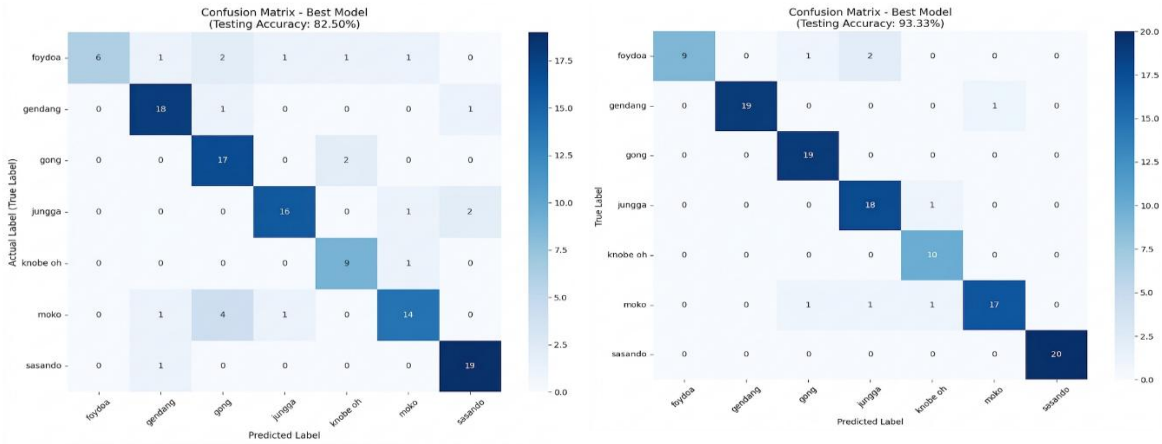


Figure 5. confusion matrix CNN & VGG19.

Evaluation using the Classification Report was also conducted to review each model's ability to recognize the unique geometric details of each cultural instrument, as shown in Figure 6. The comparative assessment of the test performance of both architectures, as shown in the Classification Report, indicates a significant advantage of the VGG19 Fine-Tuning model over the Custom CNN. Overall, the VGG19 Fine-Tuning model achieved a highly satisfactory test accuracy of 93% and a macro-average F1 score of 0.93. The highest recognition was observed in the Sasando class, with a flawless score of 1.00 (100%), followed by excellent detection in Gendang (0.97), Gong (0.95), Knobe Oh (0.91), Jungga (0.90), Moko (0.89), and Foydoa (0.86). In contrast, the Custom CNN architecture produced lower but still solid performance for a model built from scratch, with a total accuracy of 82% and a macro-average F1 score of 0.81. In the Custom CNN model, Sasando remained the class with the highest recognition rate, achieving an F1-Score of 0.90, followed closely by Gendang (0.88), Jungga (0.86), and Knobe Oh (0.82). However, the model still experienced detection sharpness fluctuations and lower scores in Gong (0.79), Moko (0.76), and the lowest in Foydoa (0.67) due to sensitivity issues in recall, which dropped to 0.50. This comprehensive comparison confirms that applying *transfer learning* via VGG19 Fine-Tuning yields a much more stable feature representation and more evenly distributed recognition across all NTT traditional instrument classes compared to the Custom CNN architecture.

CNN CLASSIFICATION REPORT

	precision	recall	f1-score	support
foydoa	1.00	0.50	0.67	12
gendang	0.86	0.90	0.88	20
gong	0.71	0.89	0.79	19
jungga	0.89	0.84	0.86	19
knobe oh	0.75	0.90	0.82	10
moko	0.82	0.70	0.76	20
sasando	0.86	0.95	0.90	20
accuracy			0.82	120
Macro Avg.	0.84	0.81	0.81	120
Weighted Avg.	0.84	0.82	0.82	120

VGG19 CLASSIFICATION REPORT

	precision	recall	f1-score	support
foydoa	1.00	0.75	0.86	12
gendang	1.00	0.95	0.97	20
gong	0.90	1.00	0.95	19
jungga	0.86	0.95	0.90	19
knobe oh	0.83	1.00	0.91	10
moko	0.94	0.85	0.89	20
sasando	1.00	1.00	1.00	20
accuracy			0.93	120
macro avg	0.93	0.93	0.93	120
weighted avg	0.94	0.93	0.93	120

Figure 6. Classification Report CNN & VGG-19

Conversely, the model model_VGG19_0.00001.h5 demonstrates far superior consistency and stability. The test accuracy of the VGG19 model successfully reached 93.33%, exceeding its average training accuracy of 89.29%. With a very small and tight gap of only 4.04%, the performance of the VGG19 model is convincingly categorized as "Stable." This trend comparison confirms that the VGG19 architecture with transfer learning is far more reliable and adaptive than the Custom CNN and has been proven to classify traditional NTT musical instruments on new data more optimally and consistently. The final evaluation results are presented in Table 2.

Table 2 . Final Evaluation Results CNN & VGG19.

No	Model Name	Val Acc	Test Acc	Gap	Status
1	Model_CNN_0.0001.h5	68.76%	82.50%	13.74%	Less Stable
2	model_VGG19_0.00001.h5	89.29%	93.33%	4.04%	Stable

4. Conclusion

This study has demonstrated that the VGG19 Fine-Tuning architecture achieved an accuracy rate of 93.33% in classifying traditional musical instruments of East Nusa Tenggara (NTT). This result significantly outperforms previous studies using standard CNNs, including those by Ferdian Bili & Matheos Sarimole (87%), Prayoga et al. (87.83%), and Rahmadhani & Marpaung (76%). Although the performance in this study surpasses that of the conventional methods, its accuracy is still below that achieved by Rahayu et al. (97.5%) and Munandar & Rozi (99.15%), who utilized more complex model architectures or massive-scale datasets. Therefore, to bridge this performance gap in the future, further research can focus on integrating cutting-edge architectures, such as Vision Transformers or hybrid methods, that offer deeper, more dynamic feature extraction to enhance classification precision for objects with complex visual variations.

Acknowledgment

The author would like to express sincere gratitude to the Institut Bisnis dan Teknologi Indonesia (INSTIKI) Denpasar as the academic institution that provided support and facilities during the implementation of this research. The author also extends heartfelt thanks to Mr. Christoforus Da Santo, Curator of the East Nusa Tenggara Provincial Museum, for granting permission and providing invaluable technical assistance in the direct collection of the primary image dataset. The moral support and valuable contributions of colleagues and all parties involved greatly contributed to the success of developing the experimental classification model for traditional NTT musical instrument images based on Convolutional Neural Networks (CNN).

References

[1] Y. Ferdian Bili and F. Matheos Sarimole, "Pengenalan dan Edukasi Motif Batik Untuk Sekolah Dasar Negeri Pondok Bahar 06 Menggunakan Metode Convolution Neural Network (CNN)," 2025. [Online]. Available: https://journal.stmiki.ac.id

[2] U. Sri Rahmadhani and N. Lysbetti Marpaung, "Klasifikasi Jamur Berdasarkan Genus Dengan Menggunakan Metode CNN," vol. 8, no. 2, 2023.

[3] Ni. L. W. Rahayu, N. Gunantara, and M. Sudarma, "SINTECH Journal | 1," 2023, [Online]. Available: https://doi.org/10.31598

[4] A. Prayoga, Maimunah, P. Sukmasetya, Muhammad Resa Arif Yudianto, and Rofi Abul Hasani, "Arsitektur Convolutional Neural Network untuk Model Klasifikasi Citra Batik Yogyakarta," *Journal of Applied Computer Science and Technology*, vol. 4, no. 2, pp. 82–89, Nov. 2023, doi: 10.52158/jacost.v4i2.486.

[5] A. N. R. Munandar and A. F. Rozi, "Analisis Arsitektur Convolutional Neural Network Untuk Klasifikasi Citra Bunga," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 3, pp. 522–531, Jul. 2024, doi: 10.47233/jteksis.v6i3.1413.

[6] P. A. R. Devi and H. Rosyid, "Pemaparan Materi Dasar Pengolahan Citra Digital untuk Upgrade Wawasan Siswa di SMK Dharma Wanita Gresik," *Jurnal Abdi Masyarakat Indonesia*, vol. 2, no. 4, pp. 1259–1264, Jun. 2022, doi: 10.54082/jamsi.405.

[7] K. Prasetyo and R. Mahenra, "Analisis Kinerja Convolutional Neural Networks Baseline untuk Identifikasi Jenis Jenis Penyakit Kentang," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 2, pp. 609–615, Mar. 2025, doi: 10.57152/malcom.v5i2.1722.

[8] J. Teknik Elektro and G. Wega Intyanto, "Klasifikasi Citra Bunga dengan Menggunakan Deep Learning: CNN (Convolution Neural Network)."

[9] I. Rusman, "Klasifikasi Jenis Rempah-Rempah Menggunakan Algoritma Convolutional Neural Network (CNN)," 2025.

[10] W. Bismi and M. Qomaruddin, "Jurnal Informatika dan Rekayasa Perangkat Lunak Klasifikasi Citra Genus Panthera Menggunakan Pendekatan Deep learning Berbasis Convolutional Neural Network (CNN)," vol. 5, no. 2, pp. 172–179, 2023.

[11] L. febby Olivia, Y. Yuhandri, and S. Arlis, "Penerapan Deep Learning Menggunakan Metode Convolutional Neural Network dan K-Means dalam Klasterisasi Citra Butiran Pasir," *Jurnal KomtekInfo*, pp. 54–62, Mar. 2025, doi: 10.35134/komtekinfo.v12i1.629.

[12] E. Sugiarto, F. Budiman, and A. Fahmi, "Implementation of Deep Learning Based on Convolution Neural Network for Batik Pattern Recognition," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, Jan. 2025, doi: 10.22219/kinetik.v10i1.2019.

[13] I. G. T. Suryawan, M. Sudarma, I. K. G. D. Putra and A. A. K. O. Sudana, "Classifying Ikat Gringsing Woven Cloth Motifs Using Convolutional Neural Network," 2023 7th International Conference on New Media Studies (CONMEDIA), Bali, Indonesia, 2023, pp. 197-202, doi: 10.1109/CONMEDIA60526.2023.10428579.

[14] I. G. T. Suryawan, A. A. N. M. Jaya, I. B. A. I. Iswara, I. P. M. K. Artha and I. K. N. Putra, "Balinese Script Handwriting Recognition Using Convolutional Neural Network,"

2024 IEEE International Symposium on Consumer Technology (ISCT), Kuta, Bali, Indonesia, 2024, pp. 185-190, doi: 10.1109/ISCT62336.2024.10791286.

[15] I. G. Totok Suryawan and I. W. M. A. Putra, "Convolutional Neural Network for Cataract Clasification Using Primary Dataset," 2024 9th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, 2024, pp. 1436-1441, doi: 10.1109/ICBIR61386.2024.10875946.

[16] I. P. G. Y. P. Putra, N. W. J. K. Dewi, P. S. W. Lesmana, I G. T. Suryawan, and P. S. U. Putra, "Comparison of ResNet-50 and DenseNet-121 Architectures in Classifying Diabetic Retinopathy," Indonesian Journal of Data and Science, vol. 6, no. 1, pp. 64-73, Mar. 2025. Available: https://doi.org/10.56705/ijodus.v6i1.232