International Journal of Informatics and Computation (IJICOM)
Vol. 7, No. 1, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Deep Learning for Malay Architectural Identification: A CNN Approach to Heritage Recognition and Preservation

Sri Winiarti[1], Sunardi, Abdul Fadhil

Abstract
This study develops a classification model of traditional Malay buildings using a deep learning approach to analyze the suitability between the design model and real objects. We utilized VGG16 with a dataset of Malay traditional building images to train and test the model. The test results show that the VGG16 model can achieve an accuracy of 98.77% with a learning rate of 0.0001, dropout of 0.20, and epochs of 25. These results indicate that VGG16 is effective as a tool in the process of identifying and preserving traditional architecture based on imagery by producing good accuracy.

Keywords:
Malay Traditional Architecture, Deep Learning, CNN, Heritage Preservation, VGG16 Model

1. Introduction

The preservation of cultural heritage is an essential aspect of maintaining national identity and historical continuity. Traditional buildings represent a society's customs, beliefs, and architectural advancements, reflecting the evolution of construction techniques and cultural values. One of Indonesia's most prominent architectural heritages is the traditional Malay house, which features distinctive design elements adapted to the region's climate, social structure, and cultural traditions. These buildings are characterized by elevated wooden structures, steeply pitched roofs, intricate carvings, and natural ventilation systems, collectively representing a unique architectural style[1].

Malay traditional architecture is deeply rooted in historical and philosophical narratives, with each structural component carrying cultural significance. For example, the shape of the roof often symbolizes the hierarchical social structure, while the house's orientation is typically aligned with environmental and spiritual considerations [2]. Despite their cultural and architectural importance, many traditional Malay buildings face the threat of degradation due to modernization, lack of preservation efforts, and insufficient documentation. Over time, the loss of these structures could lead to a decline in historical knowledge and the disappearance of a valuable part of Indonesia's heritage [3] [4].

The advancement of technology, particularly in artificial intelligence and computer vision, has provided new opportunities for heritage documentation and preservation. Deep learning, a subset of machine learning, has shown remarkable potential in image classification and pattern recognition[5]. Among the various deep learning models available, CNN has emerged as the most effective approach for analyzing visual data. CNNs have been widely utilized in medical imaging, security surveillance, and automated

1. Sri Winiarti, (Universitas Ahmad Dahlan, Indonesia, Corresponding Author: sri.winiarti@tif.uad.ac.id)
2. Sunardi (Universitas Ahmad Dahlan, Indonesia)
3. Abdul Fadhil (Universitas Ahmad Dahlan, Indonesia)

identification systems, proving their ability to extract meaningful features from images accurately[6].

The development of deep learning has brought significant changes in digital image processing, including architecture and cultural heritage preservation [7]. CNN models such as VGG16 have proven effective in classifying and extracting visual features from images and can identify architectural patterns with high accuracy [8]. This capability opens up opportunities for using deep learning to support the preservation of traditional architecture that is starting to be eroded by modernization. In architectural heritage preservation, CNNs offer a promising solution for the automated identification and classification of traditional buildings. By leveraging deep learning techniques, researchers and conservationists can create digital records of architectural styles, allowing for more efficient restoration efforts and historical studies [9]. This approach also helps minimize human subjectivity in architectural analysis, ensuring more reliable and data-driven assessments. Through proper implementation, deep learning can preserve traditional architecture while making historical knowledge more accessible [10].

One of the problems that arise in architecture is the mismatch between the results of digital design and the real form of the building, especially in traditional buildings. This also occurs in traditional Malay buildings with distinctive elements such as curved roofs, jalousie windows, and carved wooden pillars [5]. To answer this problem, an AI-based approach can be used to identify the suitability of design elements of traditional Malay buildings. As a popular method in image classification, CNN offers an accurate and efficient solution for processing visual features of architectural elements. One of the widely used CNN architectures is VGG16, which has a moderate network depth but is quite strong in detecting complex visual features such as contours, textures, and building ornament patterns[11].

This study explores the application of CNNs, specifically the VGG16 model, to classify and identify Malay traditional buildings. The research focuses on recognizing architectural elements such as roofs, windows, and pillars, key distinguishing features of Malay heritage structures. By training a deep learning model with a carefully curated dataset, this study aims to enhance the accuracy of architectural classification, providing a systematic method for identifying and documenting Malay traditional buildings.

2. Related Works

The integration of deep learning in image recognition has emerged as a transformative approach to enhance accuracy and efficiency results [39][40][41][42][43]. Various studies demonstrate the effectiveness of DL models in recognizing diverse heritage forms, from tangible monuments to intangible cultural expressions. The heritage recognition and classification (HRC) model has been created based on the DL strategy. A paper proposed a complete model using various Indian heritage images comprised of 10000 images containing four categories of heritage images namely animals, birds, monuments, and paintings. The complete process outcome resulted in a recognition accuracy of 94.32% using the binary classification technique, whereas 95.43% accuracy has been observed for further multi-classification tasks of heritage category images [35].

Another work explored DL techniques to preserve and understand heritage through comics. Detecting and analyzing important elements like speech bubbles and characters, can uncover the cultural significance of these comics in preserving indigenous heritage. The proposed Faster Region-based CNN (R-CNN) and Bidirectional Encoder Representations from Transformers (BERT) models demonstrate their accuracy and effectiveness in interpreting indigenous intangible heritage through comics [36]. Another paper introduced a deep learning-based ancient literature recognition method for cultural heritage preservation. The model of Single Shot Multibox Detector for detecting and recognizing the ancient characters focuses on the recognition of Oracle Bone Inscriptions' ancient characters. The experimental results show that Precision achieves and Recall achieves 0.86 and 0.97. The paper can prove the effectiveness of Single Shot Multibox Detector in ancient character recognition [37].

The current study presents a hybrid DL CNN and RNN to extract visual features and temporal dependencies from heritage object images, respectively, and demonstrates the potential of DL approaches for the classification of cultural heritage objects. The model utilized 10,000 heritage object images, which were divided into a 70:30 ratio of training and testing. The proposed model achieved an overall accuracy of 91.5% on the testing set, demonstrating its effectiveness for the classification of diverse heritage object classes. The comparative analysis of the model's performance on each heritage object class highlighted that the "Tombs and Mausoleums" class had the highest performance, with a precision score of 94.15%, a recall score of 92.84%, and an F1 score of 93.49% [38].

Several challenges have been identified in applying deep learning to architectural heritage classification. Ensuring deep learning models can generalize well across different variations is crucial for achieving robust classification results. Heritage documentation through deep learning can be combined with augmented reality (AR) and virtual reality (VR) [14] [15]. Another challenge is the computational cost of training complex models of CNN architectures with multiple layers, making them difficult to implement in environments with limited access to high-performance hardware. Many communities attempted to address this issue by optimizing model architectures, reducing parameter complexity, and utilizing cloud-based solutions [16][17].

3. Proposed Method

3.1 Dataset

This study collected high-resolution images of Riau Malay traditional houses in Pekanbaru, Indonesia, and categorized them into three architectural components: roofs, windows, and pillars. Researchers applied preprocessing techniques such as resizing, scaling, and augmentation to enhance the dataset and support robust training. From a total of 750 images, 649 were used for training and 101 for testing, ensuring a balanced evaluation. Using authentic, real-world images allowed the model to learn directly from actual architectural structures. After validating the dataset, the team constructed and trained a CNN model based on the VGG16 architecture to perform classification tasks.

3.2 VGG16 Architecture

This study selected the VGG16 with 3×3 convolutional filters to extract local features critical for identifying unique characteristics of Malay traditional architecture, such as roofs, windows, and pillars. To improve performance, VGG16 can undergo feature extraction and classification of architectural images by tuning hyperparameters, including learning and dropout rates, By training on real-world images, the model aims to support accurate classification that preserves cultural authenticity and informs design decisions for architects and heritage preservation agencies [12] [13].

Fig. 1 systematically illustrates the research flow diagram, including data collection, pre-processing, model training, and evaluation. This diagram presents a comprehensive overview of the methodological framework used to accurately classify components of Malay traditional buildings.

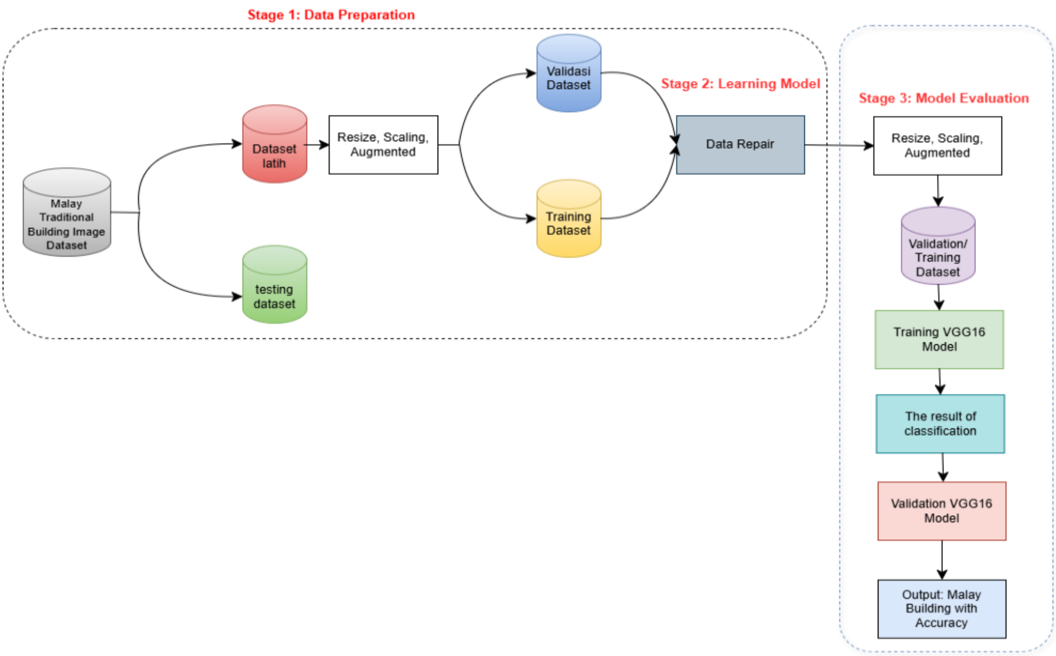


Fig. 1. Research Workflow for Malay Traditional Building Classification Using VGG16

Fig. 1 illustrates the operation of the CNN VGG16 model in classifying images of traditional Malay buildings through four key stages: data pre-processing, model training, feature extraction, and final classification. During the data pre-processing stage, the dataset—comprising hundreds of images obtained from online sources and field documentation—is categorized based on architectural elements. The dataset is then divided into training, validation, and test subsets. Pre-processing procedures include resizing all images to 224×224 pixels to comply with VGG16 input dimensions, normalizing pixel values to the [0, 1] range, and applying data augmentation techniques such as rotation, flipping, and zooming to enhance variability and improve model generalization.

The proposed VGG16 has 13 convolutional layers with 3 fully connected layers. Initially, the model loads pre-trained weights from ImageNet, which are then fine-tuned to adapt to the specific characteristics of Malay traditional building imagery. Within the convolutional layers, the model applies 3×3 filters to capture spatial features such as edges, textures, and structural patterns relevant to architectural components like roofs, windows, and pillars. In VGG16, the convolution layer is responsible for extracting spatial features from the image. The convolution process is expressed mathematically as [18][19]:

$$S(i,j) = (X * W)(i,j) = \sum_{n-0}^{M-1} \sum_{n-0}^{N-1} X(i+m,\ j+n) \,.\, \mathrm{W}(m,n) \ldots .(1)$$

where:

X : input image (e.g., building image),

W : filter/kernel,

S(i,j) : convolution result.

This convolution is performed in layers on VGG16, as many as 13 convolutional layers with a 3x3 kernel. After the convolution results, the Rectified Linear Unit (ReLU) activation function is used to add non-linearity to speed up the training process and prevent the vanishing gradient problem as follows:

$$f(x) = \max(0, x) \dots\dots (2)$$

After convolution and activation, the features are compressed with max pooling [20][21][22]:

$$F_{max} = \{x_i\}_{i=0}^{n} \dots\dots (3)$$

The compressed image is flattened and inserted into a Fully Connected Layer [20]:

$$F_{average}(x) = \frac{1}{N}\sum_{i=1}^{N} x_i \dots\dots (4)$$

For the final classification, the Softmax function is used to calculate the probability of each class [23]:

$$P\ (j \backslash x) = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}} \dots\dots (5)$$

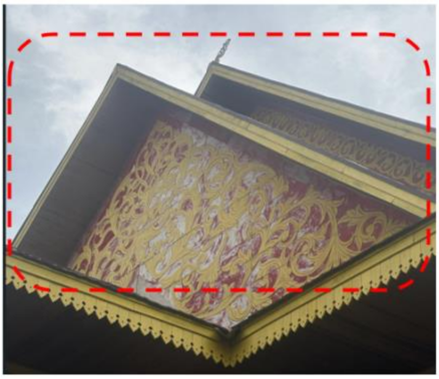
This study uses pooling layers and dropout regularization to improve CNN efficiency and generalization. Pooling reduces spatial dimensions and computational load, helping the model focus on key features while minimizing overfitting. Dropout randomly deactivates neurons during training, encouraging robust pattern learning and reducing reliance on specific pathways. This study also adopts Softmax which can provide a probability distribution by transforming raw outputs into values interpreted as unnormalized probabilities for each class. Mathematically, the softmax function maps a vector z with K components into a probability distribution, ensuring that each value falls between 0 and 1[34].

The training configuration involved experimenting with different parameter combinations to optimize the VGG16 model for classifying images of traditional Malay buildings. Learning rates of 0.00001, 0.0001, and 0.001 were tested alongside dropout rates of 0.20, 0.25, and 0.30 to prevent overfitting. The model was trained for 25 to 30 epochs using the Adam optimizer and a batch size of 32. Eighty percent of the dataset was used for training and the remaining 20% for validation. The training process was conducted on the Google Colab platform with GPU acceleration to enhance computational efficiency and reduce processing time.

4. Experimental Setup

To conduct this study, we collected the dataset directly using digital cameras and smartphones, capturing images of traditional Malay buildings and compiling them into the custom Bangunan Melayu dataset. This dataset contains 750 images, categorized into roofs, pillars, and windows, with 642 images used for training and 108 for testing. A pre-processing phase followed, involving de-noising and resizing images to 256 × 256 pixels to standardize quality and improve model performance. The study then proceeds through a systematic methodology of model development, training, and evaluation using the

VGG16 architecture. This study utilizes performance evaluation includes metrics such as accuracy, precision, recall, and F1-score,




(a) (b)

Fig. 2a. Examples of Roof Structures in Malay Traditional Buildings
Fig. 2b. Examples of Window and Pillar Structures in Malay Traditional Buildings

CNN is a subset of deep learning designed to recognize and is widely applied in object detection, facial recognition, and image classification because it can extract complex features from visual data [24] [25]. A key component of CNN architecture is the convolution layer, which performs convolution operations on the output from the previous layer. CNN has pooling layers that are responsible for refining the extracted feature maps using statistical operations based on pixel values. Pooling layers progressively reduce the dimensionality of feature maps by decreasing the output data volume. Pooling layers enhance the CNN's ability to generalize across different image variations [26][27]. The pooling ensures that essential patterns with reduced spatial resolution [28].

The activation function used in CNN is crucial in introducing non-linearity to the network. The Rectified Linear Unit (ReLU) is the most widely used activation function due to its simplicity and efficiency. ReLU applies the function f(x) = max (0, x), meaning negative input is set to zero, while positive inputs remain unchanged. This operation enhances the network's ability to model complex relationships by enabling the learning of non-linear patterns. ReLU reduces computational complexity and accelerates the training process by preventing negative values from propagating [29][30]. The derivative function of RelU is shown in Equation 2.

$$f' = \begin{cases} 1 \ jika \ x > 0 \\ 0 \ jika \ x \leq 0 \end{cases} \quad \ldots..(6)$$

ReLU can overcome the vanishing gradient problem, which often occurs in other activation functions such as sigmoid and tanh. Vanishing Gradient is a problem that occurs when training deep neural networks, where the gradient of the loss function becomes so small that the weight updates in the early layers of the network become insignificant. As a result, the neural network has difficulty learning in deeper layers [31][32].

Figure 3 illustrates the VGG16 architecture, a widely used CNN for image recognition that processes 224 × 224-pixel inputs through 13 convolutional layers with 3×3 kernels to extract features such as edges and textures. Max pooling layers follow to reduce spatial dimensions and computation while minimizing overfitting. The model then passes features through three fully connected layers for classification, with the final output representing building types like Malay traditional architecture. An additional path labeled "Moccl" appears to extract supplementary features, though its function is not explicitly defined.

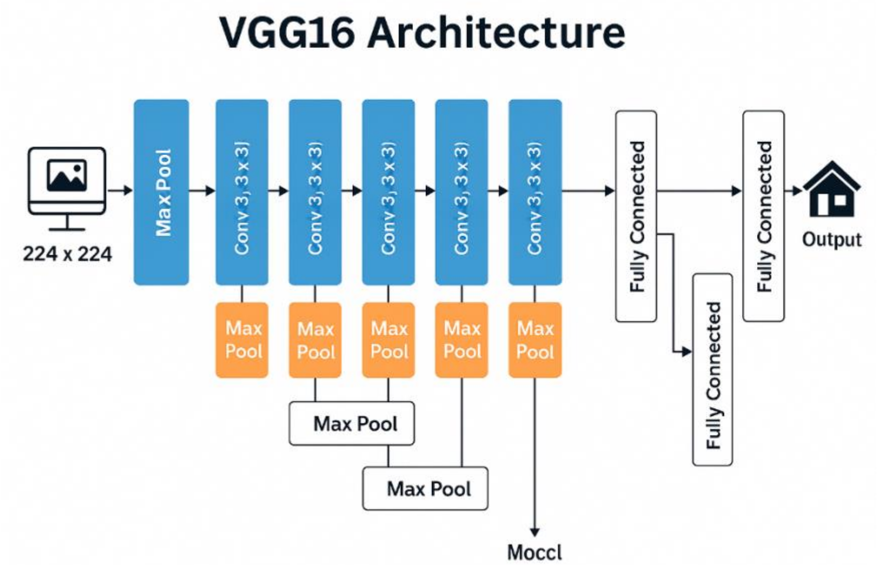


Fig. 3. Architecture of the VGG16 CNN Model

5. Result and Analysis

This study applies VGG16 to classify and assess traditional Malay building images, utilizing its deep architecture—comprising 13 convolutional layers, five max-pooling layers, and three fully connected layers with 3×3 kernel filters—to extract detailed architectural features. The model processes 224×224-pixel RGB images, maintaining spatial integrity while learning complex patterns. The first convolutional layer (Conv-1) uses 64 filters, followed by 128 filters in Conv-2, 256 in Conv-3, and 512 in Conv-4 and Conv-5. This incremental number of filters allows the model to capture low-level and high-level image features, enhancing its ability to distinguish between architectural styles.

VGG16 includes three fully connected (FC) layers, where the first two layers contain 4096 neurons each. The final classification layer consists of 1000 neurons, corresponding to the ILSVRC 1000-class classification with 13 convolutional layers and 3 fully connected layers. The dataset of hundreds of images of traditional Malay buildings from various regions is used for model training and testing. Experimental parameters include an input size of 224x224 pixels, the number of epochs between 25–30, the best learning rate of 0.001, and a dropout rate of 0.20–0.25. The evaluation uses accuracy, training loss, and validation loss metrics. The model achieved the highest validation accuracy of 98.77% at a combination of a learning rate of 0.001 and a dropout of 0.20–0.25, with the lowest training loss of 0.0076. The comparison of accuracy between model configurations shows that the most optimal parameter combination is a learning rate of 0.001 with a dropout of 0.20–0.25, where all configurations with these parameters produce consistently high validation accuracy above 98%. In contrast, when the learning rate is lowered to 0.0001 and 0.00001, accuracy is significantly decreased to 70–80%, especially when combined with a higher dropout rate (0.30–0.35). This indicates that the model is either learning too slowly (underfitting) or losing its generalization ability due to over-regularization.

Table 1 presents the experimental results in which a learning rate of 0.00100 consistently delivers the best performance, achieving a validation accuracy of 98.77% in most test scenarios. This high accuracy is particularly noticeable when the dropout rate is set to 0.20 or 0.25, as lower dropout rates help minimize validation loss. In contrast, a learning rate of

0.00010 results in lower validation accuracy, ranging between 95.37% and 97.23%, with a higher validation loss. This suggests that the model struggles to optimize effectively at this learning rate. The lowest-performing model is observed with a learning rate of 0.00001, where validation accuracy drops significantly to around 68.13% – 82.45%, indicating inefficient learning.

Table 1. Experimental Results of the VGG16 Model with Different Learning Rates, Dropout Rates, and Epochs

No.	Epoch	Learning Rate	Dropout Rate	Training Loss	Validation Loss	Validation Accuracy
1	25	0.00100	0.20	0.0076	0.0213	0.9877
2	25	0.00010	0.20	0.1552	0.1822	0.9537
3	25	0.00001	0.20	0.1088	0.1842	0.9792
4	30	0.00100	0.25	0.0156	0.0163	0.9877
5	30	0.00010	0.25	0.1275	0.2203	0.9723
6	30	0.00001	0.25	0.6293	0.7054	0.8245
7	20	0.00100	0.20	0.0173	0.1041	0.9877
8	20	0.00010	0.25	0.1833	0.2512	0.9607
9	20	0.00001	0.30	0.7695	0.8237	0.6813
10	25	0.00100	0.25	0.0128	0.0485	0.9877
11	25	0.00010	0.30	0.1608	0.2038	0.9630
12	25	0.00001	0.35	0.7685	0.7632	0.7067

When analyzing the effect of dropout rates on model performance, a dropout rate of 0.20 consistently yields the best results, particularly when combined with a high learning rate (0.00100). This combination achieves a validation accuracy of 98.77% while maintaining a low validation loss, as seen in experiment No. 7 (0.1041). While a dropout rate of 0.25 still provides good results, its performance is slightly lower than 0.20 regarding validation accuracy and loss. When using a lower learning rate (0.00001), a dropout rate of 0.25 leads to reduced validation accuracy. However, increasing the dropout rate to 0.30 or higher significantly degrades model performance, particularly at lower learning rates.

Figure 4 visualizes the performance trends further, presenting training, validation, and loss graphs for a clearer analysis. The validation accuracy across different experiments highlights that specific parameter combinations achieve optimal performance. Experiments Nos. 1, 4, 7, and 10 reach the highest validation accuracy (0.9877), mainly when using a high learning rate (0.00100) and a dropout rate of 0.20 or 0.25. In contrast, experiments with lower learning rates (0.00010 or 0.00001) consistently yield lower accuracy values.

Validation loss analysis reveals that models trained with higher learning rates tend to have lower validation loss. The lowest validation loss occurs in experiments using a learning rate of 0.00100, particularly in No. 4 (0.0163) and No. 10 (0.0485). Conversely, models trained with a very low learning rate (0.00001) suffer from high validation loss, as seen in experiments No. 6 (0.7054) and No. 12 (0.7632). These results indicate that models trained with very low learning rates struggle to converge effectively, leading to poor performance.

Training loss patterns further support these findings. Models with lower training loss tend to have lower validation loss, reinforcing the correlation between effective learning and optimal performance. In experiments with a high learning rate (0.00100), training loss remains low, as observed in No. 1 (0.0076) and No. 4 (0.0156). However, experiments such as No. 6 (0.6293) and No. 9 (0.7695) demonstrate high training loss, often correlating with low validation accuracy, indicating that the model failed to learn effectively in certain configurations.

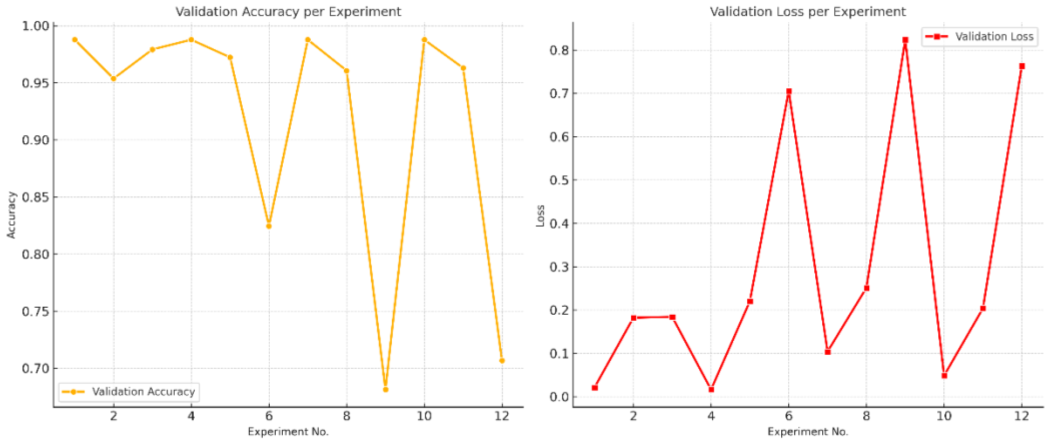


Fig. 4. Training Loss, Validation Loss, and Validation Accuracy Trends Across Experiments

Fig. 4 shows that the best parameter combination in training CNN VGG16 is obtained when using a learning rate of 0.001 and a dropout rate of 0.20–0.25, with high validation accuracy and low loss. Conversely, a learning rate that is too small and a dropout that is too large causes a decrease in model performance. This graph supports the importance of selecting optimal training parameters in CNN-based classification tasks for cultural heritage preservation. The relationship between validation accuracy and learning rate further emphasizes the importance of high learning rates combined with optimal dropout rates. Across all scenarios, higher learning rates (0.00100) consistently yield better validation accuracy. However, dropout rates exceeding 0.30 negatively impact performance, particularly at lower learning rates. This trend is observed in experiments No. 9 and No. 12, where a combination of a low learning rate (0.00001) and a high dropout rate (0.30 or more) significantly reduces accuracy.

Based on these findings, the optimal parameter combination for the VGG16 model is a learning rate of 0.00100, a dropout rate of 0.20 or 0.25, and an epoch count of 25 or 30. This configuration consistently delivers the highest validation accuracy while maintaining low validation loss, making it the most effective setup for classifying Malay traditional buildings using deep learning.

The accuracy analysis using a confusion matrix demonstrates the performance of the VGG16 model, which was evaluated through training, testing, and validation of Malay traditional building images. The best confusion matrix result achieved an accuracy of 98.77%, obtained using a learning rate of 0.0001, 25 epochs, and a dropout rate of 0.20. Avoiding excessively small learning rates, such as 0.00001, was crucial, especially when combined with a high dropout rate (0.30 or more), as this significantly degraded model performance.

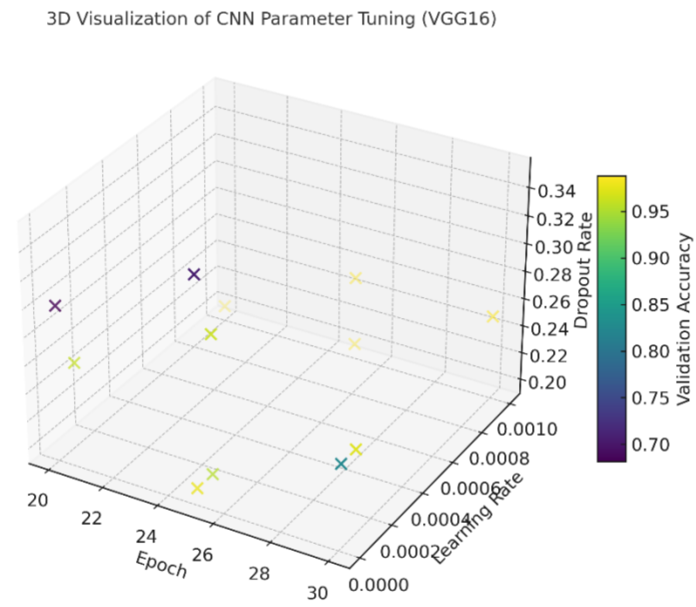


Fig. 5. 3D Visualization of Learning Rate Dropout rate, and validation Accuracy

Fig. 5 shows the relationship between three main CNN training parameters—the number of epochs, learning rate, and dropout rate—towards model performance in the form of validation accuracy. The colors on the graph indicate the accuracy level, with greenish-yellow representing higher accuracy. It can be seen that the combination of a learning rate of 0.001, a dropout rate of 0.20–0.25, and 25–30 epochs yield the highest validation accuracy of 98.77%. In contrast, the accuracy decreases significantly in configurations with very low learning rates (0.00001) and high dropout (≥0.30), indicating over-regularization and slow convergence. This visualization reinforces the importance of optimal parameter selection in training CNN models for architectural heritage classification tasks, especially in the context of the digital preservation of traditional Malay buildings.

Figure 6 visualizes the model's performance, showcasing its evaluation through different training and validation stages.

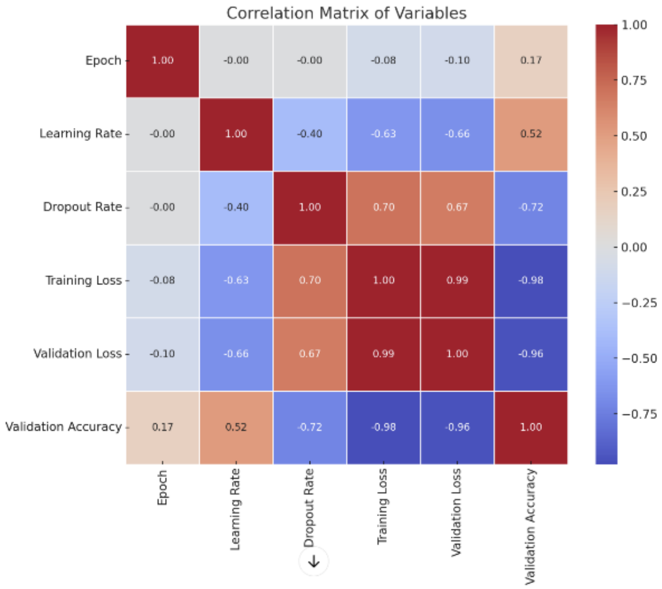


Fig. 6. Confusion Matrix and Correlation Analysis of Model Performance Metrics

Fig. 6 also presents the correlation matrix between key variables analyzed in Table 1, highlighting the relationships among validation accuracy, training loss, and validation loss. The correlation between validation accuracy and validation loss is strongly negative, meaning that as validation loss decreases, validation accuracy tends to increase, indicating better generalization of the model.

The correlation graph shows the relationship between the main variables in training the CNN VGG16 model to classify Malay traditional buildings. The analysis results reveal that the dropout rate has a strong negative correlation with the validation accuracy (-0.72), indicating that a dropout rate that is too high can reduce model performance. In addition, there is a very strong negative correlation between the training loss (-0.98) and validation loss (-0.96) with the validation accuracy, indicating that the smaller the loss value, the higher the model accuracy. Meanwhile, the learning rate shows a moderate positive correlation with accuracy (0.52), indicating that choosing the right learning rate can improve model performance. The almost perfect correlation between the training loss and validation loss (0.99) indicates the stability of the model during the training process.

The relationship between learning rate and validation accuracy shows a moderate positive correlation, meaning that a higher learning rate generally results in better validation accuracy. However, if the learning rate is too high, it may cause instability in training, whereas excessively low values slow down convergence and can lead to suboptimal results. The dropout rate and validation accuracy exhibit a weak negative correlation, implying that higher dropout rates slightly reduce validation accuracy. While dropout is essential for preventing overfitting, excessive dropout may hinder the model's ability to retain critical information during training, affecting its overall performance.

Experimental results demonstrate the effectiveness of the VGG16 model in identifying Malay traditional houses with a high level of precision. The model achieved an accuracy of 98.77%, indicating its ability to distinguish different architectural elements with minimal error. By integrating VGG16 with digital image analysis, the study can automate the recognition of architectural features and contribute significantly to the documentation and preservation of cultural heritage.

6. Conclusion

This study explored VGG16-CNN for identifying and classifying Malay traditional buildings by analyzing key architectural components such as roofs, windows, and pillars. The experimental results demonstrated the high accuracy of VGG16, achieving an optimal validation accuracy of 98.77%, confirming its reliability in recognizing traditional architectural elements. The evaluation of different hyperparameters with a learning rate of 0.00100 combined with a dropout rate of 0.20 or 0.25 consistently yielded the best results, reducing validation loss while maintaining high accuracy. In contrast, excessively low learning rates (0.00001) and high dropout rates (0.30 or more) led to a notable decline in performance. These findings emphasize the importance of careful parameter selection to enhance the robustness and generalization for classification. VGG16 has been proven effective in detecting and classifying traditional architectural elements. This approach is a modern solution for accurate, efficient, and reliable technology-based cultural preservation for developing architectural heritage documentation and conservation systems.

Future research should focus on expanding the dataset to include more diverse Malay traditional buildings and exploring alternative CNN architectures such as ResNet, InceptionV3, or MobileNet. Additionally, integrating real-time detection, augmented reality (AR), and virtual reality (VR) applications could enhance the practicality and accessibility of deep learning in architectural heritage preservation.

Acknowledgment

The authors would like to express their deepest gratitude to Dr. Murinto, Head of the Informatics Department at Universitas Ahmad Dahlan, for his invaluable support and encouragement throughout this research. Special appreciation is also extended to Heri Pramono, S.T., M.Arch., for his insightful guidance in architectural aspects related to this study. The authors are also sincerely grateful to Prof. Anton Yudhana, Head of the Institute for Research and Community Service at Universitas Ahmad Dahlan, for his continuous support in facilitating this research. Lastly, thanks to Aris Rakhmadi, a fellow doctoral colleague at DIFA, whose motivation and encouragement played a significant role in completing this work.

References

[1] R. Saputra, “Governance Frameworks and Cultural Preservation in Indonesia: Balancing Policy and Heritage,” *J. Ethn. Cult. Stud.*, vol. 11, no. 3, pp. 25–50, 2024.

[2] I. Ibrahim, N. Roslee, A. Abdullah, M. Ibrahim, and F. F. N. Murtaza, “Analysis on the socio-cultural values of the traditional Malay houses compound,” *Plan. Malaysia*, vol. 18, no. 2, pp. 109–120, 2020, doi: 10.21837/pm.v18i12.747.

[3] H. Husain And H. Tamjehi, “Preservation Of Malay Study: Selected Case Study Via Google Scholar,” *J. Pengaj. MELAYU*, vol. 35, no. 2, pp. 18–30, 2024.

[4] S. Alsheikh Mahmoud, H. Bin Hashim, M. F. Shamsudin, and H. Alsheikh Mahmoud, “Effective Preservation of Traditional Malay Houses: A Review of Current Practices and Challenges,” *Sustain.* , vol. 16, no. 11, p. 4773, 2024, doi: 10.3390/su16114773.

[5] Mario Casillo; Francesco Colace; Brij B. Gupta; Angelo Lorusso; Francesco Marongiu; Domenico Santaniello, “A Deep Learning Approach to Protecting Cultural Heritage Buildings Through IoT-Based Systems,” in *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2022.

[6] Y. A. and A. B. Yehia Miky1, 2*, “Using deep learning for enrichment of heritage BIM: Al Radwan

house in historic Jeddah as a case study," *Herit. Sci.*, vol. 12, p. 255, 2024.
[7] F. Monna *et al.*, "Deep learning to detect built cultural heritage from satellite imagery. - Spatial distribution and size of vernacular houses in Sumba, Indonesia," *J. Cult. Herit.*, vol. 52, pp. 171–183, 2021.
[8] C. H. Praharsha and A. Poulose, "CBAM VGG16: An efficient driver distraction classification using CBAM embedded VGG16 architecture," *Comput. Biol. Med.*, vol. 180, p. 108945, 2024.
[9] J. Llamas, P. M. Lerones, R. Medina, E. Zalama, and J. Gómez-García-Bermejo, "Classification of architectural heritage images using deep learning techniques," *Appl. Sci.*, vol. 7, no. 10, p. 992, 2017, doi: 10.3390/app7100992.
[10] Ece Cantemir, Ozlem Kandemir, "Use of artificial neural networks in architecture: determining the architectural style of a building with convolutional neural networks," *Neural Comput. Appl.*, vol. 36, pp. 6195–6207, 2024.
[11] H. Yazdi, S. Sad Berenji, F. Ludwig, and S. Moazen, "Deep Learning in Historical Architecture Remote Sensing: Automated Historical Courtyard House Recognition in Yazd, Iran," *Heritage*, vol. 5, no. 4, pp. 3066–3080, 2022, doi: 10.3390/heritage5040159.
[12] M. Hussain, T. Thaher, M. B. Almourad, and M. Mafarja, "Optimizing VGG16 deep learning model with enhanced hunger games search for logo classification," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-82022-5.
[13] Z. P. Jiang, Y. Y. Liu, Z. E. Shao, and K. W. Huang, "An improved VGG16 model for pneumonia image classification," *Appl. Sci.*, vol. 11, no. 23, 2021, doi: 10.3390/app112311185.
[14] J. Zhang, W. A. J. Wan Yahaya, and M. Sanmugam, "The Impact of Immersive Technologies on Cultural Heritage: A Bibliometric Study of VR, AR, and MR Applications," *Sustain.*, vol. 16, no. 15, p. 6466, 2024, doi: 10.3390/su16156446.
[15] S. Münster *et al.*, "Artificial Intelligence for Digital Heritage Innovation: Setting up an R&D Agenda for Europe," *Heritage*, vol. 7, no. 2, pp. 794–816, 2024, doi: 10.3390/heritage7020038.
[16] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00444-8.
[17] H. Li and H. Dong, "Architectural Style Classification Based on Deep Learning," *Comput. Aided. Des. Appl.*, vol. 22, pp. 16–31, 2025, doi: 10.14733/cadaps.2025.S1.16-31.
[18] Moez Krichen, "Convolutional Neural Networks: A Survey," *Computers*, vol. 12, p. 151, 2023.
[19] A. S. Khan, Asifullah, U. Zahoora, and A. S. Qureshi, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks," *Artif. Intell. Rev.*, vol. 53, pp. 5455–5516, 2020.
[20] Lei Zhao & Zhonglin Zhang, "An improved pooling method for convolutional neural networks," *Sci. Rep.*, vol. 14, p. 1589, 2024.
[21] W. Zhao, J. L. B, T. Fana, Y. Renc, and Yukun Xia, "Analyzing and visualizing scientific research collaboration network with core node evaluation and community detection based on network embedding," *Pattern Recognit. Lett.*, vol. 144, pp. 54–60, 2021.
[22] H. H. Yongliang Zhang, "Adaptive complex-valued stepsize based fast learning of complex-valued neural networks," *Neural Networks*, vol. 124, pp. 233–242, 2020.
[23] M. D. Xia Zhao, Limin Wang, Yufei Zhang, Xuming Han, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif. Intell. Rev.*, vol. 57, no. 99, 2024.
[24] D. S. Utpal Barman, Ridip Dev Choudhury, and G. G. Barman, "Computers and electronics in agriculture Book Review," *Stat. Anal. Reg. Yield Trials*, vol. 1699, no. 94, pp. 0–1, 1995.
[25] R. Garg, G. Singh, A. Singh, and M. P. Singh, "Fingerprint recognition using convolution neural network with inversion and augmented techniques," *Syst. Soft Comput.*, vol. 6, p. 200106, 2024, doi: 10.1016/j.sasc.2024.200106.
[26] R. Nirthika, S. Manivannan, A. Ramanan, and R. Wang, "Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study," *Neural Comput. Appl.*, vol. 34, no. 7, pp. 5321–5347, 2022, doi: 10.1007/s00521-022-06953-8.
[27] A. Azulay and Y. Weiss, "Why do deep convolutional networks generalize so poorly to small image transformations?," *J. Mach. Learn. Res.*, vol. 20, pp. 1–25, 2019.
[28] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Mach. Learn. with Appl.*, vol. 6, p. 100134, 2021, doi: 10.1016/j.mlwa.2021.100134.
[29] D. Yang, K. M. Ngoc, I. Shin, and M. Hwang, "DPReLU: Dynamic Parametric Rectified Linear Unit and Its Proper Weight Initialization Method," *Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, 2023, doi: 10.1007/s44196-023-00186-w.
[30] R. Parhi and R. D. Nowak, "The Role of Neural Network Activation Functions," *IEEE Signal Process. Lett.*, vol. 27, pp. 1779–1783, 2020, doi: 10.1109/LSP.2020.3027517.

[31] W. S. W. Ahmad and M. N. Jambli, “Improving a Deep Neural Network Generative-Based Chatbot Model,” *ASEAN Eng. J.*, vol. 14, no. 2, pp. 45–52, 2024, doi: 10.11113/aej.v14.20663.

[32] G. Airlangga1, “A Comparative Analysis of Deep Learning Models for SMS Spam Detection: CNN-LSTM, CNN-GRU, and ResNet Approaches,” *J. Comput. Networks, Archit. Comput. High Perform.*, vol. 6, no. 4, pp. 1952–1960, 2024.

[33] K. A. Rashedi, M. T. Ismail, S. Al Wadi, A. Serroukh, T. S. Alshammari, and J. J. Jaber, “Multi-Layer Perceptron-Based Classification with Application to Outlier Detection in Saudi Arabia Stock Returns,” *J. Risk Finance. Manag.*, vol. 17, no. 2, 2024, doi: 10.3390/jrfm17020069.

[34] T. T. Le, H. T. Duong, and H. C. Phan, “Optimization of Neural Network architecture and derivation of closed-form equation to predict ultimate load of functionally graded material plate,” *Adv. Mech. Eng.*, vol. 15, no. 5, pp. 1–17, 2023, doi: 10.1177/16878132231175002

[35] V. K. Kukreja, R. Sharma, and D. Bordoloi, "Application of deep learning strategy for multi-classification of Indian heritage images," in *Proc. IEEE Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, 2023, pp. 1-5, doi: 10.1109/icccnt56998.2023.10307029.

[36] R. Sharma and V. K. Kukreja, "Deep learning-based comic recognition and analysis for the preservation of indigenous intangible heritage," in *Proc. IEEE Int. Conf. Smart Humanit. (ICSH)*, 2023, pp. 136-140, doi: 10.1109/icsh57060.2023.10482818.

[37] L. Meng, N. Kamitoku, X. Kong, and K. Yamazaki, "Deep learning based ancient literature recognition and preservation," in *Proc. Soc. Instrum. Control Eng. (SICE)*, 2019, pp. 473-476, doi: 10.23919/SICE.2019.8860070.

[38] R. Jain, V. K. Kukreja, R. Sharma, and A. Verma, "New perspectives on heritage: A deep learning approach to heritage object classification," in *Proc. IEEE World Conf. Comput. (WCONF)*, 2023, pp. 1-6, doi: 10.1109/wconf58270.2023.10235213.

[39] Lengga S. Sandy, Sarjon Defit, and Gunadi Widi N., “Identifying Corn Leaf Diseases Using CNN Algorithm”, *ijicom*, vol. 7, no. 1, pp. 99–107, Mar. 2025.

[40] R. L. Satria Putra and M. H. Wathan, “Shallots Classification using CNN”, *ijicom*, vol. 3, no. 1, pp. 40–51, Feb. 2022.

[41] R. L. Satria and M. H. Wathan, “ConFruit: Effective Fruit Classification Using CNN Algorithm”, *ijicom*, vol. 5, no. 1, pp. 10–18, Aug. 2023.

[42] M. Diqi and S. H. Mulyani, “Implementation of CNN for Plant Leaf Classification”, *ijicom*, vol. 2, no. 2, pp. 1–10, Mar. 2021.

[43] A. Abidin, Hamzah, and M. Endah Hiswati, “Efficient Fruits Classification Using Convolutional Neural Network”, *ijicom*, vol. 3, no. 1, pp. 1–9, Oct. 2021.