Spice Image Classification Using ResNet50 and Augmentation Technique

1*Julio Francisco Bacun, 2Irfan Pratama
1,2Department of Information System, Mercu Buana University of Yogyakarta, Indonesia
Email: 1211210028@student.mercubuana-yogya.ac.id, 2irfanp@mercubuana-yogya.ac.id

Article Info

Article history:
Received Jul 02nd, 2025
Revised Oct 06th, 2025
Accepted Nov 27th, 2025

Keyword:
Convolutional Neural Network
Deep Learning
ResNet50
Spice Classification
Transfer Learning

ABSTRACT

This research aimed to develop an automatic classification system for Indonesian spices using a deep learning approach based on the ResNet50 architecture. The classification task involved 31 spice categories, each with 210 images. Two training strategies were implemented: training the model from scratch and using transfer learning with pre-trained weights from ImageNet. The model trained from scratch achieved a validation accuracy of 57%, while the transfer learning approach combined with fine-tuning of the last 33 layers resulted in a significantly higher validation accuracy of 96%. Image preprocessing, data augmentation, and class weighting were applied to improve the model's generalization and handle data imbalance. The confusion matrix analysis showed that most predictions aligned with the true labels, especially in the transfer learning model. These findings demonstrate that transfer learning with ResNet-50 can effectively classify spice images with high accuracy, even when visual similarity exists between certain classes. This research highlights the potential of deep convolutional neural networks to support automatic and reliable identification systems for biodiversity mapping and agricultural industries.

Corresponding Author:
Julio Francisco Bacun,
Department of Information Systems, Mercu Buana University Yogyakarta,
10 Wates Road, Argomulyo Village, Sedayu Subdistrict,
Bantul Regency 55753, Special Region of Yogyakarta, Indonesia.
Email: julio.bacun12@gmail.com

DOI: http://dx.doi.org/10.24014/ijaidm.v8i3.37862

1. INTRODUCTION

Spices are an essential part of Indonesia's biological wealth and have long served as cultural, economic, and traditional medicinal assets. In addition to their use as culinary flavorings, spices are widely applied in the cosmetics and pharmaceutical industries. The high diversity of spices both in terms of shape, color, and texture positions Indonesia as one of the world's largest spice producers [1]. This potential must be supported by accurate identification systems to maximize the economic value of spices across various sectors.

However, manually identifying spice types remains a considerable challenge. The visual similarities among different types of spices often lead to identification errors [2]. Conventional methods such as visual inspection and reference-based matching are subjective, slow, and inefficient when applied at an industrial scale or in biodiversity mapping efforts. Misclassification can negatively affect product quality, food safety, and even open opportunities for raw material adulteration.

To address these challenges, an intelligent classification system is required that can automatically and accurately identify spice types. One of the most effective approaches in visual data processing is the use of deep learning technology, particularly Convolutional Neural Networks (CNN). CNNs are designed to process two-dimensional data such as digital images, by mimicking the human visual system in recognizing visual patterns. Among the various CNN architectures, ResNet50 is known for its strong performance due to its use of residual learning, allowing for the training of very deep networks without loss of accuracy [3].

Several previous studies have explored the classification of spice images using CNN methods. A study by Juli & Timur [4], implemented a CNN architecture with two convolutional layers to classify three types of spices: ginseng, ginger, and galangal. Using 300 images and an 80:20 train-test split, the model employed 10 and 20 convolutional channels, 3×3 pooling layers, and used the tanh and softmax activation functions. It achieved a classification accuracy of 88.89%. Meanwhile, a study by Putra et al. [5], compared six popular CNN architectures Xception, MobileNetV2, DenseNet201, VGG16, VGG19, and ResNet50 in classifying ten types of spices. Among these, the Xception model achieved the best performance with an F1-score of 96.99%, indicating superior effectiveness compared to the other models. In another study conducted by Nisa & Candra [6], classification was performed on four types of rhizomes (ginger, turmeric, aromatic ginger, and galangal) using 1,000 images, evenly distributed across all classes. The applied CNN consisted of four convolutional layers and was tested under various training configurations. The best experiment achieved 90% accuracy, and the model was integrated into a web-based platform as a spice classification application. Maulana et al. [7], employed a CNN based on the VGG16 architecture to identify 31 types of spices, using a dataset of 6,510 images (210 images per class). The classification system achieved a peak accuracy of 86.66% after training for 50 epochs. This study also utilized data augmentation techniques to improve the model's robustness in handling visual variations across spice classes. Collectively, these four studies demonstrate that CNN-based approaches are highly effective for classifying spice images, regardless of dataset scale. They also highlight the significant potential for developing practical and efficient automated systems for identifying spices.

Based on the aforementioned background, this study aims to develop an automatic classification system for various Indonesian spices using the CNN approach, specifically employing the ResNet50 architecture. Two strategies are implemented: training the model from scratch and applying *transfer learning* with pre-trained weights from ImageNet. The dataset consists of diverse images per class that represent the natural variations in spice shape, color, and texture. To enhance the model’s generalization capability, systematic image augmentation techniques are applied. It is expected that the results of this study will contribute to the development of accurate, efficient, and practical digital classification systems particularly to support the spice industry and biodiversity conservation in Indonesia.

2. RESEARCH METHOD

This research was conducted using a cloud-based programming platform, namely Google Colaboratory (Colab). The choice of Colab was based on its ability to provide free access to GPUs, which is highly beneficial for accelerating the training process of deep learning models that require high computational power. In addition, its direct integration with Google Drive facilitates efficient dataset management, model storage, and evaluation result documentation without reliance on local hardware. The complete research methodology workflow is illustrated in Figure 1.

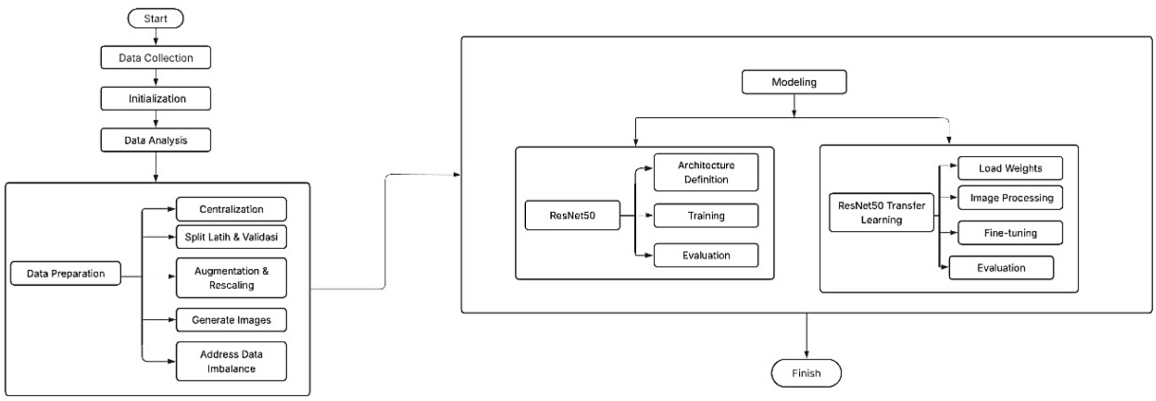

Figure 1. Research Methodology

2.1. Data Collection

This study utilized a dataset obtained from the Kaggle platform, titled 31 Spices Dataset. The dataset consists of 31 classes of spices commonly used in Indonesian cuisine, such as fennel, andaliman, tamarind, onion, shallot, garlic, coriander seeds, star anise, clove, kaffir lime leaves, basil leaves, coriander leaves, bay leaves, ginger, cumin, cardamom, cinnamon, sappanwood, candlenut, cubeb, lesser galangal, kluwek, turmeric, black pepper, galangal, nutmeg, saffron, lemongrass, vanilla, and sesame. Each class contains 210 images.

2.2. Initialization

The initial setup was carried out in the Google Colaboratory programming environment, which provides GPU support to accelerate the model training process. Library initialization included TensorFlow, Keras, NumPy, Pandas, Matplotlib, and Scikit-learn to support data modeling and visualization. The image augmentation process was facilitated by the ImageDataGenerator from Keras, while data management and analysis were handled using Pandas and NumPy.

2.3. Dataset Analysis

Dataset analysis was conducted by traversing each spice class directory, counting the number of images per class, and verifying the consistency of image dimensions. This step is essential to maintain balanced data distribution and consistent image sizes two critical factors for stable training and accurate model performance, especially since CNN models are highly sensitive to imbalanced data distribution and non-uniform image inputs.

2.4. Data Preparation

Data preparation is a crucial stage in deep learning-based image classification, ensuring that the data is prepared efficiently and consistently for use. This stage begins with centralization, which involves organizing the images into a uniform folder structure based on class labels to facilitate automated processing [8]. The dataset is then split into training and validation sets (train–validation split), typically using a ratio such as 80:20, to prevent overfitting and to assess the model's generalization performance [9].

To make the model more robust against real-world data variation, image augmentation techniques such as rotation, flipping, zooming, and blurring are applied, along with normalization of image size and pixel scale (rescaling) 13. In cases where the amount of data is limited, the ImageDataGenerator from Keras is used to generate image batches in real time at a resolution of 384×384 pixels, while applying one-hot encoding on the labels for memory efficiency and faster training [10].

To address dataset imbalance, class weights are applied to assign higher importance to minority classes. The weights are calculated inversely to the class frequency using the compute_class_weight function from Scikit-learn [11].

2.5. Modeling

2.5.1. ResNet50 from Scratch

In this study, the ResNet50 architecture was developed through a modified approach using the Keras library, without relying on the built-in architecture. The network was constructed by utilizing several key components, such as Conv2D layers to extract visual features from input images, BatchNormalization to maintain the stability of activation distributions during training, and the ReLU activation function to introduce the necessary non-linearity [12][13]. Prior to the classification stage, the model was equipped with a GlobalAveragePooling2D layer, which serves to condense spatial features into a global representation [14].

The model was trained entirely from scratch using a preprocessed image dataset of spices. The optimization process employed the Adam algorithm, known for its efficiency in adaptively adjusting weight updates, and used categorical crossentropy as the loss function, considering the task involved multiclass classification [15]. To evaluate the model's performance, validation testing was conducted using several metrics, including classification accuracy, a confusion matrix to review the distribution of predicted results across classes, and a classification report encompassing precision, recall, and F1-score values [16]. This evaluation aimed to assess how well the model could generalize to previously unseen data.

2.5.2. ResNet50 Transfer Learning

In the transfer learning approach applied in this study, the initial weights of the ResNet50 model, which had been previously trained on the ImageNet dataset, were reused [17]. The model's original output layer was removed to allow for the adjustment of the network structure according to the number of categories in the spice dataset used. Before being processed by the model, all images were resized to a fixed dimension of 384×384 pixels. In addition, normalization was performed using the preprocess_input function from Keras. An applications library, which helps align the input format with that of the pretrained model.

The next step was fine-tuning, which involved unfreezing several top layers of the network to allow them to be retrained using the new data. This adjustment enables the model to recognize more dataset-specific visual patterns [18]. A fully connected (Dense) layer was added at the end of the network to perform classification according to the number of spice classes in the dataset [19]. The modified model was then evaluated using the same metrics as in the training-from-scratch approach accuracy, confusion matrix, and classification report. This evaluation aimed to assess the effectiveness of transfer learning in enhancing classification performance compared to full training from scratch [20].

3. RESULTS AND ANALYSIS

3.1. Data Collection and Sample Visualization

The dataset used in this study consists of 6,510 images categorized into 31 spice classes, with each class containing 210 images. The images were organized in a directory-based structure and loaded using the flow_from_directory method. All images were resized to 384×384 pixels to match the input size required by the ResNet50 model. To illustrate the visual diversity of the dataset, an example of an image of the spice category is shown in Figure 2.

Figure 2. Sample spice images

3.2. Augmented Data

To improve the model's generalization capability and minimize the risk of overfitting, two distinct data augmentation strategies were applied to the training images. The first strategy was employed during the training of the model from scratch, while the second was applied in the Transfer Learning approach using the ResNet50 architecture, which incorporated a specific preprocessing function. The configurations of the augmentation techniques used in each scenario are presented in Tables 1 and 2.

Table 1. Data Augmentation Configuration for ResNet50 (From Scratch)

Augmentation Type	Parameter
rotation_range	45
width_shift_range	0.15
height_shift_range	0.15
zoom_range	0.15
horizontal_flip	True
vertical_flip	True
shear_range	0.05
brightness_range	0.9, 1.1
channel_shift_range	10
fill_mode	nearest
preprocessing Function	none

Table 2. Data Augmentation Configuration for ResNet50 (Transfer Learning)

Augmentation Type	Parameter
rotation_range	60
width_shift_range	0.15
height_shift_range	0.15
zoom_range	0.20
horizontal_flip	True
vertical_flip	True
shear_range	0.05
brightness_range	0.9, 1.1
channel_shift_range	10
fill_mode	nearest
preprocessing Function	(ResNet50)

These augmentation strategies helped simulate realistic variations in lighting, orientation, and scale, thereby increasing the diversity of the training data without manual collection. The preprocessing function in the Transfer Learning scenario ensures that the input images are standardized according to the expectations of the pre-trained ResNet50 weights.

3.3. Evaluation of the ResNet50 Model Trained from Scratch

The ResNet50 model was trained from scratch for 25 epochs using the Adam optimizer and the categorical cross-entropy loss function. The training results of the ResNet50 model are presented in Figure 3.

The figure illustrates the evolution of loss and accuracy during the training process of the CNN model. The left graph shows a rapid decrease in training loss toward near-zero values, while the validation loss initially spikes but quickly drops and stabilizes, indicating an effective learning process after a few epochs. The right graph demonstrates a consistent increase in training accuracy, reaching over 65%, whereas validation accuracy fluctuated but tended to increase, reaching around 58%. This pattern indicates that the model learned well from the training data but still faced challenges in generalizing to the validation data, likely due to class imbalance or data complexity.

Figure 3. Loss Evolution and Accuracy Evolution Graph of ResNet50

3.3.1. Confusion Matrix

The confusion matrix analysis was conducted to examine in detail the distribution of correct predictions (true positives) and incorrect predictions (false positives and false negatives) for each class. The confusion matrix of the ResNet50 model is shown in Figure 4.

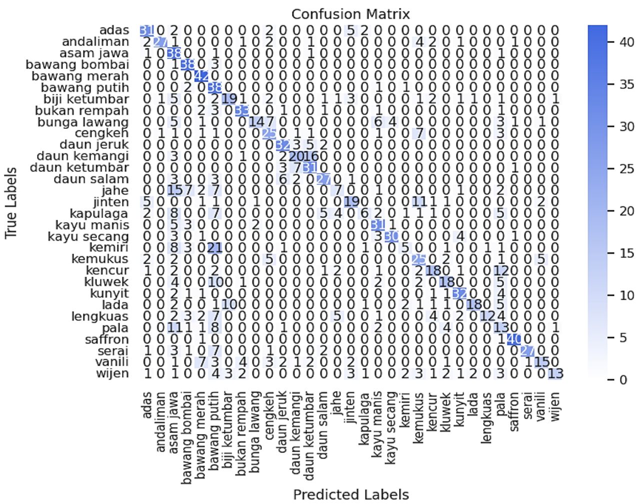

Figure 4. Confusion Matrix of the ResNet50

The confusion matrix illustrates that the ResNet50 model was able to classify visually distinctive classes such as galangal, vanilla, and sappanwood accurately. However, significant misclassifications occurred in classes with similar visual features, such as ginger, turmeric, and lesser galangal, as well as in minority classes such as saffron and kluwek. This pattern reflects the presence of feature similarity and data imbalance, indicating the need for improvements such as data augmentation and class reweighting to enhance the model's overall accuracy.

3.4. Evaluation of ResNet50 Model (Transfer Learning)

Transfer learning with pre-trained weights from ImageNet demonstrated a significant improvement in performance compared to training from scratch. In this experiment, fine-tuning was applied by unfreezing the last 33 layers of the ResNet50 architecture (starting from layer 143 to the final layer), while the remaining layers were kept frozen. The training results of the ResNet50 model with transfer learning are shown in Figure 5.

The loss and accuracy graphs indicate that the ResNet50 model using the transfer learning approach achieved excellent performance. Training loss dropped sharply to near-zero values, while validation loss remained stable at a low level without any significant increase, indicating the absence of overfitting. Furthermore, training accuracy increased consistently, reaching nearly 100%, and validation accuracy reached approximately 96%, maintaining stability until the end of training. This pattern suggests that the model was able to generalize well to new data, thanks to the effective combination of ImageNet pre-trained weights and fine-tuning strategies applied to the top layers.

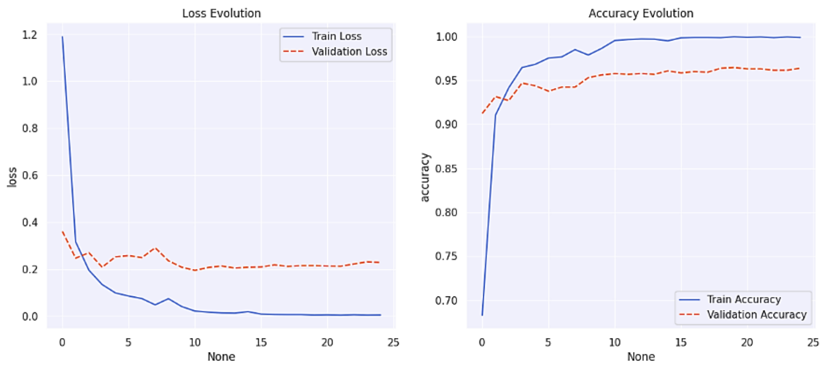

Figure 5. Loss Evolution and Accuracy Evolution Graph of ResNet50 with Transfer Learning

3.4.1. Confusion Matrix

Confusion matrix analysis was conducted to examine in detail the distribution of correct predictions (true positives) and incorrect predictions (false positives and false negatives) for each class. The confusion matrix of the ResNet50 model with transfer learning is shown in Figure 6.

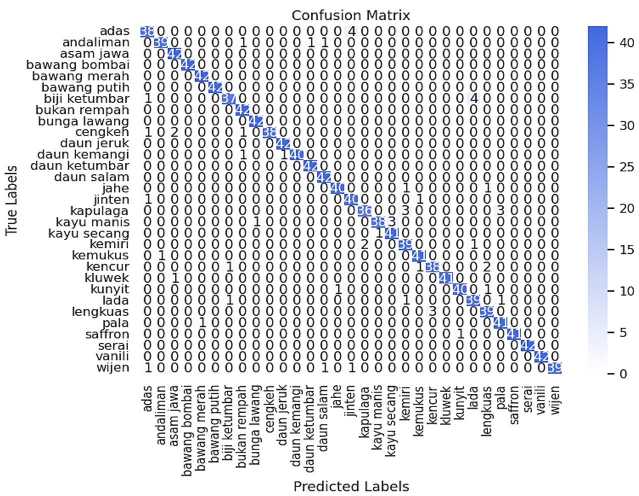

Figure 6. Confusion Matrix of the ResNet50 TL

The confusion matrix illustrates that the ResNet50 model with transfer learning achieved highly accurate classification results. Almost all values are located along the diagonal line, indicating that true positives dominate and the level of misclassification is very low.

To facilitate a clearer understanding of each model's performance, the evaluation results are presented in Table 3. This table compares the ResNet50 model trained from scratch with the ResNet50 model using transfer learning, based on key evaluation metrics: accuracy, precision, recall, and F1 Score.

Table 3. Model Evaluation Results

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet50 Model Trained from Scratch	57	64	57	56
ResNet50 Transfer Learning	96	96	96	96

The evaluation results indicate that the ResNet50 model using transfer learning significantly outperforms the model trained from scratch. Transfer learning achieved accuracy, precision, recall, and F1-score values of 96%, whereas the model trained from scratch only reached around 56% to 64%. This performance gap is due to transfer learning, leveraging pretrained weights from large-scale datasets such as ImageNet, allowing the model to recognize fundamental visual features. In contrast, training a model from scratch requires substantially more data and training time to learn similar patterns. Moreover, transfer learning enables faster convergence and better generalization, particularly when the training dataset is limited.

3.5. Comparison with Previous Research

To assess the advantages of the proposed approach, a comparison was made with previous studies that utilized CNN-based methods, as shown in Table 4.

Table 4. Comparison Was Made with Previous Studies

Study	Method	Dataset (Classes)	Accuracy (%)
Identification of Indonesian Spice Types Using CNN with VGG16 Architecture [7]	VGG16	31 Classes	57
This Study	ResNet50 TL	31 Classes	96

The evaluation conducted in this study demonstrates that the overall performance of the model surpasses that of previous research employing the VGG16 architecture. This improvement is largely attributed to the use of ResNet50, a more modern and efficient architecture, combined with the application of transfer learning during the training process. One of the key strengths of ResNet50 lies in its residual connections and shortcut paths that allow information to bypass several layers without alteration. This mechanism is particularly effective in addressing the vanishing gradient problem often encountered in deep networks, enabling the model to learn more complex features in a stable and efficient manner. Furthermore, the use of pretrained weights from ImageNet offers a significant advantage, as the model already possesses foundational knowledge of common visual patterns such as shapes, colors, and textures. This is especially beneficial when working with relatively small datasets, as the model does not need to learn all features from scratch. As a result, the training process becomes more efficient, and the classification outcomes are more accurate. The combination of a deeper yet stable architecture, the benefits of transfer learning, and potentially more optimized training parameters collectively contribute to the superior accuracy, precision, recall, and F1-score achieved by the ResNet50 model in this study compared to earlier approaches.

Although the ResNet50 model with transfer learning demonstrated excellent classification performance, several limitations should be noted. Class imbalance and visual similarity among spice classes such as ginger, turmeric, and lesser galangal led to misclassification in some cases. Additionally, the limited dataset size (210 images per class) constrained the model’s ability to generalize to new data. The fine-tuning strategy applied only to the last 33 layers may not have fully utilized the complete potential of the ResNet50 architecture. Therefore, future improvements using larger datasets, feature-specific tuning, or an end-to-end training approach could serve as effective solutions to enhance the model’s accuracy and robustness.

5. CONCLUSION

This research successfully developed an automated classification system for 31 Indonesian spices using the ResNet50 deep learning architecture. Two training strategies were explored: training from scratch and applying transfer learning with pre-trained ImageNet weights. The transfer learning model significantly outperformed the one trained from scratch, achieving 96% accuracy compared to only 57%. Key techniques such as image augmentation, class weighting, and fine-tuning of the top layers were effectively implemented to enhance model generalization. The transfer learning model demonstrated consistent and strong results across all evaluation metrics, including precision, recall, and F1-score. Although some misclassifications occurred in visually similar classes like ginger, turmeric, and galangal, the overall classification performance remained high. Compared to previous studies using architectures like VGG16, the ResNet50 with transfer learning provided better accuracy and robustness. In conclusion, this study confirms that the combination of a deeper architecture, pre-trained knowledge, and proper training strategies makes ResNet50 with transfer learning a powerful tool for spice classification. It holds great potential for real-world applications in the food industry, agriculture, and biodiversity conservation.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Universitas Mercu Buana Yogyakarta for their continuous support and the resources provided throughout the course of this research project. I am also deeply thankful to the Information Systems Study Program for fostering a rigorous academic environment and a spirit of collaboration, both of which have played a vital role in the development and success of this work.

REFERENCES

[1] C. Nelly, L. Fitriyana, T. D. Santi, and Saudah, “Diversity of traditional vegetables and spices as local food security for the Gayo Tribe, Aceh, Indonesia,” Biodiversitas, vol. 25, no. 12, pp. 4699–4711, 2024, doi: 10.13057/biodiv/d251206.

[2] I. N. Suandana and W. Apriandari, “Pemanfaatan CNN (Convolutional neural network) dan MobileNetV2 dalam klasifikasi rempah-rempah lokal di Indonesia,” vol. 8, no. 5, pp. 10109–10116, 2024.

[3] K. He, “Deep Residual Learning for Image Recognition”.

[4] V. N. Juli and N. T. Timur, “Klasifikasi Citra Digital Bumbu dan Rempah Dengan Algoritma Convolutional Neural Network (CNN) 1 . Klasifikasi Citra Digital : artikel atau elemen dalam gambar terkomputerisasi ke dalam klasifikasi atau kelas yang telah,” vol. 2, no. 3, 2024.

[5] A. E. Putra, M. F. Naufal, and V. R. Prasetyo, “Klasifikasi Jenis Rempah Menggunakan Convolutional Neural Network dan Transfer Learning,” J. Edukasi dan Penelit. Inform., vol. 9, no. 1, p. 12, 2023, doi: 10.26418/jp.v9i1.58186.

[6] C. Nisa and F. Candra, “Klasifikasi Jenis Rempah-Rempah Menggunakan Algoritma Convolutional Neural Network,” MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 1, pp. 78–84, 2023, doi: 10.57152/malcom.v4i1.1018.

[7] R. Maulana, R. Dwi Zahra Putri, T. Ade Amelia, H. Syahputra, and F. Ramadhani, “Identifikasi Jenis Rempah-Rempah Indonesia Dengan Convolutional Neural Network (Cnn) Menggunakan Arsitektur Vgg16,” JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 4, pp. 6034–6039, 2024, doi: 10.36040/jati.v8i4.10138.

[8] H. P. Tran, H. T. Diem Tuyet, T. Q. Dang Khoa, L. N. Lam Thuy, P. T. Bao, and V. N. Thanh Sang, “Microscopic Video-Based Grouped Embryo Segmentation: A Deep Learning Approach,” Cureus, vol. 15, no. 9, 2023, doi: 10.7759/cureus.45429.

[9] S. Jasmine and P. Marichamy, “MNR2NeXt-50: Segmentation and quantification of epicardial fat from cardiac CT images using transfer learning with an optimized ensemble model,” J. Radiat. Res. Appl. Sci., vol. 18, no. 3, p. 101648, 2025, doi: 10.1016/j.jrras.2025.101648.

[10] “Classification of Images To Detect Distracted Drivers By Using,” 2023.

[11] H. Khalifeh, “Olive Leaf Disease Detection Using Deep Learning Empowering Palestinian Agriculture with AI,” no. June, 2025, doi: 10.13140/RG.2.2.27931.37923/1.

[12] J. I. Mestre, S. Barrachina, D. Quezada, and M. F. Dolz, “Deep learning inference optimisation for IoT: Conv2D-ReLU-BN layer fusion and quantisation,” J. Supercomput., vol. 81, no. 4, 2025, doi: 10.1007/s11227-025-07107-y.

[13] B. Kim et al., “Deep Learning Activation Layer-Based Wall Quality Recognition Using Conv2D ResNet Exponential Transfer Learning Model,” Mathematics, vol. 10, no. 23, 2022, doi: 10.3390/math10234602.

[14] L. Wen, Z. Xiao, X. Xu, and B. Liu, “Disaster Recognition and Classification Based on Improved ResNet-50 Neural Network,” 2025.

[15] S. Thite, D. Godse, K. Patil, and P. Chumchu, “Facilitating spice recognition and classification : An image dataset of Indian spices Facilitating spice recognition and classification : An image dataset of Indian spices,” Data Br., vol. 57, no. October, p. 110936, 2024, doi: 10.1016/j.dib.2024.110936.

[16] S. Kanakala and S. Ningappa, “Detection and Classification of Diseases in Multi-Crop Leaves using LSTM and CNN Models,” J. Innov. Image Process., vol. 7, no. 1, pp. 161–181, 2025, doi: 10.36548/jiip.2025.1.008.

[17] D. Juodelyte, A. Jiménez-Sánchez, and V. Cheplygina, “Revisiting Hidden Representations in Transfer Learning for Medical Imaging,” Trans. Mach. Learn. Res., vol. 2023, pp. 1–19, 2023.

[18] C. N. N. Fine-tuning, “Optimizing South Kalimantan Food Image Classification Through,” vol. 10, no. 4, pp. 897–913, 2025, doi: 10.26555/jiteki.v10i4.30325.

[19] L. Campbell, “Transfer Learning in Image Classification Tasks for Small Datasets,” no. June, 2025.

[20] R. Ren, S. Zhang, H. Sun, and T. Gao, “Research on Pepper External Quality Detection Based on Transfer Learning Integrated with Convolutional Neural Network,” 2021.

BIBLIOGRAPHY OF AUTHORS

Julio Francisco Bacun is an eighth-semester student in the Information Systems program at Mercu Buana University Yogyakarta. He is currently conducting research on the application of Convolutional Neural Networks (CNNs) for digital image processing, focusing on developing an automatic classification system for various types of Indonesian spices using the ResNet-50 architecture, aiming to support the spice industry and biodiversity conservation.

Irfan Pratama is a professional in the field of Data Mining with expertise in extracting valuable insights from big data to support data-driven decision-making. With an educational background in Information Technology and a concentration in Data Mining, he is proficient in various data analysis techniques, including clustering, classification, regression, and association rule mining. He is also skilled in applying machine learning algorithms to process and analyze complex data.