ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Optimizing Scalability in Spice Identification through Transfer Learning with Convolutional Neural Networks

I Nyoman Switrayana[1*], Muhamad Azwar[2]
[1,2]Computer Science, Faculty of Engineering, Bumigora University
[1*]nyoman.switrayana@universitasbumigora.ac.id, [2]azwar@universitasbumigora.ac.id

Abstract. Indonesia is renowned for its rich diversity of spices, which hold significant cultural and economic value. However, public knowledge of these spices remains limited, making their identification challenging. Addressing this issue, this study aims to develop a scalable spice identification system using Convolutional Neural Networks (CNN) with a Transfer Learning approach. The system is designed to recognize 30 types of spices while maintaining high accuracy, utilizing the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework for systematic development. The dataset was collected through open sources and web scraping from Google Images. Four CNN models (ResNet50, EfficientNetB0, Xception, and MobileNet) were evaluated under three data splits: 90:10, 80:20, and 70:30. Performance metrics including accuracy, precision, recall, and F1-score were used for evaluation. Among these models, Xception achieved the best performance in the 90:10 split, with an accuracy of 84.51%, followed by EfficientNetB0 at 83.57%. The results demonstrate that transfer learning effectively enhances model accuracy and scalability, enabling reliable spice identification across diverse categories. This system has practical implications for promoting public awareness, supporting culinary industries, and preserving Indonesia's rich spice heritage. The proposed approach highlights the potential of CNN-based systems for addressing classification challenges in resource-constrained settings, offering a foundation for future research and real-world applications.
Purpose: The purpose of this study is to address the limitations of previous research on spice classification, which have been constrained by small datasets and a narrow range of spice varieties, limiting the scalability and generalization of existing models. This research aims to develop a CNN-based system capable of identifying 30 types of Indonesian spices by leveraging a larger and more diverse dataset, along with advanced CNN architectures. By focusing on scalability and maintaining high accuracy, the study seeks to provide a robust and efficient solution to the growing need for spice classification systems. Additionally, this research contributes insights into the data requirements and performance of different architectures under varying conditions, bridging the gap in existing studies and advancing the field of spice recognition.
Methods/Study design/approach: This study utilized a systematic approach based on the CRISP-DM framework to develop a scalable spice identification system using Convolutional Neural Networks (CNN) and Transfer Learning. The dataset, comprising 30 types of Indonesian spices, was sourced through web scraping and divided into three splits: 90:10, 80:20, and 70:30. Four CNN models, namely ResNet50, EfficientNetB0, Xception, and MobileNet, were fine-tuned using pre-trained weights to optimize their performance for spice classification.
Result/Findings: The results highlight a relationship between model architecture, data availability, and performance. Xception and EfficientNetB0 performed best with ample training data, achieving high accuracy and resilience. However, models like ResNet50 showed significant performance drops with reduced data. MobileNet demonstrated the highest stability across all splits, maintaining consistent results even with limited data. Larger splits (e.g., 90:10) enabled complex architectures to excel, while smaller splits (e.g., 70:30) exposed the limitations of less generalizable models.
Novelty/Originality/Value: This research lies in the comparative analysis of four advanced CNN models: ResNet50, EfficientNetB0, Xception, and MobileNet, applied to a diverse dataset of 30 Indonesian spices. This study stands out by utilizing a larger and more varied dataset compared to previous research, which typically involved fewer spice categories. By evaluating the performance of these models under different data availability conditions, this study provides a deeper understanding of how data size influences model effectiveness.

Keywords: Convolutional Neural Network, CRISP-DM, Optimizing, Spices Identification, Transfer Learning

Received January 2025 / **Revised** June 2025 / **Accepted** June 2025

INTRODUCTION

Indonesia is globally recognized for its rich biodiversity and natural resources [1], especially its vast array of spices, which are essential in both traditional Indonesian cuisine and the global spice trade. These spices, derived from various parts of plants such as stems, leaves, bark, roots, seeds, and flowers [2], play a crucial role in the cultural heritage and culinary practices of the nation. In 2016, according to data from

DOI: 10.24014/coreit.v11i1.35453

the Food and Agriculture Organization (FAO), Indonesia ranked fourth among the world's top spice-producing countries [3]. However, despite their significance, many of Indonesia's spices share similar physical characteristics, making them challenging to distinguish accurately, especially for those unfamiliar with them. This poses a significant issue, as public awareness, particularly among younger generations, regarding the various types of spices remains limited, with few educational initiatives addressing this gap.

Given the growing demand for accurate spice identification, there is an urgent need for innovative technological solutions to assist in overcoming these challenges. With the rapid advancement of digital technologies and data-driven algorithms, considerable breakthroughs have been made in the field of object recognition and classification. Deep Learning, a subfield of artificial intelligence, has emerged as a powerful approach for handling complex image recognition tasks, offering high accuracy and efficiency [4]. Convolutional Neural Networks (CNN), a popular technique in Deep Learning, are particularly effective for image classification tasks due to their ability to learn hierarchical feature representations from high-dimensional data, enabling them to perform well even on large and complex datasets [5], [6]. These capabilities make CNN an ideal tool for developing systems aimed at identifying and classifying spices.

Several studies have explored the application of Convolutional Neural Networks (CNNs) in spice classification, each contributing valuable insights to the field. For instance, [7] and [8] demonstrated promising results, achieving 85% and 88% testing accuracy with a relatively small dataset consisting of 300 images. This study used CNNs to classify 3 categories of spices based on their visual features, showing the potential of deep learning in addressing the challenges of spice identification. Similarly, [9] applied Transfer Learning to a dataset of 1,000 images across 10 different spice classes, reporting a notable accuracy of 97%. Their work highlights the efficacy of Transfer Learning in improving model performance, especially when working with limited data. However, the study was still limited by the number of spice classes and the dataset's diversity. The impact of using a large model on the dataset partitioning for training and testing is not discussed. Specifically, in this case, a model with a diverse architecture is used.

Furthermore, [10] investigated the effect of data augmentation on spice classification with 4 classes, showing significant improvements in accuracy—from 54% to 80%. This research underscored the importance of augmenting limited datasets to prevent overfitting and improve model robustness. However, the study was based on 4 categories of spices, which raises questions about the model's scalability and its performance with larger and more diverse datasets. Moreover, the study only used basic augmentation techniques, leaving room for exploring more sophisticated methods to enhance the model's generalization. In another study, [11], [12], [13], applied the K-Nearest Neighbor (KNN) algorithm and [14] using Naïve Bayes to classify spices based on RGB color features and texture extraction. The RGB values of each channel are used as features. The researchers employed a preprocessing step where images were converted to grayscale, and texture features such as energy and homogeneity were extracted. While this approach demonstrated promising results, it was reliant on handcrafted features, which often fail to capture the full range of variations inherent in spice images, especially those with similar colors and textures. Additionally, KNN's performance in higher-dimensional spaces may degrade, as it becomes computationally expensive with large datasets.

Previous studies have shown that the modified VGG16 model, with adjustments to the number of layers and parameters, achieved 85% accuracy in classifying five spice categories, outperforming AlexNet in extracting features from small datasets [15]. The simplified model proved to capture features more effectively with limited data. This was evident when the unmodified VGG16 model was applied, as it experienced overfitting [16]. A more complex model, VGG19, demonstrated better performance, particularly in capturing relevant features from images when data variations increased [17]. Additionally, transfer learning using MobileNetV2 [2] and DenseNet121 [18] yielded high performance. By implementing both models, [19] reported accuracies of 93% and 94%, respectively, for classifying five spice categories. These studies also examined the impact of the number of epochs, finding that higher epoch counts generally improved model performance. Based on these findings, this research will use 50 epochs to optimize results. Given the proven effectiveness of MobileNet variants in previous studies, this research will explore the capability of the baseline MobileNet model to classify a larger number of spice categories and evaluate its robustness against data variations.

The study by [20] on the classification of spice types using Convolutional Neural Networks (CNN) involved 4 spice categories. The CNN model achieved an accuracy of 90%. Following this, the work by [21] successfully classified 18 spice types with an accuracy rate of 75%. A total of 2,700 images were used during the training phase. However, the CNN model in this study still struggled to reliably recognize a large variety of spices. From previous research, it is evident that an increase in the diversity of spice types leads to a significant decrease in model performance. In particular, while the CNN model performs well with fewer categories, its ability to generalize effectively declines as the number of spice types increases. This can be attributed to the challenges posed by the substantial intra-class variation and the lack of sufficient training samples for each class, which impairs the model's ability to differentiate subtle differences between similar spice types. Furthermore, the complexity of the feature extraction process becomes more pronounced as the number of classes rises, requiring more sophisticated techniques to maintain or improve accuracy.

The literature highlights a common limitation in spice classification studies: reliance on small datasets and a narrow range of spice varieties. While CNNs have demonstrated great potential in this domain, their scalability and generalization remain constrained by the lack of diverse, larger datasets. Addressing these challenges requires employing advanced techniques such as Transfer Learning and data augmentation. To tackle these limitations, this study investigates the performance of four CNN architectures: ResNet50, EfficientNetB0, Xception, and MobileNet. These models are chosen for their proven efficacy in diverse image classification tasks. The research focuses on developing a robust system capable of classifying 30 types of Indonesian spices, leveraging a more extensive and varied dataset to enhance both scalability and accuracy. Furthermore, the study examines the data requirements of each architecture, offering insights into their performance under different conditions. By advancing the scalability and robustness of CNN-based spice classification, this research seeks to bridge existing gaps in the field and contribute meaningfully to automated spice recognition.

METHODS

This chapter outlines the implementation steps for developing an Indonesian spice recognition system using Convolutional Neural Networks (CNN) and adopting the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. This framework is divided into six key stages, as shown in Figure 1: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [22]. However, this study is limited to the first five stages. The final stage, Deployment, is beyond the scope of this research. The focus is on building and assessing the model to ensure reliable performance for recognizing Indonesian spices.

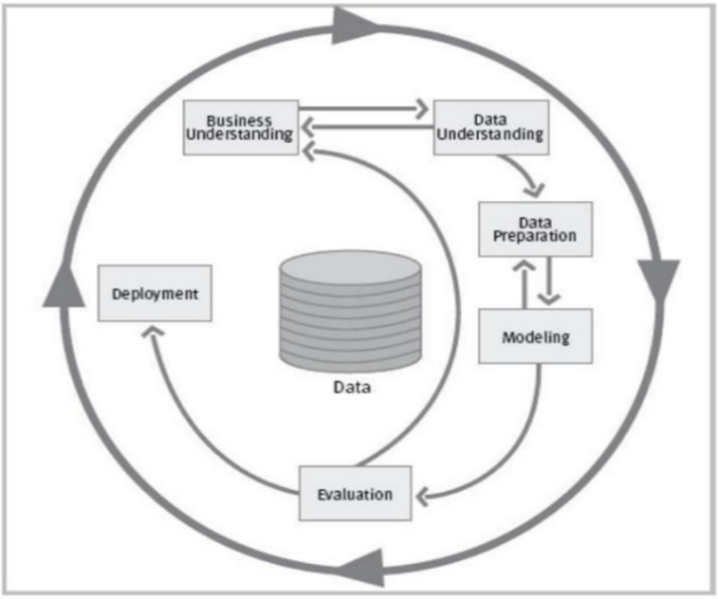

Figure 1. Metodologi CRISP-DM [23]

1) Business Understanding

Business understanding is a crucial initial stage in every artificial intelligence-based system development project. This stage aims to define business needs, system objectives, and success indicators for its implementation. In the context of this study, business understanding helps identify the importance of Indonesian spice recognition as a solution to simplify the identification of various commonly used spices while simultaneously raising public awareness of spice diversity. With this understanding, a research framework is designed, covering data management, model selection, and system performance evaluation.

2) Data Understanding

After achieving a deep business understanding, the next step is data understanding, which is an essential key in the development of the Convolutional Neural Network (CNN)-based spice recognition system. This step involves in-depth analysis of the dataset to be used for training and testing the CNN model. To obtain a representative dataset, identifying secondary data sources becomes critical. Secondary data has been obtained from Kaggle and supplemented by data scraping from Google Images. The data scraping from Google Images will use keywords for the list of Indonesian spices as shown in Table 1. During the data understanding process, the focus is on the dataset's characteristics, including image resolution, color variation, and the representation of each spice class.

Table 1. List of Indonesian Spices to Be Recognized by the System

No	Spice Names	No	Spice Names	No	Spice Names
1	Andaliman	11	Bunga Lawang	21	Kunyit
2	Biji Ketumbar	12	Wijen	22	Kayu Secang
3	Daun Ketumbar	13	Pala	23	Adas
4	Kapulaga	14	Bawang Bombai	24	Lengkuas
5	Kluwek	15	Asam Jawa	25	Bawang Putih
6	Kencur	16	Daun Kemangi	26	Kemiri
7	Vanili	17	Daun Jeruk	27	Daun Salam
8	Bawang Merah	18	Kayu Manis	28	Jinten
9	Jahe	19	Cengkeh	29	Kemukus
10	Lada	20	Saffron	30	Serai

3) Data Preparation

This phase involves several crucial steps to ensure that the model can learn effectively and make accurate predictions. Data cleaning is performed to remove any irrelevant or corrupted images. Pixel value normalization is then applied to standardize the image data, ensuring that the model processes the input in a consistent manner. Data augmentation is employed to artificially expand the dataset by generating new images through transformations such as rotation, scaling, and flipping. Image resizing ensures that all images have a consistent size, which is important for CNN processing. Lastly, the data is split into training and testing sets, commonly with ratios like 90:10, 80:20, or 70:30, to train the model and evaluate its performance. By preparing the image data carefully through these steps, the CNN model can be trained effectively, optimizing its accuracy in recognizing and classifying various spices.

4) Modeling

In the modeling phase of the CRISP-DM methodology, several deep learning models will be utilized for the task of Indonesian spice recognition. These models include ResNet50, EfficientNetB0, Xception, and MobileNet, all of which will be trained using transfer learning. Transfer learning is a technique where a model pre-trained on a large, general dataset (such as ImageNet) is fine-tuned for a specific task without the need for retraining from scratch. ResNet50 leverages its deep architecture with residual connections, which allow it to train deeper networks without degradation in performance [24]**.** This approach enables the model to capture complex features effectively. EfficientNetB0, designed for optimal efficiency, balances model width, depth, and resolution to provide an efficient learning framework [25]**.** Using transfer learning with EfficientNetB0, the model can quickly adapt to new datasets by leveraging shared features from ImageNet. Xception, known for its depthwise separable convolutions, reduces the number of parameters while maintaining performance, which is particularly beneficial for tasks requiring less computational power [26]**.** MobileNet, optimized for mobile devices, uses a lightweight architecture to maintain high accuracy with reduced computational requirements [27]**.** Transfer learning with all models involves pre-training the model on ImageNet and then fine-tuning it on the spice recognition dataset, allowing it to learn relevant features efficiently. By utilizing these models with transfer learning, the study aims to enhance the accuracy and efficiency of the deep learning models in recognizing a wide variety of Indonesian spices.

5) Evaluation

The evaluation stage is carried out to determine the model's performance. Metrics such as accuracy, precision, recall, and F1-score, as outlined by [28], are employed to evaluate the model's effectiveness in identifying spices within the test set. The formulas for these metrics are provided in Equations 1-4. True Positives (TP) represent correctly identified positive classes, True Negatives (TN) refer to correctly identified negative classes, False Positives (FP) occur when negative classes are misclassified as positive, and False Negatives (FN) arise when positive classes are misclassified as negative.

$$\text{Precision } = \frac{TP}{TP+FP} \tag{1}$$

$$\text{Recall } = \frac{TP}{TP+FN} \tag{2}$$

$$\text{F1}-\text{Score } = \frac{2\ x\ Precision\ x\ Recall}{Precision+Recall} \tag{3}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

RESULT AND DISCUSSION

This section will discuss the main results obtained through experiments and analysis conducted during the modeling and evaluation phases of the study. The primary objective of these results is to explore the effectiveness of the developed model for recognizing Indonesian spices using Convolutional Neural Networks (CNN) based on the CRISP-DM methodology. The analysis will cover various aspects, including accuracy, prediction errors, and comparisons between the methods used. Through a deep discussion of these results, the researcher can identify the strengths and weaknesses of the model and provide further insights into its potential applications in real-world scenarios. This discussion will also include implications of these findings for improving the quality of the Indonesian spice recognition system in the future.

1) Business Understanding

Business understanding in the context of Indonesian spice recognition serves as a critical foundation for this project. This stage identifies the primary objectives of the system, which are to provide ease in identifying commonly used spice varieties and enhance public knowledge about spices. One of the project’s significant aspects is the recognition of 30 distinct spice types, representing the rich diversity of Indonesian spices. To achieve these goals, the system was developed using a *transfer learning* approach with state-of-the-art *deep learning* models, including ResNet50, EfficientNetB0, Xception, and MobileNet. These models were trained to recognize and differentiate a wide range of spice varieties. Success metrics, such as recognition accuracy, are crucial to evaluate the system’s performance, particularly given the complexity introduced by the inclusion of numerous spice types.

2) Data Understanding

The dataset for this project was obtained from public open-source platforms, such as Kaggle, and through data scraping from Google Images. Each class contains approximately 210 images, resulting in a total of 6,398 images representing 30 types of Indonesian spices. As part of the data understanding process, it was essential to ensure that the dataset reflects the diversity and characteristics representative of various Indonesian spices. In the context of data scraping, data understanding also involves determining the most relevant attributes or features to train the model. This is necessary for the CNN model to recognize and understand the unique characteristics of each spice type. With this approach, the model development is expected to achieve high accuracy and be responsive to potential variations in the dataset, ensuring reliability in the spice recognition process in daily life. Figure 2 shows a sample of the dataset used in this study, which consists of images representing 30 types of Indonesian spices. Each image is labeled with its corresponding spice category, providing the necessary ground truth for the training and evaluation of the recognition model.

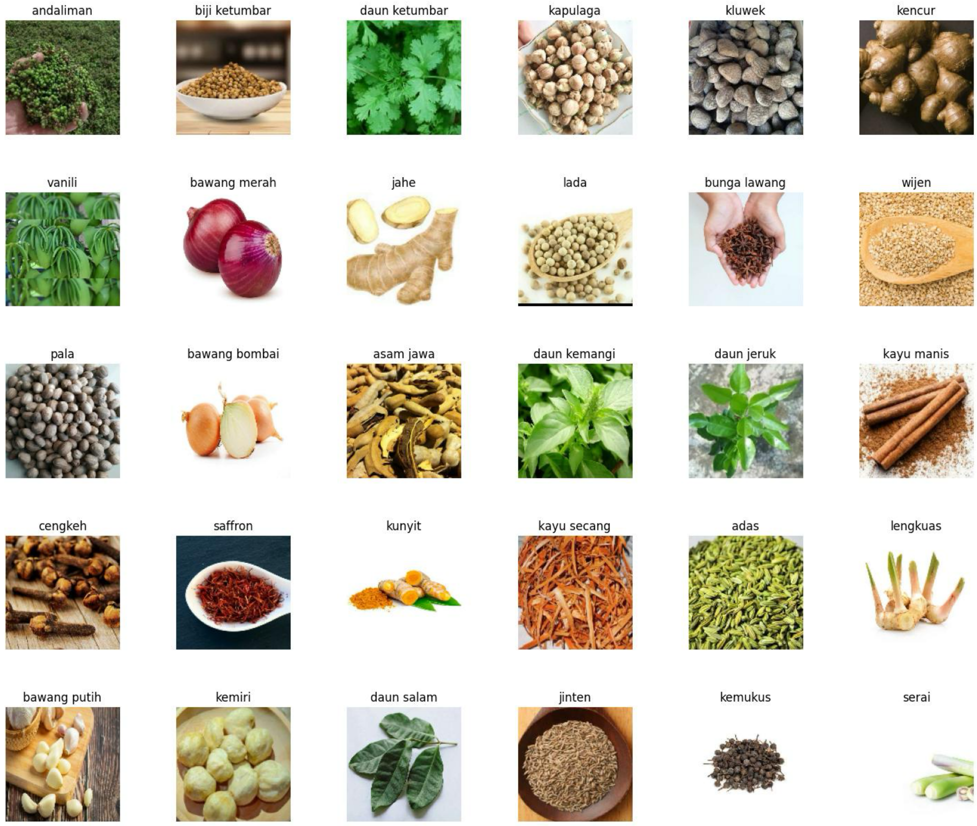

Figure 2. Samples of Dataset

3) Data Preparation

To enhance the model's generalization capabilities and prevent overfitting, several data augmentation techniques were employed: random flipping, random rotation, and rescaling. Random flipping was used to simulate different perspectives of the same spice by horizontally or vertically flipping the images. Figure 3 demonstrates the data preparation process applied to the spice dataset before training with the CNN model.

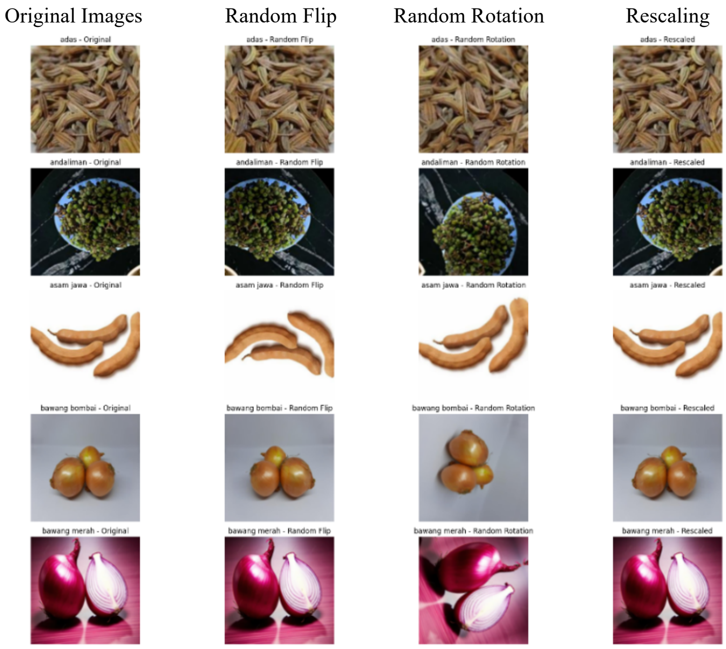

Figure 3. Example of data preparation random flip, random rotation, and rescaling results

Random rotation was applied to rotate images at small angles, allowing the model to learn rotation-invariant features, which is particularly beneficial for spices, where orientation can vary in real-world scenarios. Rescaling standardized all images to a consistent size, ensuring that variations in image resolution do not affect the model's performance. This consistent scaling across images helps maintain visual integrity and prepares the dataset for effective CNN training. These preprocessing steps are essential for creating a robust, diverse dataset that trains a model capable of handling various spice images effectively. After all preparations, the data is then split into training and testing datasets with ratios of 90:10, 80:20, and 70:30.

4) Modeling

During the modeling phase, the primary focus is on designing and implementing a Convolutional Neural Network (CNN) for spice recognition. To enhance the system's performance, several pre-trained models are utilized as backbones through the transfer learning approach. These models include ResNet50, EfficientNetB0, Xception, and MobileNet. Each model serves as the feature extraction layer, leveraging its pre-trained weights on the ImageNet dataset to extract high-level features from spice images efficiently. The use of CNNs as an automatic feature extraction tool generates important features [29], which facilitate the classifier in making more accurate predictions. Before the image dataset is trained on the model, the data needs to be prepared during the data preparation stage. Each model processes input images with a fixed size of 128×128×3, where 128x128 represents the spatial resolution, and 3 corresponds to the RGB color channels. Figure 4 illustrates the proposed model architecture for the Indonesian spice recognition system.

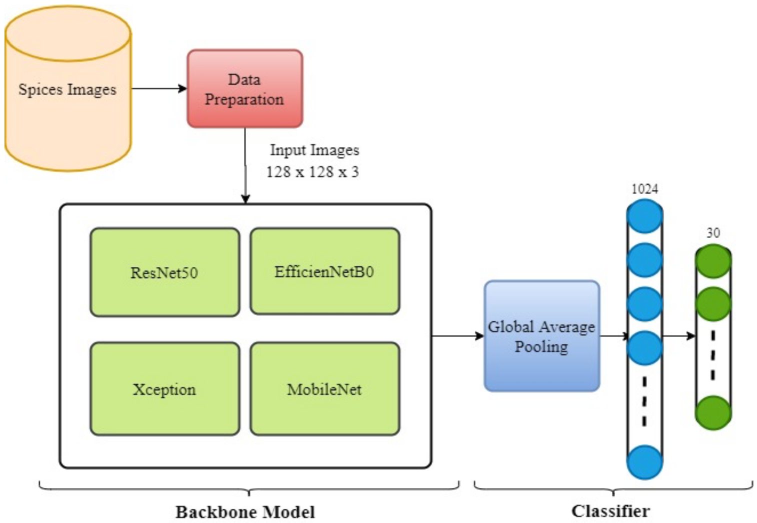

Figure 4. Proposed Model Architecture

The transfer learning approach enables the system to capitalize on the robustness of these pre-trained models while adapting them to the specific task of spice classification. The models are modified by removing their top layers and freezing the convolutional base to retain the learned features. Subsequently, custom classification layers are added on top of the backbone to tailor the network for the specific task. For example, the ResNet50-based architecture includes a global average pooling layer, a fully connected dense layer with 1024 units and ReLU activation, a dropout layer for regularization, and a final dense layer with 30 neurons corresponding to the 30 spice classes. The output layer uses a softmax activation function to predict the probabilities for each class. All models follow a similar structure for their classifier component. The training process optimizes the network using the Adam optimizer with a learning rate of 0.001 and sparse categorical cross-entropy as the loss function. The models were trained on the prepared dataset for 50 epochs with a batch size of 32, using the TensorFlow framework. The experiments were conducted in a computing environment equipped with 24 GB of RAM, an NVIDIA GeForce RTX 3060 GPU with 6 GB of VRAM, and an AMD Ryzen 7 5800H processor with integrated Radeon graphics running at 3.20 GHz.

By using multiple backbones, the system evaluates and compares the performance of each model in terms of accuracy, precision, recall, and F1-score. This approach allows for identifying the most suitable model architecture for spice recognition, ensuring scalability and robustness across diverse datasets.

5) Evaluation

The evaluation of spice recognition systems using various CNN architectures across three data split scenarios (90:10, 80:20, and 70:30) provides valuable insights into the influence of data availability and architectural complexity on model performance. The 90:10 split, characterized by a large training dataset, yielded the best overall performance for most models. This result highlights the importance of abundant training data for deep learning models, especially those with high architectural complexity. These models excelled in this scenario, with Xception achieving the highest accuracy (84.51%) and EfficientNetB0 close behind (83.57%). Their strong performance can be attributed to their ability to learn complex feature representations when provided with sufficient data. The evaluation results of the models across all split scenarios are presented in Table 2.

Table 2. Model Evaluation Results

Model	Accuracy	Precision	Recall	F1-Score
90:10 Scenario				
Resnet50	76,21	77,08	76,21	75,23
EfficientNetB0	83,57	85,82	83,57	83,75
Xception	**84,51**	86,96	84,51	84,52
Mobilenet	74,02	79,55	74,02	73,89
80:20 Scenario				
Resnet50	68,02	73,12	68,02	66,69
EfficientNetB0	**80,84**	84,06	80,84	81,08
Xception	78,97	83,57	78,97	79,48
Mobilenet	76,62	81,69	76,62	76,28
70:30 Scenario				
Resnet50	56,80	67,39	56,80	56,83
EfficientNetB0	72,90	78,30	72,90	72,65
Xception	**81,03**	84,37	81,03	80,92
Mobilenet	78,11	81,00	78,11	77,87

In the 80:20 split, where the training dataset size was reduced, there was a noticeable drop in performance for most models, reflecting the impact of data availability on training. EfficientNetB0 continued to perform robustly with 80.84% accuracy, followed by Xception (78.97%). These models demonstrated resilience, likely due to their efficient use of feature extraction and their ability to generalize better with slightly less data. The other architectures like ResNet50 showed significant performance degradation, dropping to 68.02% accuracy, indicating a reliance on larger training datasets to achieve optimal results. The 70:30 split, with the largest test set and smallest training set, posed the greatest challenge for all models, emphasizing the importance of training data size. Despite this, Xception remained the top performer with 81.03% accuracy, followed by Mobilenet (78.11%). These models' ability to maintain competitive performance under data constraints demonstrates their architectural robustness and advanced feature extraction capabilities, especially Mobilenet. Mobilenet was more stable in its performance across all scenarios, even achieving the highest performance in the 70:30 scenario.

The results reveal a clear relationship between model architecture, data availability, and performance. Architectures like Xception and EfficientNetB0—characterized by their depth and sophisticated feature extraction mechanisms—excel in scenarios with ample training data but show resilience even as data availability decreases. A complex models or those less optimized for efficient feature extraction, such as ResNet50 is more vulnerable to performance drops with reduced data. This pattern underscores the importance of selecting appropriate architectures based on data availability and the complexity of the recognition task. Notably, MobileNet demonstrates the highest stability across all experiments, maintaining consistent performance even under limited data conditions. Overall, the findings emphasize that larger data splits (e.g., 90:10) allow complex transfer learning architectures to achieve optimal performance, leveraging their ability to learn nuanced features from abundant data. Conversely, smaller splits (e.g., 70:30) expose the limitations of models that are not designed to generalize well with constrained training data. These insights advocate for the use of advanced data augmentation, transfer learning, and hyperparameter optimization to mitigate the challenges posed by limited datasets, ensuring robust and reliable spice recognition systems. Figure 5 provides a detailed performance comparison of different models (Resnet50, EfficientNetB0, Xception, MobileNet) across varying train-test split ratios (70:30, 80:20, 90:10).

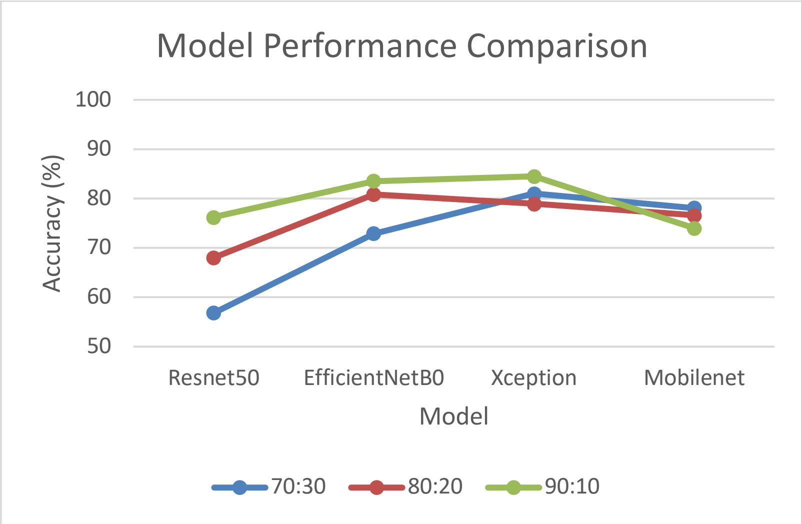

Figure 5. Model Performance Comparison

Based on figure 5, this study aligns with the findings of [25], which show that the Xception model performs the best, followed by EfficientNetB0 and ResNet50 in ImageNet classification tasks. The superiority of Xception is attributed to its larger number of parameters, which results in better performance compared to EfficientNetB0. However, despite having fewer parameters, EfficientNetB0's performance is not significantly different from Xception. Interestingly, although ResNet50 has more parameters to train, its performance is lower than both Xception and EfficientNetB0. Furthermore, the MobileNet model outperforms ResNet50 in this study. To improve the quality of the Indonesian spice recognition system in the future, it is recommended to implement advanced data augmentation techniques. Cross-validation can also provide a more comprehensive evaluation, ensuring model performance consistency across different splits and reducing potential biases. By addressing these areas, future development can enhance the accuracy and reliability of the system, particularly in real-world scenarios with varying data conditions. The confusion matrix in Figure 6 provides a detailed view of the Xception model's performance on the testing dataset for the Indonesian spice recognition task. This matrix compares the true labels (actual spices) along the vertical axis with the predicted labels (model's predictions) along the horizontal axis.

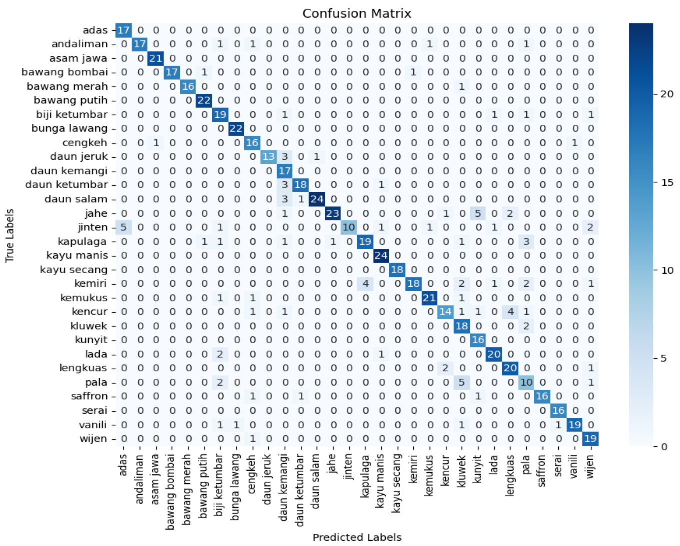

Figure 6. Confusion Matrix of Xception Model

The diagonal elements, which represent correctly classified samples, show high accuracy, particularly for spices like adas, asam jawa, bawang putih, bunga lawang, daun kemangi, kayu manis, kayu secang, kunyit, and serai. However, there are instances of misclassification, such as jahe being predicted as kunyit (5 instances), jinten being predicted as Adas (5 instances), kemiri being predicted as kapulaga (4 instances), kencur being predicted as lengkuas (4 instances), pala being predicted as biji ketumbar (2 instances) and kluwek (5 instances), reflecting the visual similarity between these spices. The heatmap's color intensity highlights areas where predictions deviate from the true labels, indicating where the model might benefit from further refinement. Despite these misclassifications, the Xception model shows strong overall recognition accuracy, demonstrating the effectiveness of the model but also suggesting areas for improvement in handling specific spice variants.

Figure 7. Prediction Results on Testing Data Using the Best Model (Xception)

Figure 7 showcases the prediction results obtained on the testing dataset using the Xception model, which was identified as the best-performing model during the evaluation phase. The figure includes sample input images from the testing dataset, along with their predicted labels and corresponding ground truth labels. Correct predictions are highlighted in green, while incorrect predictions are marked in red. One notable misclassification is the case where "biji ketumbar" was predicted as "pala", likely due to visual similarities in shape and texture between the two spices. Similar to the previous study in [30], using the VGG16 model with 31 classes faced the challenge of the model struggling to distinguish spices with similar visual features. This visualization demonstrates the model's overall effectiveness in recognizing various spice types while also providing insights into specific instances where misclassifications occurred, highlighting areas for further improvement.

CONCLUSION

This study successfully developed a spice recognition system leveraging various Convolutional Neural Network (CNN) architectures through transfer learning, with Xception and EfficientNetB0 showing superior performance across different data splits. The results underscore the importance of dataset quantity and quality, where larger splits like 90:10 facilitated more effective training for complex models. The more complex the model, the larger the data required. Future research should focus on expanding data augmentation techniques to increase dataset diversity, enhancing the model's ability to generalize to similar caracteristics and unseen data. Additionally, incorporating a greater volume of primary data captured in real-world conditions, such as through smartphone cameras, will provide more varied and realistic inputs. This approach ensures the model is robust and practical for real-time applications. Exploring ensemble methods and architecture-specific hyperparameter tuning can further improve model accuracy and reliability. Deploying the system in real-world settings and integrating user feedback will also help refine its capabilities and usability for practical spice recognition tasks.

ACKNOWLEDGEMENT

This research is funded by Bumigora University, and we would like to express our sincere gratitude for their support in making this study possible. We also appreciate the contributions of all individuals and departments involved in facilitating this project.

REFERENCES

[1] M. Sanjaya and E. Nurraharjo, "Deteksi Jenis Rempah-Rempah Menggunakan Metode Convolutional Neural Network Secara Real Time," *J. Sains Komput. Inform. (J-SAKTI*, vol. 7, no. 1, pp. 22–31, 2023.

[2] I. N. Suandana, A. Asriyanik, and W. Apriandari, "Pemanfaatan CNN (Convolution Neural Network) Dan Mobilenet V2 Dalam Klasifikasi Rempah-Rempah Lokal Di Indonesia," vol. 8, no. 5, pp. 10109–10116, 2024.

[3] Kemenparekraf/Baparekraf RI, "Rempah-Rempah Khas Indonesia yang Banyak Diekspor," *Minggu, 22 Agustus*, 2021. https://www.kemenparekraf.go.id/hasil-pencarian/rempah-rempah-khas-indonesia-yang-banyak-diekspor (accessed Dec. 25, 2024).

[4] O. Diaz Annesa, C. Kartiko, and A. Prasetiadi, "Identifikasi Spesies Reptil Menggunakan Convolutional Neural Network (CNN)," *Rekayasa Sist. dan Teknol. Inf.*, vol. 1, no. 3, pp. 899–906, 2019.

[5] B. Dwi Mardiana, W. Budi Utomo, U. Nur Oktaviana, G. Wasis Wicaksono, and A. Eko Minarno, "Herbal Leaves Classification Based on Leaf Image Using CNN Architecture Model VGG16," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 1, pp. 20–26, 2023, doi: 10.29207/resti.v7i1.4550.

[6] M. Harahap, Em Manuel Laia, Lilis Suryani Sitanggang, Melda Sinaga, Daniel Franci Sihombing, and Amir Mahmud Husein, "Deteksi Penyakit Covid-19 Pada Citra X-Ray Dengan Pendekatan Convolutional Neural Network (CNN)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 70–77, 2022, doi: 10.29207/resti.v6i1.3373.

[7] I. Wulandari, H. Yasin, and T. Widiharih, "Klasifikasi Citra Digital Bumbu Dan Rempah Dengan Algoritma Convolutional Neural Network (CNN)," *J. Gaussian*, vol. 9, no. 3, pp. 273–282, 2020, doi: 10.14710/j.gauss.v9i3.27416.

[8] R. Alondeo Boimau and Y. R. Kaesmetan, "Klasifikasi Citra Digital Bumbu dan Rempah Dengan Algoritma Convolutional Neural Network (CNN)," vol. 2, no. 3, 2024.

[9] 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.

[10] W. M. Pradnya D and A. P. Kusumaningtyas, "Analisis Pengaruh Data Augmentasi Pada Klasifikasi Bumbu Dapur Menggunakan Convolutional Neural Network," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2022, 2022, doi: 10.30865/mib.v6i4.4201.

[11] Kaharuddin, Kusrini, and E. T. Luthfi, "Klasifikasi Jenis Rempah-Rempah Berdasarkan Fitur Warna Rgb Dan Tekstur Menggunakan Algoritma K-Nearest Neighbor," *J. Inf. Interaktif*, vol. 4, no. 1, pp. 17–22, 2019.

[12] S. Y. Riska and L. Farokhah, "Klasifikasi Bumbu Dapur Indonesia Menggunakan Metode K-Nearest Neighbors (K-NN)," *Smatika J.*, vol. 11, no. 01, pp. 37–42, 2021, doi: 10.32664/smatika.v11i01.568.

[13] F. A. Mufarroha and D. Abdul Fatah, "Klasifikasi Jenis Rempah Penghasil Minyak Atsiri Menggunakan Metode Machine Learning," *J. Simantec*, vol. 11, no. 1, pp. 123–130, 2022, doi: 10.21107/simantec.v11i1.19743.

[14] N. P. Batubara, D. Widiyanto, and N. Chamidah, "Klasifikasi Rempah Rimpang Berdasarkan Ciri Warna RGB Dan Tekstur GLCM Menggunakan Algoritma Naive Bayes," *Inform. J. Ilmu Komput.*, vol. 16, no. 3, p. 156, 2020, doi: 10.52958/iftk.v16i3.2196.

[15] E. Tanuwijaya and A. Roseanne, "Modifikasi Arsitektur VGG16 untuk Klasifikasi Citra Digital Rempah-Rempah Indonesia," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 189–196, 2021, doi: 10.30812/matrik.v21i1.1492.

[16] S. Hermawan and N. Agustina, “Implementasi Convolutional Neural Network untuk Klasifikasi Rempah-Rempah Khas Indonesia,” *DoubleClick J. Comput. Inf. Technol.*, vol. 7, no. 1, pp. 1–7, 2023.
[17] P. Kinanti, R. Gustriansyah, and Z. R. Mair, “Penggunaan Convolutional Neural Network (CNN) Untuk Klasifikasi Jenis Rempah – Rempah,” no. November, 2024.
[18] C. Fadillah and S. Supatman, “Klasifikasi Jenis Rempah-Rempah Alami Untuk Kecantikan Menggunakan Densnet 121,” vol. 9, no. 1, pp. 923–929, 2025.
[19] Y. Hatur Puspita and A. Sabri, “Transfer Learning Model Pralatih MobileNetV2 dan DenseNet121 untuk Klasifikasi Tanaman Rempah,” *J. Ilm. Komputasi*, vol. 23, no. 1, pp. 67–74, 2024, doi: 10.32409/jikstik.23.1.3502.
[20] 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.
[21] Hajriansyah, “Identifikasi Jenis Rempah-Rempah MenggunakanMetode CNN Berbasis Android,” *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 8, no. 1, pp. 223–232, 2023.
[22] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
[23] I. N. Switrayana, D. Ashadi, H. Hairani, and A. Aminuddin, “Sentiment Analysis and Topic Modeling of Kitabisa Applications using Support Vector Machine (SVM) and Smote-Tomek Links Methods,” *Int. J. Eng. Comput. Sci. Appl.*, vol. 2, no. 2, pp. 81–91, 2023, doi: 10.30812/ijecsa.v2i2.3406.
[24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” vol. 19, no. 2, pp. 107–117, 2015, [Online]. Available: http://image-net.org/challenges/LSVRC/2015/.
[25] M. Tan and Q. V Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks Mingxing,” 2019.
[26] F. Chollet and G. Inc., “Xception: Deep Learning with Depthwise Separable Convolutions,” *Comput. Vis. Pattern Recognit.*, pp. 1251–1258, 2017.
[27] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” no. April 2017, 2017, [Online]. Available: https://www.researchgate.net/publication/316184205.
[28] I. N. Switrayana, S. Hadi, and N. Sulistianingsih, “A Robust Gender Recognition System using Convolutional Neural Network on Indonesian Speaker,” vol. 13, pp. 1008–1021, 2024.
[29] I. N. Switrayana and N. U. Maulidevi, “Collaborative Convolutional Autoencoder for Scientific Article Recommendation,” *Proc. - 2022 9th Int. Conf. Inf. Technol. Comput. Electr. Eng. ICITACEE 2022*, pp. 96–101, 2022, doi: 10.1109/ICITACEE55701.2022.9924130.
[30] 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.