Vol. 12 No.1 Tahun 2026, pp 44-56
Jurnal Teknologi dan Manajemen Informatika
http:// http://jurnal.unmer.ac.id/index.php/jtmi
P-ISSN: 1693-6604 E-ISSN: 2580-8044

Traditional Snack Image Classification Using ResNet50 and EfficientNetB0: A Comparative Study for Smart Culinary Tourism in Jakarta

Muhammad Zaki Alfadilah[1], Guntur Eka Saputra[2], Armaini Akhirson[3]

[1,2] Informatika, Fakultas Teknologi Industri, Universitas Gunadarma, Indonesia
[3] Pariwisata, Fakultas Sastra dan Budaya, Universitas Gunadarma, Indonesia

Article Info

Article History
Received: 08-06-2026
Revised : 23-05-2026
Accepted: 26-06-2026

Keywords
Convolutional Neural Network;
Transfer Learning;
Traditional Snack Classification;
Smart Tourism;
Deep Learning;

Corresponding Author
Guntur Eka Saputra
Universitas Gunadarma
guntur@staff.gunadarma.ac.id

ABSTRACT

This study is motivated by the significant potential of Jakarta's traditional culinary sector in tourism, which is not yet supported by interactive digital identification media for tourists. This research aims to implement a Convolutional Neural Network (CNN) algorithm to automatically classify traditional snacks as part of a smart tourism system. A primary dataset consisting of 2,600 images across 14 snack classes was collected from Ciracas Market and Setu Babakan in April 2025. Data preprocessing was conducted using augmentation techniques, followed by model development using transfer learning with two architectures, ResNet50 and EfficientNetB0. The experimental results show that ResNet50 achieved the highest accuracy of 99.08% with a loss value of 0.0239, outperforming EfficientNetB0, which decreased to 94.48% accuracy. The best model was deployed in a Streamlit-based web application that provides interactive information, including the name, history, and ingredients of traditional snacks. This system facilitates tourists in recognizing Jakarta's local culinary heritage and supports the implementation of smart tourism.

INTRODUCTION

The rapid advancement of Information and Communication Technology (ICT) has accelerated the adoption of intelligent computing systems across various sectors, including smart tourism [1]. The importance of this sector is reinforced by data from the Central Statistics Agency, which shows a consistent increase in tourist mobility each year [2]. This trend is aligned with government initiatives promoting digital transformation through the Tourism 5.0 agenda, as outlined in the National Medium-Term Development Plan (RPJMN) 2025–2029, as well as the fourth pillar of Asta Cita, which emphasizes strengthening science and technology development to achieve national independence [3].

Jakarta's traditional culinary sector represents not only cultural heritage but also a key driver of the regional creative economy. The integration of artificial intelligence in tourism has demonstrated significant potential to enhance tourist experiences through personalized and interactive services [4], [5]. However, in practice, many culinary tourism destinations in Jakarta have not yet implemented interactive digital tools to provide comprehensive information to visitors. Tourists often encounter limitations in existing applications when identifying specific types of traditional snacks, leading to inaccuracies in recognizing local cultural products [6].

This is an open access article under the CC BY-SA license.
DOI: https://doi.org/10.26905/jtmi.v12i1.17000

This information gap limits the optimal utilization of culinary tourism potential. Moreover, existing systems often exhibit low accuracy in distinguishing the diverse textures and visual characteristics of traditional foods, indicating the need for more precise and adaptive technological solutions [7].

Previous studies have explored food classification using deep learning approaches. Research based on the FoodX-251 dataset has demonstrated success in fine-grained food classification [8]. Similarly, the ChineseFoodNet dataset has been utilized for large-scale Chinese food recognition [9], while the MyFood system applies image segmentation techniques for nutritional monitoring [10]. CNNs have been widely recognized as the state-of-the-art approach in image classification tasks [11], leveraging deep learning mechanisms that mimic human neural systems [12]. Furthermore, transfer learning using pre-trained models on ImageNet has proven effective in accelerating convergence and improving performance on small to medium-sized datasets [13], [14]. Comparative studies also highlight that the selection of CNN architecture significantly influences classification accuracy [15].

However, most existing studies rely on global culinary datasets, which are not fully representative when applied to Jakarta's traditional snacks that possess unique visual characteristics. In addition, class imbalance and limited dataset size remain critical challenges, where data augmentation plays an essential role in improving model generalization [16], [17]. Therefore, this study aims to evaluate and compare the performance of ResNet50 and EfficientNetB0 for classifying Jakarta's traditional snacks using a transfer learning approach. ResNet50 offers advantages in training deep networks through residual learning mechanisms [18], while EfficientNetB0 provides improved scalability and computational efficiency [19]. Model performance is evaluated using standard metrics, including accuracy, precision, recall, and F1-score [20].

This study proposes a deep learning-based classification system for Jakarta traditional snacks using a primary dataset consisting of 14 locally collected snack categories. Unlike previous studies, this research emphasizes cultural specificity and practical implementation within the smart tourism domain. This study contributes by providing a comparative evaluation of several deep learning architectures to identify the most effective model for traditional snack classification and by implementing the best-performing model in a Streamlit-based web application. The application delivers interactive information about traditional snacks, including their names, history, and ingredients. The proposed system is expected to enhance tourist experience while supporting the preservation and promotion of Jakarta's culinary heritage through digital innovation [21].

METHODS

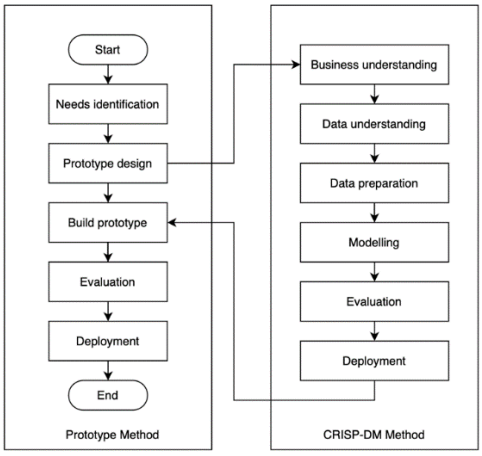


Figure 1. Research Methods

This study employed a hybrid approach by combining the Cross-Industry Standard Process for Data Mining (CRISP-DM) and the prototype development method. CRISP-DM was

used to guide the process of developing the image classification model through six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. In parallel, the Prototype method was applied during the development of the Streamlit-based web application to support iterative system design and feature improvement. Through this approach, the research aims to develop an accurate classification model for Jakarta traditional snack images and implement it in an interactive application that supports smart culinary tourism.

Dataset and Data Collection

The dataset used in this study is a primary dataset collected in April 2025. Data selection focused on 14 types of Jakarta's traditional snacks, such as Kembang Goyang, Kerak Telor, and various traditional cakes, which were chosen based on their popularity in local culinary centers, including Ciracas Market and Setu Babakan. A total of more than 2,600 images were independently captured using a 200 MP smartphone camera at an approximate distance of 30 cm. The dataset was subsequently expanded to 2,895 images through data augmentation, covering all 14 snack classes. This specification was designed to ensure that texture details and visual characteristics of each object were clearly captured to support classification accuracy.

Data Processing Procedure

The initial stage involves data preparation, where all images were resized to 256 × 256 pixels to standardize the input dimensions for the model. Furthermore, data augmentation techniques including rotation, flipping, zooming, shifting, and shearing were applied to classes with fewer than 200 images. This approach aims to increase data variability and prevent overfitting. The dataset was then divided into training, validation, and testing sets using a ratio of 70:15:15, resulting in 2,027 training images, 433 validation images, and 435 testing images. Representative samples of the dataset are illustrated in Table 1.

Table 1. Sample images from 14 traditional snack classes highlighting variations in texture, color, and shape

No	Class Name	Sample Image	Number of Images	Number of Augmented Images
1	Kue lumpur		203 Images	203 Images
2	Kue cucur		179 Images	200 Images
3	Kue pancong		209 Images	209 Images

4	Ongol ongol		206 Images	206 Images
5	Kue gemblong		206 Images	206 Images
6	Kue cente		76 Images	200 Images
7	Uli bakar		210 Images	210 Images
8	Putu mayang		121 Images	200 Images
9	Kue rangi		202 Images	202 Images
10	Kue wajik		157 Images	200 Images

11	Kembang goyang		211 Images	211 Images
12	Kerak telor		216 Images	216 Images
13	Selendang mayang		215 Images	215 Images
14	Kue cubit		217 Images	217 Images

System Architecture and Hyperparameters

The proposed system is based on CNNs by comparing two pre-trained models: ResNet50 and EfficientNetB0. ResNet50 is selected due to its capability to address the vanishing gradient problem through residual learning mechanisms with 50 convolutional layers, while EfficientNetB0 is employed as a comparison model in terms of network scaling efficiency with approximately 5.3 million parameters. Both models utilize pre-trained weights from ImageNet. After extracting features from the base models, several additional layers were incorporated, including Global Average Pooling, Batch Normalization, Dropout, and a Dense layer with ReLU activation. The final classification layer uses Softmax activation to classify the images into 14 categories.

Table 2. Model Hyperparameters Configuration

Hyperparameters	Value
Optimizer	Adam
Input Size	256 × 256 pixel (RGB)
Learning Rate (Transfer Learning)	10^{-4}
Learning Rate (Fine-Tuning)	10^{-5}
Batch Size	32
Maximum Epochs	30
Early Stopping (patience)	5

Activation (Hidden)	ReLU
Activation (Output)	Softmax
Regularization	Dropout + Batch Normalization
Pre-trained Weights	ImageNet

Evaluation Metrics

Model performance in classifying traditional snack images was evaluated using metrics derived from the confusion matrix, including accuracy, precision, and recall. Accuracy is used to measure the overall effectiveness of the model in correctly classifying all instances, both positive and negative classes. It is calculated as the ratio of correctly predicted observations to the total number of observations, as defined in the accuracy equation.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

Furthermore, precision is used to evaluate the accuracy of the model in predicting positive instances by comparing the predicted results with the actual positive class labels. It quantifies the proportion of true positive predictions among all predicted positive instances.

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

In addition to precision, recall (also referred to as sensitivity) is used to measure the model's ability to correctly identify all relevant instances within each class. It reflects how well the model captures actual positive cases.

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

In these evaluations, True Positive (TP) and True Negative (TN) indicate correct classifications, whereas False Positive (FP) and False Negative (FN) indicate misclassifications produced by the model.

The F1-score is used to provide a balanced measure between precision and recall.

$$F1\ Score = \frac{2TP}{(2TP + FP + FN)} \quad (4)$$

RESULTS AND DISCUSSION

Model Training Analysis

Model training was conducted using two primary architectures, namely EfficientNetB0 and ResNet50, by splitting the dataset into training and validation sets. Based on the experimental results, both models demonstrated strong capability in learning and recognizing the visual features of 14 types of Jakarta's traditional snacks. Model performance was monitored through accuracy curves across training epochs.

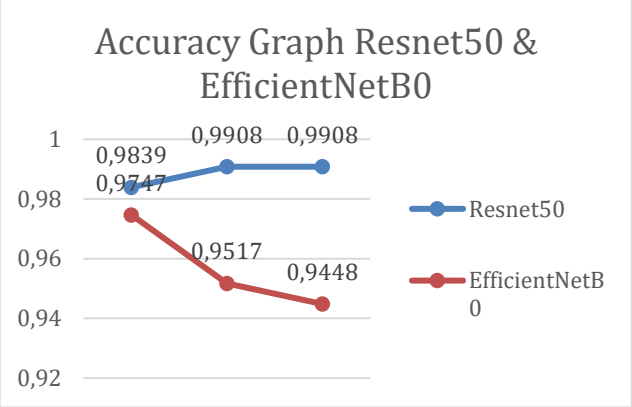


Figure 2. Model Accuracy Results

The results indicate that ResNet50 achieves faster and more stable convergence compared to EfficientNetB0 on the given dataset. The validation accuracy remains consistently high throughout the training process, suggesting that the model does not experience significant overfitting and is able to generalize effectively to unseen image data. This performance can be attributed to the residual learning mechanism in ResNet50, which facilitates better gradient flow and enables deeper network training without degradation. In contrast, although EfficientNetB0 is designed to optimize computational efficiency through compound scaling, it may require further hyperparameter tuning to achieve comparable performance on this dataset. This result suggests that ResNet50 is more robust in handling visual variability in small to medium-sized datasets. Furthermore, the consistency between training and validation performance indicates that the proposed model maintains robustness and stability during the learning process, making it suitable for real-world implementation in smart tourism applications.

Model Performance Comparison

The experiment was conducted by comparing two CNN architectures, namely ResNet50 and EfficientNetB0. The evaluation was carried out over three iterations to assess the stability, consistency, and reliability of each model in classifying 14 types of Jakarta's traditional snacks. This iterative approach aims to ensure that the observed performance is not influenced by random variation during the training process.

Table 3. The Comparative Performance Results of ResNet50 and EfficientNetB0 across three Experimental Iterations

Model	Accuracy per Iteration-1	Accuracy per Iteration - 2	Accuracy per Iteration - 3
ResNet50	0.9839	0.9908	0.9908
EfficientNetB0	0.9747	0.9517	0.9448

Table 4. Comparison of Loss Values for ResNet50 and EfficientNetB0

Model	Loss per Iteration -1	Loss per Iteration-2	Loss per Iteration -3
ResNet50	0.0456	0.0400	0.0239
EfficientNetB0	0.1411	0.2067	0.2489

Table 5. Paired T-Test Results of ResNet50 and EfficientNetB0

Model Comparison	t- Statistic	P- Value	Significance ($\alpha = 0.05$)	Conclusion
ResNet50 vs EfficientNetB0	2.783	0.108	Not Significant ($p > 0.05$)	No statistically significant difference detected; ResNet50

	demonstrates consistently superior practical performance across all iterations

As shown in Tables 3, 4, and 5, the ResNet50 model demonstrates a consistently improving performance trend, achieving the highest accuracy of 99.08% at the third iteration, accompanied by a decreasing loss value of 0.0239. In contrast, EfficientNetB0 shows a gradual decline in performance, where accuracy drops from 97.47% in the first iteration to 94.48% in the third iteration, while the loss value increases to 0.2489. This pattern indicates that ResNet50 is more stable and robust in learning complex visual features of traditional snack images. The decreasing loss and increasing accuracy suggest effective model convergence, whereas the opposite trend observed in EfficientNetB0 may indicate suboptimal generalization or sensitivity to dataset variability.

To statistically validate these findings, a paired t-test was conducted across three experimental iterations. The results yielded a t-statistic of 2.783 and a p-value of 0.108. Although the p-value exceeds the significance threshold of 0.05, which may be attributed to the limited number of iterations (n=3), the consistent accuracy gap observed across all iterations provides practical evidence of ResNet50's superior performance. These findings suggest that the performance difference, while not reaching statistical significance due to sample size constraints, is consistent and meaningful in the context of traditional snack classification.

Hyperparameter Configuration

To optimize model performance during the training process, a set of hyperparameters was determined based on the most stable experimental results. The training process employed the Adam optimizer, which is widely recognized for its efficiency in handling large-scale image data. The learning rate was adjusted dynamically, starting at 10^{-4} during the transfer learning phase to adapt the pre-trained ImageNet weights to the traditional snack dataset, and subsequently reduced to 10^{-5} during the fine-tuning stage to achieve smoother convergence. The batch size was set to 32, with a maximum of 30 training epochs. In addition, an Early Stopping mechanism was applied with a patience parameter of 5 to automatically terminate training when no improvement in validation loss was observed, thereby preventing overfitting.

Classification Matrix Evaluation

To gain a deeper understanding of model performance across each class, an evaluation was conducted using a classification report on the best-performing model, ResNet50. ResNet50 employs a residual learning mechanism to address the vanishing gradient problem in very deep neural networks. Unlike conventional architectures, ResNet50 utilizes identity shortcut connections (skip connections), allowing gradients to flow directly across layers without significant information degradation. The architecture consists of 50 layers, including convolutional layers, bottleneck blocks, and a global average pooling layer before being passed to a fully connected layer. In this implementation, the model accepts input images with a resolution of 256 × 256 pixels, enabling the extraction of complex features from the textures and shapes of Jakarta's traditional snacks.

To evaluate the model's performance, a classification matrix was employed, incorporating precision, recall, and F1-score metrics to ensure balanced recognition across all snack classes.

Table 6. Classification Report ResNet50

Class	Precision	Recall	F1-Score	Support
kembang_goyang	1.00	1.00	1.00	32
kerak_telor	1.00	1.00	1.00	32
kue_cente	1.00	0.97	0.98	30

kue_cubit	1.00	1.00	1.00	33
kue_cucur	1.00	1.00	1.00	30
kue_gemblong	0.94	1.00	0.97	31
kue_lumpur	1.00	0.97	0.98	31
kue_pancong	0.94	1.00	0.97	31
kue_rangi	1.00	1.00	1.00	30
kue_wajik	1.00	1.00	1.00	30
ongol_ongol	1.00	1.00	1.00	31
putu_mayang	1.00	1.00	1.00	30
selendang_mayang	1.00	0.91	0.95	32
uli_bakar	1.00	1.00	1.00	32
Accuracy	-	-	0.99	435
Macro Avg	0.99	0.99	0.99	435
Weighted Avg	0.99	0.99	0.99	435

The classification report provides a comprehensive evaluation of the model's performance for each class, highlighting its ability to maintain high precision and recall across most categories. This indicates that the model is not biased toward specific classes and demonstrates strong generalization capability in multiclass classification tasks.

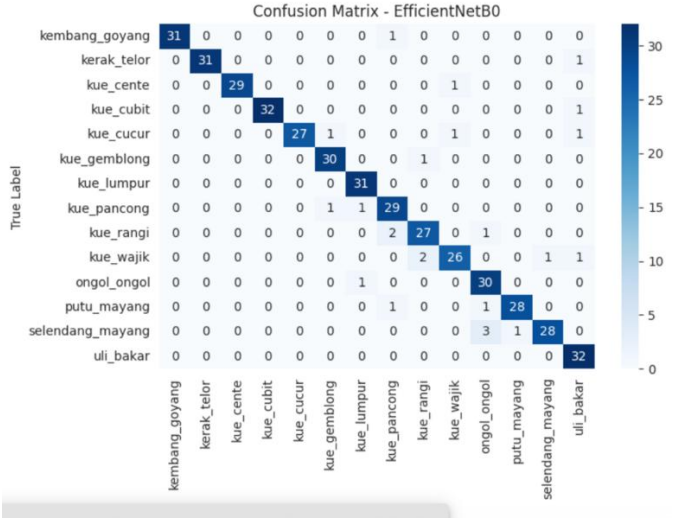


Figure 3. Confusion Matrix

The results presented in Table 5 and Figure 3 indicate that the ResNet50 model achieves near-perfect classification performance across almost all traditional snack classes. The high average recall value of 0.99 confirms the model's strong capability in correctly identifying relevant instances while minimizing false negatives, even in cases where visual similarity between classes is high. Nevertheless, the *selendang_mayang* and *kue_gemblong* classes exhibit slightly lower F1-scores of 0.95 and 0.97, respectively. This performance gap can be attributed to inherent visual similarities between these classes and other snack categories within the dataset. Specifically, *selendang_mayang* was misclassified as *ongol_ongol* in three instances, which is visually explainable by the shared characteristics between the two classes:

both snacks feature soft, layered textures with similar color profiles derived from natural ingredients such as pandan and palm sugar, producing comparable green, white, and brown tones.

Furthermore, both are typically presented as flat, sliced portions, making their visual boundaries difficult to distinguish under varying lighting or camera angles. Similarly, one instance of *selendang_mayang* was misclassified as *putu_mayang*, as both snacks share green pandan-based coloring and a similarly soft, multi-layered appearance when photographed from above. In the case of *kue_gemblong*, one misclassification occurred toward *kue_pancong*, likely due to their comparable golden-brown surface coloration resulting from frying and grilling processes, as well as their rounded shape profiles that may appear visually convergent at certain capture angles. These findings suggest that visual similarity in color, texture, and shape among traditional snack classes remains a key challenge in fine-grained food classification, and may be addressed in future work through more targeted augmentation strategies or attention-based mechanisms that emphasize discriminative local features.

These findings highlight the effectiveness of the proposed approach, where the integration of transfer learning, fine-tuning strategies, and well-optimized hyperparameter settings contributes significantly to improving feature discrimination and overall model performance. This result demonstrates the robustness of ResNet50 in handling multiclass image classification tasks with subtle interclass variations.

Table 7. Comparative Analysis of Classification Performance Between Proposed Method and Prior Work

Model		Accuracy	Loss	Description
[7]	ResNet50	62.34	1.52	Trained on the general Food-101 dataset without domain-specific fine-tuning, resulting in lower accuracy due to high visual variability across food categories.
	EfficientNetB0	75.14	0.6	Outperformed ResNet50 on Food-101 through compound scaling efficiency, yet still limited by the diversity and generality of the benchmark dataset.
Our method	ResNet50	0.9908	0.0295	Trained on a domain-specific primary dataset of 14 Jakarta traditional snacks with transfer learning, fine-tuning, and augmentation, achieving superior accuracy and stable convergence.
	EfficientNetB0	0.9448	0.2489	Applied under the same training conditions as ResNet50, but exhibited declining accuracy across iterations, suggesting sensitivity to dataset variability and the need for further hyperparameter tuning.

The comparative analysis presented in Table 7 specifically references the study by Liu [7] as a benchmark due to its methodological similarity, as both studies employ ResNet50 and EfficientNetB0 architectures for food image classification tasks. However, notable differences in experimental outcomes can be observed. Liu [7] reported that EfficientNetB0 outperformed ResNet50 on the general Food-101 dataset, achieving 75.14% accuracy compared to 62.34%. In contrast, the present study yields the opposite trend, where ResNet50 achieves a substantially higher accuracy of 99.08%, outperforming EfficientNetB0 at 94.48%. This reversal suggests that architectural superiority is not absolute but rather dataset-dependent.

The Food-101 dataset encompasses a broad and highly diverse range of food categories with significant inter-class visual variability, a condition that favors EfficientNetB0's compound scaling efficiency. Conversely, the domain-specific nature of Jakarta's traditional

snack dataset, combined with the application of targeted augmentation, fine-tuning strategies, and optimized hyperparameter configurations, appears to favor ResNet50's residual learning mechanism, which demonstrates stronger feature discrimination capability under more controlled visual conditions. These findings highlight that the choice of classification architecture should be guided not only by general benchmark performance but also by the visual characteristics and domain specificity of the target dataset.

Web Application Implementation

After obtaining the best-performing model, the system was deployed as a web-based application using the Streamlit framework. The application was evaluated by uploading randomly selected images of traditional snacks to assess its real-time responsiveness and prediction accuracy.

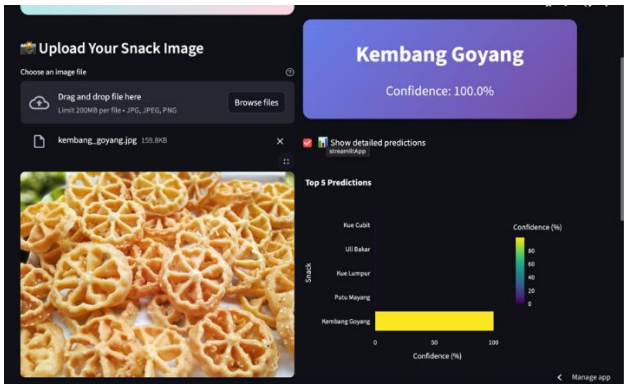


Figure 4. User Interface of Traditional Snack Classification Application



Figure 5. User Interface of Historical and Recipe Information

The application not only presents classification results but also integrates contextual information, such as the history and ingredients of each traditional snack. This additional feature enhances user engagement and provides educational value, making the system highly relevant for smart tourism applications. By combining image recognition with cultural information, the proposed system contributes to promoting and preserving Jakarta's culinary heritage through digital innovation.

CONCLUSIONS AND RECOMMENDATIONS

This study successfully implemented a CNN approach for classifying 14 types of traditional Jakarta snacks with high performance. Based on three experimental iterations, the ResNet50 model combined with transfer learning and fine-tuning demonstrated superior performance compared to EfficientNetB0, achieving the highest accuracy of 99.08% and a loss value of 0.0239. The application of data augmentation, batch normalization, dropout, and early stopping effectively contributed to model stability and prevented overfitting.

Furthermore, the deployment of the model into a web-based application using Streamlit and Hugging Face demonstrates the practical applicability of artificial intelligence in supporting smart culinary tourism. The system not only enables automatic classification but also provides

educational value by presenting historical and ingredient-related information interactively. However, this study is limited by the relatively small dataset size and its focus on a specific regional culinary domain. Future work is recommended to expand the dataset with more diverse traditional food categories and to optimize the model for mobile-based deployment. Additionally, integrating geolocation (GPS) features to identify nearby food vendors could significantly enhance the user experience in real-world tourism scenarios.

REFERENCES

[1] U. Gretzel, M. Sigala, Z. Xiang, and C. Koo, "Smart tourism: foundations and developments," *Electronic Markets*, vol. 25, no. 3, pp. 179–188, 2015. doi: https://doi.org/10.1007/s12525-015-0196-8

[2] Badan Pusat Statistik, "Jumlah perjalanan wisatawan nusantara menurut provinsi tujuan (perjalanan)," 2025. [Online]. Available: https://www.bps.go.id

[3] Kementerian Sekretariat Negara, "Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2025–2029 dan Program Asta Cita," *Jakarta, 2025*. [Online]. Available: https://www.setneg.go.id/baca/index/rpjmn_2025_2029_fondasi_awal_wujudkan_visi_indonesia_emas_2045

[4] S. Wang, Q. Wang, Q. Cui, and T. Lan, "Artificial Intelligence in Tourism: A Systematic Literature Review and Future Research Agenda," *Sustainability*, vol. 17, no. 20, p. 9080, 2025. doi: https://doi.org/10.3390/su17209080

[5] D. Buhalis and M. Amaranggana, "Smart tourism destinations," in *Information and Communication Technologies in Tourism*, 2014, pp. 553–564. doi: https://doi.org/10.1007/978-3-319-03973-2_40

[6] U. Gretzel, H. Werthner, C. Koo, and C. Lamsfus, "Conceptual foundations for understanding smart tourism ecosystems," *Computers in Human Behavior*, vol. 50, pp. 558–563, 2015. doi: https://doi.org/10.1016/j.chb.2015.03.043

[7] Y. Liu, "Automatic food recognition based on efficientnet and ResNet," *Journal of Physics: Conference Series*, vol. 2646, p. 012037, 2023. doi: https://doi.org/10.1088/1742-6596/2646/1/012037

[8] P. Kaur, K. Sikka, W. Wang, S. Belongie, and A. Divakaran, "FoodX-251: A dataset for fine-grained food classification," *arXiv preprint arXiv:1907.06167*, 2019. doi: https://doi.org/10.48550/arXiv.1907.06167

[9] X. Chen, Y. Zhu, H. Zhou, L. Diao, and D. Wang, "ChineseFoodNet: A large-scale image dataset for Chinese food recognition," *arXiv preprint* arXiv:1705.02743, 2017, doi: https://doi.org/10.48550/arXiv.1705.02743

[10] C. N. C. Freitas, F. R. Cordeiro, and V. Macario, "MyFood: A food segmentation and classification system to aid nutritional monitoring," *in 2020 33rd SIBGRAPI*, 2020, pp. 234–241. doi: https://doi.org/10.1109/SIBGRAPI51738.2020.00039

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 436–440, 2012.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. doi: https://doi.org/10.1038/nature14539

[13] N. Abou Baker, N. Zengeler, and U. Handmann, "A Transfer Learning Evaluation of Deep Neural Networks for Image Classification," *Machine Learning and Knowledge Extraction*, vol. 4, no. 1, pp. 22–41, 2022. doi: https://doi.org/10.3390/make4010002

[14] V. S. Mahalle, N. M. Kandoi, and S. B. Patil, "Transfer Learning by Fine-Tuning Pre-trained Convolutional Neural Network Architectures for Image Recognition," in *Data Science and Big Data Analytics*, IDBA 2023, Springer, Singapore, 2024. doi: https://doi.org/10.1007/978-981-99-9179-2_21

[15] J. H. Dewan, R. Das, S. D. Thepade, H. Jadhav, N. Narsale, A. Mhasawade, and S. Nambiar, "Image Classification by Transfer Learning using Pre-Trained CNN Models," in 2023 *Int. Conf. on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*, IEEE, 2023. doi: https://doi.org/10.1109/RAEEUCCI57140.2023.10134069

[16] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, pp. 1–48, 2019. doi: https://doi.org/10.1186/s40537-019-0197-0

[17] K. Alomar, H. I. Aysel, and X. Cai, "Data Augmentation in Classification and Segmentation: A Survey and New Strategies," *Journal of Imaging*, vol. 9, no. 2, p. 46, 2023. doi: https://doi.org/10.3390/jimaging9020046

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: https://doi.org/10.1109/CVPR.2016.90

[19] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Machine Learning (ICML)*, 2019, pp. 1–10.

[20] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," International Journal of Data Mining and Knowledge Management Process, vol. 5, no. 2, p. 1, 2015. doi: https://doi.org/10.5121/ijdkp.2015.5201

[21] C. E. Mbonu, K. Anigbogu, D. Asogwa, and T. Belonwu, "An Explorative Analysis of SVM Classifier and ResNet50 Architecture on African Food Classification," arXiv preprint arXiv:2505.13923, 2025, doi: https://doi.org/10.48550/arXiv.2505.13923