Development of a Mobile Application Using Convolutional Neural Networks for Recognizing Indonesian Traditional Snacks

Njoto Benarkah[1*], Joko Siswantoro[2], Muhammad Ikhsan[3]

[1,2,3]Informatics Engineering, Faculty of Engineering, University of Surabaya, Surabaya, East Java, Indonesia
E-mail: [1*]benarkah@staff.ubaya.ac.id, [2]joko_siswantoro@staff.ubaya.ac.id, [3]iksan8991@gmail.com

(Received: 22 Apr 2025, revised: 9 May 2025, accepted: 12 May 2025)

Abstract

Indonesian traditional snacks constitute a vital element of the country's cultural heritage. However, growing modernization has contributed to a decline in public familiarity, particularly among younger generations. This study presents a mobile-based image classification desigend to automatically recognize Indonesian traditional snacks using convolutional neural networks (CNNs). A dataset of 3,240 images across 16 snack categories was collected using a smartphone camera. Five CNN architectures, which are, AlexNet, EfficientNetV2M, MobileNetV2, ResNet50V2, and VGG19, were evaluated for classification performance. MobileNetV2 achieved the highest accuracy and F1-score, both reaching 100%. The final model was deployed in a mobile application environment, with the backend developed using Flask and integrated into the Android platform. This research work demonstrates the potential of lightweight CNN models in preserving cultural knowledge through accessible mobile technology.

Keywords: CNN, Traditional Snack, Deep Learning, Mobile Application, Classification.

I. INTRODUCTION

The safeguarding of intangible cultural heritage (ICH) has evolved significantly since the 2003 UNESCO Convention, emphasizing not the static preservation of traditions, but their dynamic transmission and adaptation across generations. Food-related heritage, though once contested due to its commercial and mutable nature, has become a prominent focus within ICH, with examples such as the dishes from French region, Mediterranean Diet and Traditional Mexican Cuisine symbolizing both cultural identity and shared human values [1], [2]. As digital and creative tools become central to heritage preservation, initiative like the EURICA projects show how digitizing ritual recipes and utilizing social media can enhance the accessibility and longevity of gastronomic heritage [3]. In rural Mediterranean regions, creative and digital tools are being used to reconnect communities with their food traditions, promote sustainable development, and preserve cultural identities [4]. These efforts highlight the growing role of food in cultural representation, identity formation, and tourism development, principles that increasingly relevant for countries like Indonesia.

Indonesian traditional snacks represent a significant part of the nation's cultural heritage, reflecting its diverse ethnicities, languages, and customs. Each region has unique snack varieties, making them attractive to both locals and tourists [5]. Preserving and promoting these culinary assets is crucial in positioning Indonesia as a gastronomic tourism destination [6]. A study in a West Java village identified 35 traditional snacks, highlighting the richness of local culinary traditions [7]. Despite their delicious taste and visually appealing presentation, many traditional snacks remain underrecognized, particularly among the younger generation.

A preliminary survey was conducted to assess the familiarity of young Indonesians with traditional snacks. A total of 39 respondents aged 15-25, from Jakarta, Surabaya, and Makassar, participated in an online questionnaire distributed via Google Forms. The age range was selected to represent the younger generation, who are often less exposed to traditional culinary knowledge. The survey presented 30 images of various traditional snacks, primarily from Java island, and asked participants to identify each of snack from multiple-choice options. Only six snacks had identification rates above 70%, such as *klepon, lumpia, onde-onde, bolu kukus, cucur*, and *kue lapis*. Furthermore, 87% of respondents reported difficulty recognizing most of the snacks. Efforts to preserve these snacks often rely on local initiatives such as community festivals and hands-on teaching, which have limited reach. Technology offers a promising avenue to promote traditional snacks both nationally and globally.

CNNs are popular deep learning models widely used in food image classification task, including ResNet50, AlexNet, VGG-16, and InceptionV3 [8], [9]. An existing research work on CNN-based method had been conducted to recognized 10

DOI: 10.34148/teknika.v14i2.1236

CC BY SA

types of fruits and estimating nutritional values using YOLOv5 [10]. Recent studies also have employed convolutional neural networks (CNNs) for image-based recognition of traditional Indonesian snacks. Accuracy levels have ranged from 80% using VGG16 with augmentation [11], to 98% using MobileNetV2 on a dataset of five snack types [12]. Another study using MobileNetV2 for eight snack categories achieved 92.5% accuracy [13]. Despite promising results, prior studies were often limited by real-world mobile conditions. A brief overview of the underlying CNN architectures used in these studies provides context for their performance and applicability.

CNN consist of multiple convolutional, pooling, and fully connected layers [14], have become a popular tool for image recognition. For example, AlexNet introduced the ReLU activation function and used five convolutional and three fully connected layers [15]. VGG19, with its 19-layer architecture, improves upon VGG16 in handling complex visual tasks [16]. ResNet50V2 utilizes residual blocks to reduce feature dimensions and improve training efficiency [17], [18]. MobileNetV2 is known for its lightweight architecture, suitable for mobile devices [19]. EfficientNetV2, a more recent architecture, offers fast training and fewer parameters. The research work was classifying 18 traditional foods, EfficientNetV2L achieved 99.4% accuracy on 1,800 images [20].

While prior studies have shown promising results in recognizing Indonesian traditional snacks using CNNs, they have generally been limited in scope, often focusing on fewer snack categories and using datasets not representative of real-world mobile conditions. For example, previous work using MobileNetV2 achieved 92.5% accuracy on eight snack classes [13], while another study achieved 98% using a small-scale dataset of five types [12]. These efforts, though effective, did not incorporate mobile application integration with real-time classification capabilities.

This study addresses these limitations by evaluating five CNN models: AlexNet [21], VGG19 [22], ResNet50V2 [21], MobileNetV2 [23], and EfficientNetV2M [24], on a locally collected dataset comprising 3,240 images across 16 traditional snack categories, captured using a smartphone. The best-performing model is deployed within the application. A locally collected dataset using mobile phone cameras is used to ensure relevance and practical applicability.

II. RESEARCH METHODOLOGY

This research aims to develop an image classification model capable of accurately identifying various traditional Indonesian snacks using deep learning techniques. This research was conducted using Python on a Windows 11 Pro 64-bit platform. The system used an AMD Ryzen Threadripper Pro 3975WX (320 cores) CPU and NVIDIA T600 GPU. The research methodology includes dataset construction, preprocessing, model development and training, performance evaluation, and deployment through a mobile application interface. The research methodological workflow is illustrated in Figure 1.

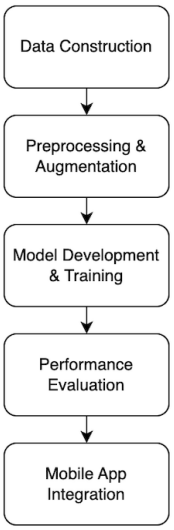


Figure 1. Research Methodological Workflow

A. Dataset Construction

The dataset comprises 3,240 images of 16 types of Indonesian traditional snacks, captured using smartphone cameras. Except for the *cucur* and *serabi solo* classes, each snack class consists of 180 images. *Cucur* and *serabi solo* each contain 360 images. A 70:20:10 split was applied for training, validation, and testing respectively, resulting in 2,252 images for training, 324 for testing, and 664 for validation. Table 1 shows the distribution of images used in each class in the dataset. There are 16 class of different Indonesian traditional snacks, in a from of image, used in this study: c*ucur, klepon, kue lapis, kue lumpur, lumpia, serabi solo, wajik, putu ayu, lanting, getuk, onde-onde, madu mongso, kue tok, nagasari, lemper*, and *gipang*.

Table 1. Traditional Indonesian Snack

No	Snack	#Train	#Test	#Validation
1.	*Cucur*	251	36	73
2.	*Getuk*	125	18	37
3.	*Gipang*	125	18	37
4.	*Klepon*	125	18	37
5.	*Kue Lapis*	125	18	37
6.	*Kue Lumpur*	125	18	37
7.	*Kue Tok*	125	18	37
8.	*Lanting*	125	18	37
9.	*Lemper*	125	18	37
10.	*Lumpia*	125	18	37
11.	*Madu Mongso*	125	18	37
12.	*Nagasari*	125	18	37
13.	*Onde-onde*	125	18	37
14.	*Putu Ayu*	125	18	37
15.	*Serabi Solo*	251	36	73
16.	*Wajik*	125	18	37
Total				3,240

DOI: 10.34148/teknika.v14i2.1236

Example images of each of the Indonesian traditional snack in the dataset are shown in Figure 2. All of images use JPG format and RGB color model.

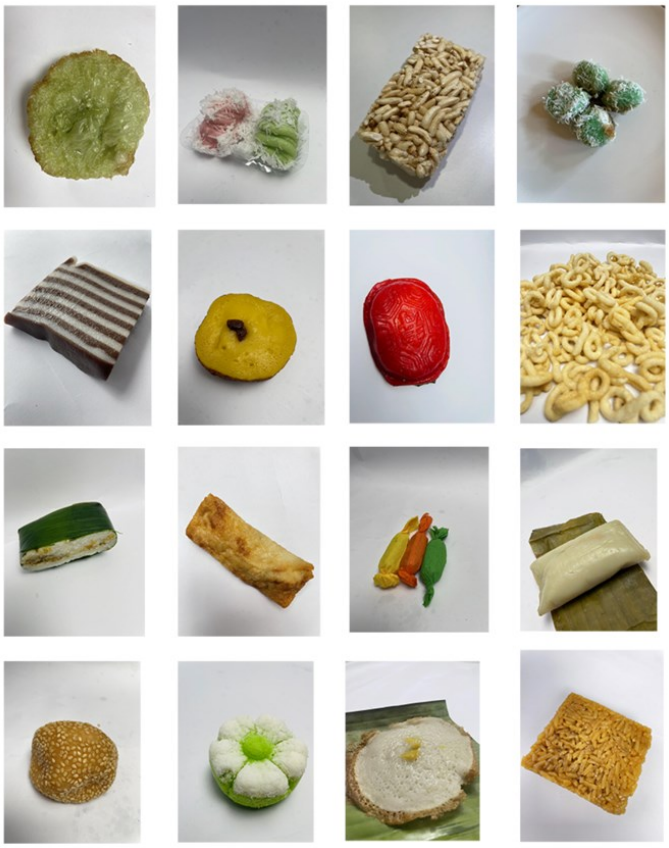


Figure 2. Images of Each Class in the Dataset, From Left to Right: *Cucur, Getuk, Gipang, Klepon, Kue Lapis, Kue Lumpur, Kue Tok, Lanting, Lemper, Lumpia, Madu Mongso, Nagasari, Onde-Onde, Putu Ayu, Serabi Solo*, and *Wajik*

B. Data Preprocessing and Augmentation

Prior to training, the images were normalized to scale pixel values to the [0, 1] range, which accelerates convergence during training. To improve generalization and mitigate overfitting, image augmentation was applied using TensorFlow's ImageDataGenerator, including the following transformations:

- Rotation range: 40 degrees to simulate camera tilt
- Width/height shift: 0.2 to simulate misalignment
- Horizontal flip: True to handle mirrored versions
- Zoom range: 0.2 to simulate varying distances
- Shear range: 0.2 to simulate angular distortions

These augmentations help the model learn invariance to real-world image distortions and contribute to better generalization on unseen data.

The images underwent preprocessing to improve quality and standardization. Techniques applied include denoising, sharpening, and resizing all images to 224×224 pixels.

C. Model Development

Each CNN model was trained and evaluated separately using the same data split (70% training, 10% testing, 20% validation). Hyperparameters were fine-tuned for each model (see Table 2).

Table 2. Hyperparameters Setting for Each Model

M	Dense Layer Nodes	Drop Out Layer Rate	Learning Rate
1	128 – 4,096	0.25 – 0.5	0.000001 – 0.001
2	64 - 512	0.2 – 0.5	0.000001 – 0.001
3	16 - 512	0.0 – 0.5	0.000001 – 0.001
4	16 - 512	0.0 – 0.5	0.000001 – 0.001
5	16 - 512	0.0 – 0.5	0.000001 – 0.001

M = model ID

Additional fine-tuning for EfficientNetV2M includes 50–150 layers, with 2–3 dense/dropout layers applied across all models. The five CNN architectures used in this study are illustrated in Figure 3 through 7. Each model varies in complexity, depth, and number of trainable parameters.

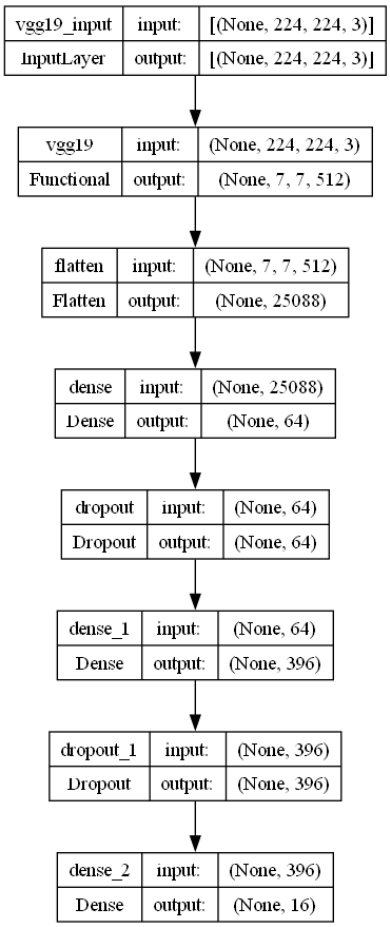


Figure 3. VGG19 Architecture in the Experiment to Classify Indonesian Traditional Snacks

DOI: 10.34148/teknika.v14i2.1236

CC BY SA

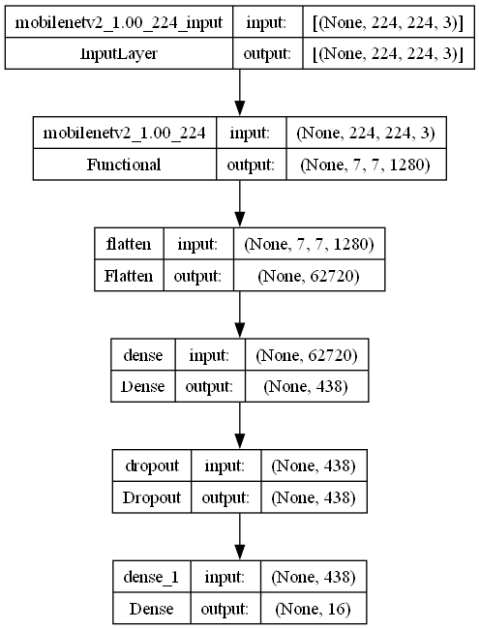


Figure 4. MobileNetV2 Architecture in the Experiment to Classify Indonesian Traditional Snacks

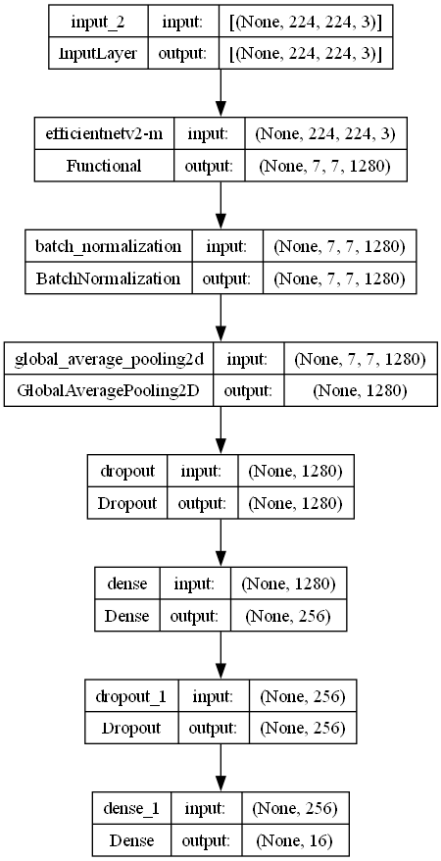


Figure 5. EfficientNetV2M Architecture Used in the Experiment to Classify Indonesian Traditional Snacks

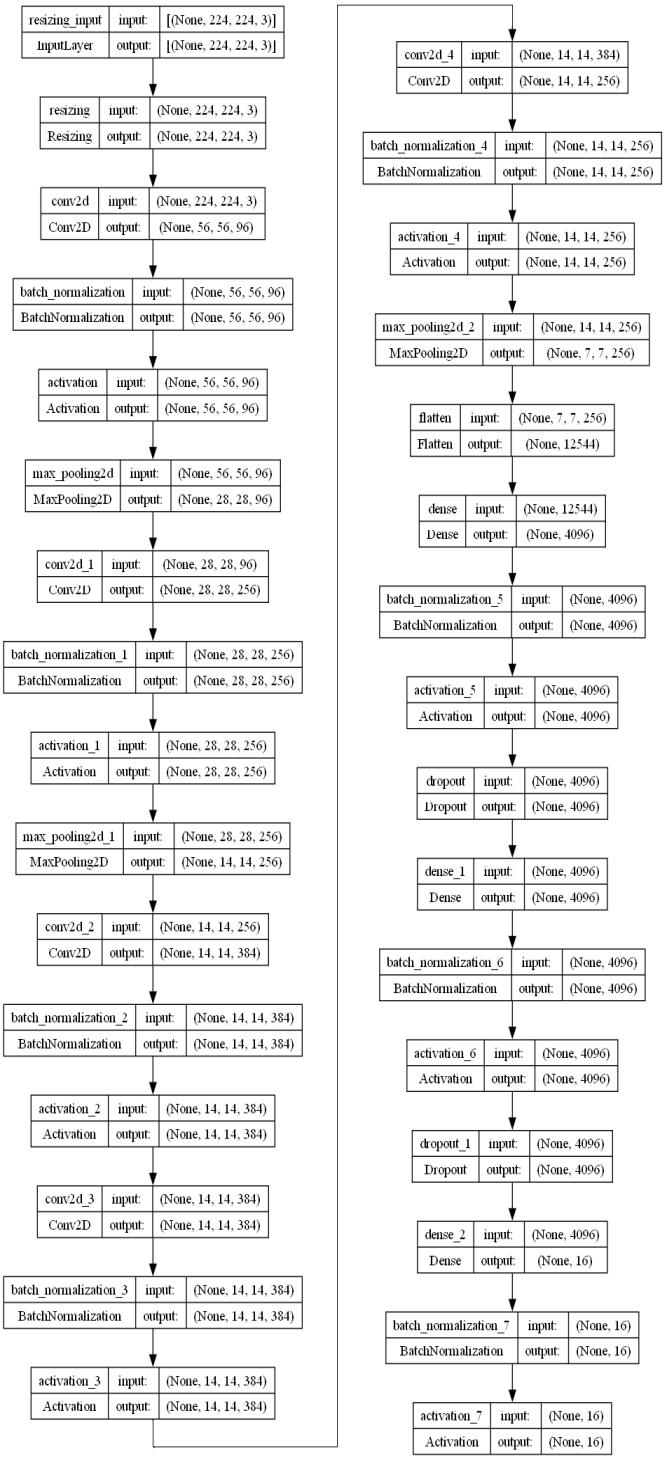


Figure 6. AlexNet Architecture Used in the Experiment to Classify Indonesian Traditional Snacks

DOI: 10.34148/teknika.v14i2.1236

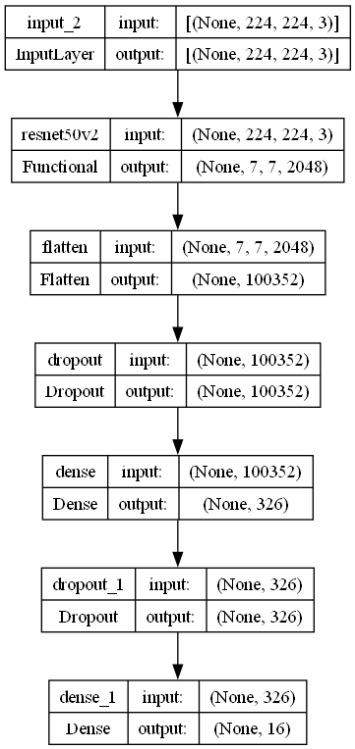


Figure 7. ResNet50V2 Architecture in the Experiment to Classify Indonesian Traditional Snacks

D. Mobile Training Configuration

Each model was trained for up to 30 epochs with a batch size of 32, using the categorical cross-entropy loss function appropriate for multi-class classification tasks. Training was performed using Adam optimizer with an initial learning rate of 0.00003 and categorical cross-entropy loss function was used since this is a multiclass classification. The training process emphasized early stopping to prevent overfitting.

A key component of the training strategy was the implementation of early stopping, which monitored the validation loss during training. If the validation loss did not improve for 5 consecutive epochs, training was halted, and the model's weights were automatically restored to the best-performing state (restore_best_weights=True). This method ensures that the final model used for evaluation retains the optimal performance achieved during training, avoiding overfitting due to excessive epochs.

E. Mobile Application Integration

The best-performing model was integrated into a Flask-based backend server, enabling mobile access through API calls for efficient snack recognition. A choice between MobileNetV2 and ResNet50V2. MobileNetV2 has less number of parameters than ResNet50V2. The mobile application used trained model from MobileNetV2. A custom-built Android mobile application was developed to allow users to classify snack images in real-time. The app features:

- Camera integration for image capture
- On-device inference using the embedded model
- Result display including snack name and confidence score

Inference is performed directly on the device without requiring a network connection, ensuring responsiveness and data privacy. The mobile app was tested on an Android device for inference speed. Application shows a splash screen upon opening followed by the main page as shown on Figure 8.

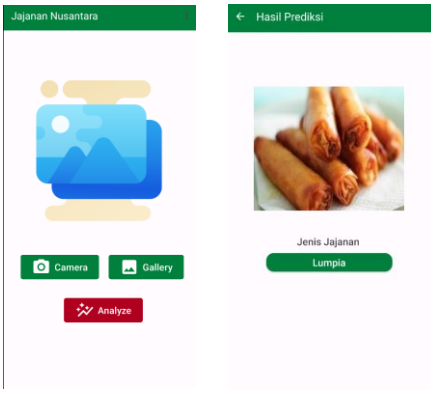


Figure 8. Application Interface When Predicting One of Indonesian Traditional Snack, Lumpia.

III. RESULTS AND DISCUSSION

The training process for each convolutional neural network (CNN) model demonstrated strong convergence behavior, with most models stabilizing in performance between epochs 15 and 30. All five architectures—AlexNet, EfficientNetV2M, MobileNetV2, ResNet50V2, and VGG19—achieved over 90% training accuracy, indicating effective learning from the dataset.

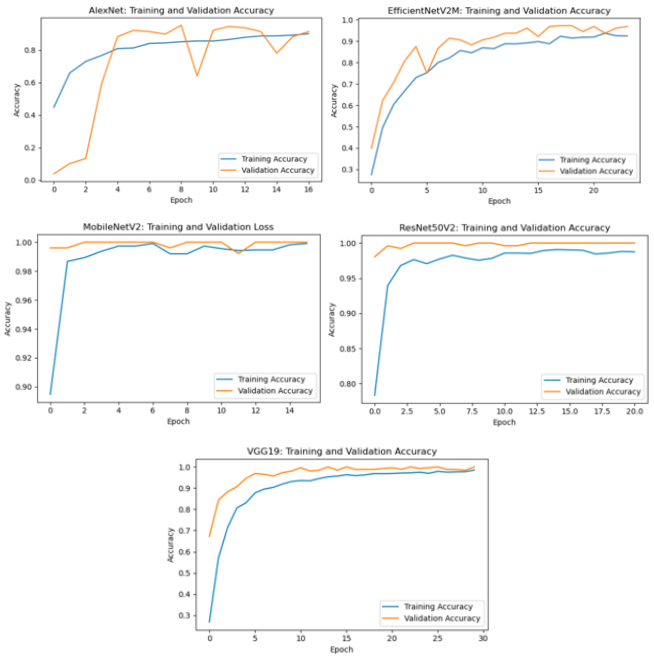


Figure 9. Graphics for Training Accuracy of Each Model Experiment, From Top, Left to Right: AlexNet, EfficientNetV2M, MobileNetV2, ResNetV2, and VGG19

DOI: 10.34148/teknika.v14i2.1236

CC BY SA

Figure 9 illustrates the accuracy and loss curves for all models throughout the training process, showing steady improvement and minimal signs of overfitting. Each model was evaluated based on four performance metrics: accuracy, precision, recall, and F1-score. Table 3 presents the performance evaluation results for each CNN model.

Table 3. Experiment Results of Five CNN Models to Classify Indonesian Traditional Snack

Model	A (%)	P (%)	R (%)	F (%)
AlexNet	91.95	93.81	91.05	90.34
EfficientNetV2M	97.84	98.04	97.84	97.83
MobileNetV2	100.00	100.00	100.00	100.00
ResNet50V2	100.00	100.00	100.00	100.00
VGG19	99.69	99.71	99.69	99.69

A=accuracy, P=precision, R=recall, F=F1-score

Table 4. The number of parameters from each model in the experiments

Model	Total Parameters
AlexNet	72,016,720
EfficientNetV2M	53,487,556
MobileNetV2	29,736,806
ResNet50V2	56,285,110
VGG19	21,662,172

Among all models tested, MobileNetV2 and ResNet50V2 delivered the best overall performance, achieving 100% across all four metrics—accuracy, precision, recall, and F1-score—on the test set. This suggests both models are highly capable of accurately classifying Indonesian traditional snacks from the image dataset used in this study.

On the other hand, EfficientNetV2M also performed exceptionally well, achieving an accuracy of 97.84%, a precision of 98.04%, recall of 97.84%, and an F1-score of 97.83%. These results further support EfficientNetV2M as a robust model for image-based food classification tasks, albeit with a higher parameter count than MobileNetV2.

AlexNet, while historically significant and relatively simple in architecture, yielded an accuracy of 91.95%, with the lowest F1-score among the models at 90.34%. This model had the highest parameter count at over 72 million, which, combined with its lower performance, makes it less suitable for mobile deployment.

VGG19, despite having the smallest number of parameters (21.6 million), achieved moderately high performance with an accuracy above 90%, suggesting it is a viable option for lightweight applications, although still outperformed by MobileNetV2 in both parameter efficiency and classification accuracy.

Table 4 summarizes the total number of parameters for each model. It is evident that MobileNetV2, with only 29.7 million parameters, provides an optimal balance between model size and performance, making it the most suitable candidate for deployment in mobile environments.

Overall, the experimental results confirm that CNN-based models can effectively classify Indonesian traditional snacks with high accuracy. In particular, MobileNetV2 stands out as the ideal model for real-time mobile applications due to its lightweight architecture and perfect performance metrics.

IV. CONCLUSION

The experimental results demonstrate that MobileNetV2 achieves the highest accuracy and F1-score in classifying Indonesian traditional snacks, outperforming other CNN architectures in this study, and it has fewer number of parameters comparing to ResNet50V2. Its lightweight design and excellent performance metrics make it the most suitable model for deployment in an Android application. This mobile application could significantly contribute to promoting Indonesian traditional snacks, especially among younger generations, and help boost gastronomy tourism by increasing awareness of Indonesia's rich culinary heritage. Future work can involve expanding the dataset to include a wider variety of traditional snacks and more images, which would further enhance the model's prediction capabilities. Additionally, modifying the ResNet50V2 architecture to reduce its parameter count could make it another strong candidate for mobile applications, while VGG19 also presents potential, with further optimizations to improve performance. Ultimately, the successful deployment of such an application would provide a tool to preserve and showcase Indonesia's diverse gastronomic traditions, supporting cultural preservation and tourism.

REFERENCES

[1] J. L. Herman and R. Matta, "Food and UNESCO's Intangible Cultural Heritage Journey," in *Oxford Research Encyclopedia of Food Studies*, Oxford University Press, 2025. doi: 10.1093/acrefore/9780197762530.013.145.

[2] V. Cuffy and J. Carr, "Intangible Cultural Heritage: Safeguarding Embodied Creole Cultures," in *Creole Cultures, Vol. 1*, Cham: Springer International Publishing, 2023, pp. 1–18. doi: 10.1007/978-3-031-24275-5_1.

[3] M. Ramazanova, R. S. Cardoso, and I. V. de Freitas, "The Role of Social Media in the Conservation and Safeguard of Gastronomy as Intangible Cultural Heritage," 2024, pp. 89–97. doi: 10.1007/978-3-031-49536-6_10.

[4] E. Del Soldato and S. Massari, "Creativity and digital strategies to support food cultural heritage in Mediterranean rural areas," *EuroMed Journal of Business*, vol. 19, no. 1, pp. 113–137, Mar. 2024, doi: 10.1108/EMJB-05-2023-0152.

[5] S. D. Yubianto, "Understanding The Role of Local Culture and Local Food in Indonesia's Gastronomy Tourism," *Jurnal Hospitality dan Pariwisata*, vol. 9, no. 1, Feb. 2023, doi: 10.30813/jhp.v9i1.4213.

[6] D. Juniarti, "Kearifan Lokal Makanan Tradisional: Tinjauan Etnis Dan Fungsinya Dalam Masyarakat Suku

Pasmah di Kaur," *Bakaba*, vol. 9, no. 2, pp. 44–53, Dec. 2021, doi: 10.22202/bakaba.2021.v9i2.4833.

[7] C. Ningsih and D. Turgarini, "Pelestarian dan Pengembangan Kampung Nikmat Cigugurgirang Sebagai Destinasi Wisata Warisan Budaya Gastronomi Sunda," *Tourism Scientific Journal*, vol. 5, no. 2, pp. 265–275, Jun. 2020, doi: 10.32659/tsj.v5i2.97.

[8] A. Kazi and S. P. Panda, "Determining the freshness of fruits in the food industry by image classification using transfer learning," *Multimed Tools Appl*, vol. 81, no. 6, pp. 7611–7624, Mar. 2022, doi: 10.1007/s11042-022-12150-5.

[9] A. Chaitanya, J. Shetty, and P. Chiplunkar, "Food Image Classification and Data Extraction Using Convolutional Neural Network and Web Crawlers," *Procedia Comput Sci*, vol. 218, pp. 143–152, 2023, doi: 10.1016/j.procs.2022.12.410.

[10] L. Luo, "Research on Food Image Recognition of Deep Learning Algorithms," in *2023 International Conference on Computers, Information Processing and Advanced Education (CIPAE)*, IEEE, Aug. 2023, pp. 733–737. doi: 10.1109/CIPAE60493.2023.00142.

[11] C. Lubis, D. Sumarlie, and T. Handhayani, "Pengenalan Kue Tradisional Indonesia Menggunakan Algoritma Convolutional Neural Network," *Computatio : Journal of Computer Science and Information Systems*, vol. 6, no. 2, pp. 164–171, Feb. 2023, doi: 10.24912/computatio.v6i2.21098.

[12] R. Faturrahman, Y. S. Hariyani, and S. Hadiyoso, "Klasifikasi Jajanan Tradisional Indonesia berbasis Deep Learning dan Metode Transfer Learning," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 11, no. 4, p. 945, Oct. 2023, doi: 10.26760/elkomika.v11i4.945.

[13] Z. Abidin, R. I. Borman, F. B. Ananda, P. Prasetyawan, F. Rossi, and Y. Jusman, "Classification of Indonesian Traditional Snacks Based on Image Using Convolutional Neural Network (CNN) Algorithm," in *2021 1st International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, IEEE, Oct. 2021, pp. 18–23. doi: 10.1109/ICE3IS54102.2021.9649707.

[14] G. G. Patil and R. K. Banyal, "Techniques of Deep Learning for Image Recognition," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, IEEE, Mar. 2019, pp. 1–5. doi: 10.1109/I2CT45611.2019.9033628.

[15] S. Li, L. Wang, J. Li, and Y. Yao, "Image Classification Algorithm Based on Improved AlexNet," *J Phys Conf Ser*, vol. 1813, no. 1, p. 012051, Feb. 2021, doi: 10.1088/1742-6596/1813/1/012051.

[16] D. Hindarto, N. Afarini, and E. T. Esthi H, "Comparison Efficacy of VGG16 and VGG19 Insect Classification Models," *JIKO (Jurnal Informatika dan Komputer)*, vol. 6, no. 3, pp. 189–195, Dec. 2023, doi: 10.33387/jiko.v6i3.7008.

[17] J. S. Iskandar and R. P. Kristianto, "Pengenalan dan Klasifikasi Ragam Kue Indonesia menggunakan Arsitektur ResNet50V2 pada Convolutional Neural Network (CNN)," in *Seminar Nasional AMIKOM Surakarta (SEMNASA) 2023*, Surakarta, Nov. 2023.

[18] D. Hindarto, "Use ResNet50V2 Deep Learning Model to Classify Five Animal Species," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 7, no. 4, pp. 758–768, Dec. 2023, doi: 10.35870/jtik.v7i4.1845.

[19] K. Dong, C. Zhou, Y. Ruan, and Y. Li, "MobileNetV2 Model for Image Classification," in *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, IEEE, Dec. 2020, pp. 476–480. doi: 10.1109/ITCA52113.2020.00106.

[20] E. E. Citra, D. H. Fudholi, and C. K. Dewa, "Implementasi Arsitektur EfficientNetV2 Untuk Klasifikasi Gambar Makanan Tradisional Indonesia," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 2, p. 766, Apr. 2023, doi: 10.30865/mib.v7i2.5881.

[21] A. C. S, "Advancements in CNN Architectures for Computer Vision: A Comprehensive Review," in *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*, IEEE, Nov. 2023, pp. 1–7. doi: 10.1109/AICERA/ICIS59538.2023.10420413.

[22] B. Li, W. Jiang, J. Gu, K. Liu, and Y. Wu, "Research on Convolutional Neural Network in the Field of Object Detection," in *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, IEEE, Jul. 2020, pp. 820–827. doi: 10.1109/ICPICS50287.2020.9202194.

[23] W. Almubarok, U. D. Rosiani, and R. Andrie Asmara, "MobileNetV2 Pruning for Improved Efficiency in Catfish Classification on Resource-Limited Devices," in *2024 IEEE 10th Information Technology International Seminar (ITIS)*, IEEE, Nov. 2024, pp. 271–277. doi: 10.1109/ITIS64716.2024.10845279.

[24] J. Chi, C. Kim On, H. Zhang, and S. S. Chai, "A Review of Deep Convolutional Neural Networks in Mobile Face Recognition," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 23, pp. 4–19, Dec. 2023, doi: 10.3991/ijim.v17i23.40867.