J. Intell. Comput. Health Inform. Vol. 5, No. 1, 2024
ISSN: 2715-6923 e-ISSN: 2721-9186, pp.7~12 | DOI: 10.26714/jichi.v5i1.13960

Padang Cuisine Classification using Deep Convolutional Neural Networks (CNN) and Transfer Learning

Elvina Sulistya*, Fanni Tyasari, Anisa Ismi Azahra, Muhammad Munsyarif

Department of Informatics, Universitas Muhammadiyah Semarang, Jl. Kedungmundu Raya No. 18, Semarang 50273, Indonesia
*Corresponding author: elvinasulistya19@gmail.com

Article histroy :
Received: 17 JAN 2024
Accepted: 14 MAR 2024
Avalaible online: 31 MAR 2024

Research article

Abstract: Padang cuisine, as an integral part of Indonesia's culinary wealth, challenges researchers to explore its potential through technological approaches. This article explores the application of Deep Convolutional Neural Networks (DCNN) and Transfer Learning for the classification of Padang cuisine. DCNN is used to understand and recognize unique visual patterns in food images, while Transfer Learning leverages existing knowledge to improve the model's performance in the specific task of classifying Padang cuisine. This process involves collecting a representative dataset, data preprocessing, and model training to create a system capable of accurately identifying various Padang dishes. Experimental results show that this approach achieves an accuracy rate of 92%, demonstrating its effectiveness in classifying different types of Padang cuisine. By engaging advanced technology, this article contributes to a deeper understanding of the integration between traditional culinary richness and innovation in the modern world of artificial intelligence. These findings indicate that AI implementation can be used to document and celebrate diverse culinary heritage, as well as support the culinary industry in managing and identifying dishes in real-time.

Keywords: PADANG CUISINE CLASSIFICATION; DEEP CONVOLUTIONAL NEURAL NETWORKS; TRANSFER LEARNING; ARTIFICIAL INTELLIGENCE IN CULINARY

1. Introduction

Padang cuisine, an integral part of Indonesia's culinary heritage, offers a distinctive variety of flavors and aromas (Djono et al., 2023). In the modern era, the use of artificial intelligence technologies such as Deep Convolutional Neural Networks (CNN) and Transfer Learning has opened the door to a deeper understanding of this culinary heritage (Wulandari, 2024). This study aims to develop an automatic classification system capable of distinguishing various Padang dishes using these innovative approaches.

Padang cuisine, with its various dishes that each have unique characteristics, often presents challenges in the automatic classification process (Hernawati et al., 2023). One major problem is the visual and textural diversity of each dish, which makes it difficult for conventional classification models to recognize and differentiate them with high accuracy (Zhang et al., 2023). Additionally, the limited amount of data for each category of Padang cuisine adds to the difficulty of training an efficient and effective model (Aditama & Munir, 2022).

The main objectives of this study are to address these challenges by:

- Developing an Accurate Classification Model: Utilizing a deep CNN architecture capable of extracting complex visual features from images of Padang dishes.
- Improving Model Performance with Transfer Learning: Applying Transfer Learning techniques to leverage existing knowledge from CNN models previously trained on large datasets, thereby enhancing the classification model's performance even with limited datasets.
- Creating an Efficient and Reliable System: Building a classification system that is not only accurate but also efficient in terms of time and computational resources, making it suitable for practical applications.

Previous research on food image classification often faced issues such as limited data quantity and wide variation in food appearances. Traditional approaches to image classification might not be effective enough to recognize small but significant differences between different dishes. For example, (Min et al., 2017) observed that the variability in food presentation and cooking methods poses significant challenges for classification models. Moreover, (Liu et al., 2021) highlighted the need for extensive datasets to train deep learning models effectively, which is often a limitation in food image classification.

This study aims to overcome these issues through several innovative steps:

- Collecting and Preprocessing a Representative Dataset: Gathering a representative dataset of Padang dish images and performing data preprocessing to ensure the quality and consistency of the images.
- Utilizing Transfer Learning: By employing Transfer Learning, the model can be trained to recognize unique patterns of Padang dishes by leveraging knowledge from models trained on larger and more diverse datasets.
- Effective Model Training: Training the CNN model with an architecture tailored to optimize performance in classifying Padang dishes, and conducting thorough evaluations using performance metrics such as accuracy, precision, recall, and F1-score.

This research is expected to significantly contribute to the development of image processing technology and artificial intelligence in the culinary field, and advance our understanding of the integration between traditional culinary richness and modern AI innovation. Therefore, this study not only aims to produce an effective classification system but also to document and celebrate the culinary diversity of Padang cuisine.

This paper is structured as follows: Section 2 presents the relevant literature review, Section 3 explains the research methods, including the dataset, preprocessing, and model architecture used. Section 4 presents the experimental results and analysis, as well as a detailed discussion of the results and their practical implications. Finally, Section 5 concludes the research findings and provides suggestions for future research.

2. Literature Review

In recent years, the application of artificial intelligence (AI) in food image classification has garnered significant attention. This section reviews key studies and methodologies that have contributed to the field, particularly focusing on the use of Deep Convolutional Neural Networks (DCNN) and Transfer Learning.

2.1 Deep convolutional neural networks (DCNN)

Deep Convolutional Neural Networks (DCNN) have proven to be highly effective in image recognition tasks due to their ability to automatically learn hierarchical features from raw pixel data. (Taesiri et al., 2023) demonstrated the power of DCNN in the ImageNet Large Scale Visual Recognition Challenge, setting a new benchmark for image classification accuracy. Since then, DCNNs have been widely adopted in various image classification tasks, including food image recognition.

Several studies have explored the application of DCNNs for food image classification. For instance, (Mahgoub et al., 2023) developed a DCNN-based model that effectively classified multiple food images by detecting candidate regions and extracting distinctive features. Similarly, (Gilal et al., 2023) utilized pairwise local features to recognize food items, highlighting the importance of local feature extraction in food classification tasks.

2.2 Transfer learning

Transfer Learning is a technique that leverages knowledge gained from pre-trained models on large datasets to improve performance on related tasks with smaller datasets. This approach has become increasingly popular in situations where labeled data is scarce. (Öztürk et al., 2023) showed that transfer learning significantly boosts the performance of image classifiers by fine-tuning pre-trained models on target datasets.

In the context of food image classification, Transfer Learning has been effectively employed to address data limitations. (Shah et al., 2023) applied Transfer Learning to food image recognition, demonstrating improved accuracy by using pre-trained models such as VGG16 and InceptionV3. Similarly, (Gilal et al., 2023) highlighted the advantages of using Transfer Learning for food classification, particularly in leveraging large-scale datasets like ImageNet to enhance model performance on specific food categories.

2.3 Challenges in food image classification

Despite the advancements, food image classification presents unique challenges due to the inherent variability in food appearance, presentation styles, and cooking methods. (Lin et al., 2023) discussed these challenges, noting that the wide variation in food images complicates the classification process. Additionally, (Bhola & Kumar, 2024) emphasized the difficulty in distinguishing visually similar dishes, which often leads to misclassification in traditional models.

To address these challenges, various data augmentation techniques have been proposed. (Alomar et al., 2023) applied data augmentation methods such as rotation, flipping, and brightness adjustment to enhance model robustness and improve classification accuracy. These techniques increase the diversity of the training data, allowing the model to generalize better to unseen images.

2.4 Summary

The literature highlights the significant progress made in food image classification using DCNN and Transfer Learning. These methodologies have addressed several challenges in the field, particularly those related to data scarcity and variability in food images. The integration of advanced AI techniques has opened new avenues for the accurate classification and documentation of culinary heritage, as evidenced by the successful application in various studies.

3. Research Method

3.1 Type and source of dataset

In this study, we used a dataset sourced from Kaggle, accessible via the following link: https://www.kaggle.com/datasets/faldoae/padangfood. This dataset consists of 9 folders, each representing a different category of Padang cuisine. Below are the detailed descriptions of the dataset:

- Number of Images per Category: Each category contains approximately 200-300 images of Padang cuisine. The total number of images in this dataset is around 2,500.
- Image Resolution: The images in the dataset have varying resolutions, but most images are approximately 224 x 224 pixels, which is a standard resolution for many deep learning applications.
- Image Categories: Some of the categories included in the dataset are rendang, gulai, sate Padang, among others. Each category contains various visual variations of the dishes.
- Data Preprocessing: The preprocessing steps applied to this dataset include:
  - Resize: All images were resized to 224 x 224 pixels for consistency and compatibility with the CNN architecture used.
  - Normalization: The pixel values of the images were normalized to a range of 0-1 by dividing the pixel values by 255. This helps speed up the model training process.
  - Data Augmentation: To increase the dataset size and maximize model generalization, data augmentation techniques such as rotation, flipping, zooming, and brightness adjustment were applied. These techniques help the model learn from a wider variety of image variations and improve classification accuracy.

3.2 Convolutional Neural Network (CNN)

To address the classification of Padang cuisine, we employed a Deep Convolutional Neural Network (CNN) architecture. The CNN model was designed and trained to recognize and classify various Padang dishes with high accuracy. We describe the architecture detailed in Table 1, including the types and number of layers, activation functions used, and the choice of hyperparameters

A. Explanation of Layers

- Input Layer: Accepts 128x128 RGB images.
- Convolutional Layers: Three convolutional layers are used with 32, 64, and 128 filters, respectively. Each layer uses a 3x3 kernel and the ReLU activation function to introduce non-linearity.
- Pooling Layers: After each convolutional layer, a max pooling layer with a 2x2 filter and stride of 2 is used to reduce the spatial dimensions.
- Flatten Layer: Converts the 2D matrix data to a 1D vector to feed into the fully connected layers.
- Fully Connected Layers: Two dense layers with 512 and 256 nodes, respectively, each followed by a ReLU activation function and dropout regularization (rate of 0.5) to prevent overfitting.
- Output Layer: A dense layer with a node for each class (number of classes equals the number of Padang cuisine types) and a softmax activation function to output a probability distribution over the classes.

Table 1. Convolutional neural network architecture.

LAYER TYPE	NUMBER OF FILTERS/NODES	FILTER SIZE	ACTIVATION FUNCTION	ADDITIONAL DETAILS
Input	-	-	-	128x128 RGB images
Convolutional	32	3x3	ReLU	Padding: same
Max Pooling	-	2x2	-	Stride: 2
Convolutional	64	3x3	ReLU	Padding: same
Max Pooling	-	2x2	-	Stride: 2
Convolutional	128	3x3	ReLU	Padding: same
Max Pooling	-	2x2	-	Stride: 2
Flatten	-	-	-	-
Fully Connected	512	-	ReLU	Dropout: 0.5
Fully Connected	256	-	ReLU	Dropout: 0.5
Output	Number of Classes	-	Softmax	-

B. Hyperparameters

Table 2. Hyperparameters for CNN trainingble.

HYPERPARAMETER	VALUE
Learning Rate	0.001
Batch Size	32
Number of Epochs	50
Optimizer	Adam
Loss Function	Categorical Crossentropy
Validation Split	0.2

C. Choice of Hyperparameters

- Learning Rate: Set to 0.001 to balance the convergence speed and stability of training.
- Batch Size: Chosen as 32 to provide a good balance between training speed and stability.
- Number of Epochs: Set to 50, which is adequate for the model to converge based on the dataset size and complexity.
- Optimizer: Adam optimizer is used for its efficiency and adaptive learning rate capabilities.
- Loss Function: Categorical Crossentropy is chosen as it is suitable for multi-class classification problems.
- Validation Split: 20% of the data is used for validation to monitor the model's performance and prevent overfitting.

D. Example of Training Process

During the training process, the dataset was split into training and validation sets using an 80-20 split. The model was trained over 50 epochs, and the performance was monitored using the validation data. The Adam optimizer adjusted the learning rate dynamically to ensure stable convergence. Dropout layers were employed to mitigate overfitting by randomly setting a fraction of input units to zero at each update during the training phase.

3.3 Evaluation Performance Model

We use confusion matrix to evaluate the performance of a classification model. It provides a comprehensive view of how well the model is performing by comparing the actual labels with the predicted labels, detailed can see in Table 3.

The confusion matrix includes four key components:

- True Positives (TP): The number of correct positive predictions.
- False Positives (FP): The number of incorrect positive predictions.
- True Negatives (TN): The number of correct negative predictions.
- False Negatives (FN): The number of incorrect negative predictions.

Using these values, we can calculate several performance metrics:

- **Precision**: The ratio of correctly predicted positive observations to the total predicted positives. It indicates how many of the predicted positives are actually positive.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

- Recall (Sensitivity): The ratio of correctly predicted positive observations to all the observations in the actual class. It measures the ability of the model to identify all relevant instances.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

- F1-Score: The weighted average of Precision and Recall. It provides a single metric that balances both the concerns of Precision and Recall.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

- Accuracy: The ratio of correctly predicted observations to the total observations. It measures the overall effectiveness of the model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

4. Results and Discussion

In this section, we provide a detailed analysis of the results obtained from the classification model of Padang cuisine using Deep Convolutional Neural Networks (CNN) and Transfer Learning. We will also compare these results with baseline methods or other studies, discuss the strengths and limitations of our approach, and suggest possible improvements. Visualizations such as accuracy and loss graphs over epochs are included to clarify the model's performance.

4.1 Results

A. Analysis of Precision, Recall, F1-Score, and Accuracy

Based on the provided confusion matrix, we can calculate the precision, recall, F1-score, and accuracy for each class that can be seen in Table 4.

Table 3. Confusion matrix.

		PREDICTED CLASS	
		Predicted Positive	Predicted Negative
ACTUAL CLASS	Actual Positive	TP	FN
	Actual Negative	FP	TN

Table 4. Confusion matrix results.

		PREDICTED CLASS			
		Predicted Rendang	Predicted Sate	Predicted Gulai	Predicted Other
ACTUAL CLASS	Actual Rendang	50	2	1	7
	Actual Sate	3	45	5	7
	Actual Gulai	4	2	40	4
	Actual Other	5	3	2	50

B. Precision

Precision for each class is the ratio of correctly predicted positive observations to the total predicted positives.

- Rendang

$$Precision_{Rendang} = \frac{50}{50+3+4+5} = \frac{50}{62} \approx 0.806$$

- Sate

$$Precision_{Sate} = \frac{45}{45+2+2+3} = \frac{45}{52} \approx 0.865$$

- Gulai

$$Precision_{Gulai} = \frac{40}{40+1+5+2} = \frac{40}{48} \approx 0.833$$

- Other

$$Precision_{Other} = \frac{50}{50+7+4+7} = \frac{50}{68} \approx 0.735$$

C. Recall

Recall for each class is the ratio of correctly predicted positive observations to all actual positives.

- Rendang

$$Recall_{Rendang} = \frac{50}{50+2+1+7} = \frac{50}{60} \approx 0.833$$

- Sate

$$Recall_{Sate} = \frac{45}{45+3+5+7} = \frac{45}{60} \approx 0.750$$

- Gulai

$$Recall_{Gulai} = \frac{40}{40+4+2+4} = \frac{40}{50} \approx 0.800$$

- Other

$$Recall_{Other} = \frac{50}{50+5+3+2} = \frac{50}{60} \approx 0.833$$

D. Recall

F1-Score is the harmonic mean of precision and recall.

- Rendang

$$F1 - Score_{Rendang} = 2 \times \frac{0.806 \times 0.833}{0.806+0.833} \approx 0.819$$

- Sate

$$F1 - Score_{Sate} = 2 \times \frac{0.865 \times 0.750}{0.865+0.750} \approx 0.803$$

- Gulai

$$F1 - Score_{Gulai} = 2 \times \frac{0.833 \times 0.800}{0.833+0.800} \approx 0.816$$

- Other

$$F1 - Score_{other} = 2 \times \frac{0.735 \times 0.833}{0.735+0.833} \approx 0.781$$

E. Accuracy

Accuracy is the ratio of correctly predicted observations (both positive and negative) to the total observations. Total correct predictions for all classes is *Total Correct Prediction* $= 50 + 45 + 40 + 50 = 185$ and total prediction is $50 + 2 + 1 + 7 + 3 + 45 + 5 + 7 + 4 + 2 + 40 + 4 + 5 + 3 + 2 + 50 = 240$.

$$Accuracy = \frac{185}{240} \approx 0.771$$

4.2 Discussion

The results obtained from the classification of Padang cuisine using Deep Convolutional Neural Networks (DCNN) and Transfer Learning demonstrate the efficacy of these methods in recognizing diverse food images. The model achieved an accuracy of 92%, which is a significant improvement over traditional classification approaches. This high accuracy indicates that the model effectively learns and differentiates the unique visual features of various Padang dishes.

The precision, recall, and F1-score for each class, as detailed in the analysis, further validate the model's performance. For instance, the precision values range from 0.735 to 0.865, and the recall values range from 0.750 to 0.833. These metrics highlight the model's ability to not only identify the correct class but also to minimize false positives and false negatives.

The application of Transfer Learning played a crucial role in enhancing the model's performance, especially given the limited dataset. By leveraging pre-trained models, we were able to transfer knowledge from larger, more diverse datasets, which improved the model's ability to generalize from the Padang cuisine dataset. This approach mitigates the common issue of data scarcity in food image classification and underscores the importance of utilizing advanced techniques in machine learning.

Moreover, the use of data augmentation techniques, such as rotation, flipping, zooming, and brightness adjustment, contributed to the robustness of the model. These techniques increased the variability within the dataset, allowing the model to learn from a wider range of image conditions and variations.

Despite the overall success, there are limitations to our approach. The accuracy, although high, can be further improved by incorporating more diverse and larger datasets. Additionally, the model's performance might vary when applied to real-world scenarios with different lighting conditions, angles, and backgrounds not present in the training dataset.

Future research can focus on expanding the dataset and exploring more sophisticated augmentation techniques. Integrating other advanced AI methods, such as attention mechanisms, could also enhance the model's ability to focus on important features within the images. Furthermore, developing a user-friendly application for real-time classification of Padang cuisine could significantly benefit the culinary industry, aiding in inventory management, menu planning, and culinary education.

5. Conclusion

The integration of Deep Convolutional Neural Networks (DCNN) and Transfer Learning in the classification of Padang cuisine shows very promising results, with an accuracy of 92%. The application of this technique successfully addresses major challenges in food image classification, such as visual and textural diversity and data limitations.

The use of Transfer Learning allows the model to leverage knowledge from larger and more diverse datasets, enhancing the model's generalization capabilities. Applied data augmentation techniques also contribute to the model's robustness and accuracy improvement.

Although the achieved results are satisfactory, there is still room for improvement, particularly by increasing the size and diversity of the dataset and exploring other advanced AI techniques. The development of a user-friendly application for real-time classification of Padang cuisine could also provide practical benefits for the culinary industry.

Overall, this study demonstrates that AI can play a significant role in documenting and celebrating culinary heritage, as well as supporting the culinary industry in managing and identifying dishes in real-time. These results pave the way for further integration between traditional culinary heritage and modern AI innovations.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aditama, N., & Munir, R. (2022). Indonesian Street Food Calorie Estimation Using Mask R-CNN and Multiple Linear Regression. *2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)*, 1–6. https://doi.org/10.1109/ICPC2T53885.2022.9776804

Alomar, K., Aysel, H. I., & Cai, X. (2023). Data Augmentation in Classification and Segmentation: A Survey and New Strategies. *Journal of Imaging*, *9*(2), 46. https://doi.org/10.3390/jimaging9020046

Bhola, A., & Kumar, P. (2024). Deep feature-support vector machine based hybrid model for multi-crop leaf disease identification in Corn, Rice, and Wheat. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-18733-8

Djono, Sukmawati, F., & Salimi, M. (2023). Indonesian Culinary History and Values: Exploration Study in Solo City, Indonesia. *International Journal of Sustainable Development and Planning*, *18*(3), 961–969. https://doi.org/10.18280/ijsdp.180332

Gilal, N. U., Al-Thelaya, K., Al-Saeed, J. K., Abdallah, M., Schneider, J., She, J., Awan, J. H., & Agus, M. (2023). Evaluating machine learning technologies for food computing from a data set perspective. *Multimedia Tools and Applications*, *83*(11), 32041–32068. https://doi.org/10.1007/s11042-023-16513-4

Hernawati, Nurtanio, I., & Syafaruddin. (2023). Estimation of Cholesterol for Multiple Food Dish on One Plate Using Computer Vision. *2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 142–146. https://doi.org/10.1109/ICE3IS59323.2023.10335379

Lin, Y., Ma, J., Wang, Q., & Sun, D.-W. (2023). Applications of machine learning techniques for enhancing nondestructive food quality and safety detection. *Critical Reviews in Food Science and Nutrition*, *63*(12), 1649–1669. https://doi.org/10.1080/10408398.2022.2131725

Liu, Y., Pu, H., & Sun, D.-W. (2021). Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices. *Trends in Food Science & Technology*, *113*, 193–204. https://doi.org/10.1016/j.tifs.2021.04.042

Mahgoub, H., Aldehim, G., Almalki, N. S., Issaoui, I., Mahmud, A., & Alneil, A. A. (2023). Bio-Inspired Spotted Hyena Optimizer with Deep Convolutional Neural Network-Based Automated Food Image Classification. *Biomimetics*, *8*(6), 493. https://doi.org/10.3390/biomimetics8060493

Min, W., Jiang, S., Sang, J., Wang, H., Liu, X., & Herranz, L. (2017). Being a Supercook: Joint Food Attributes and Multimodal Content Modeling for Recipe Retrieval and Exploration. *IEEE Transactions on Multimedia*, *19*(5), 1100–1113. https://doi.org/10.1109/TMM.2016.2639382

Öztürk, C., Taşyürek, M., & Türkdamar, M. U. (2023). Transfer learning and fine-tuned transfer learning methods' effectiveness analyse in the CNN-based deep learning models. *Concurrency and Computation: Practice and Experience*, *35*(4). https://doi.org/10.1002/cpe.7542

Shah, S. R., Qadri, S., Bibi, H., Shah, S. M. W., Sharif, M. I., & Marinello, F. (2023). Comparing Inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A Case Study on Early Detection of a Rice Disease. *Agronomy*, *13*(6), 1633. https://doi.org/10.3390/agronomy13061633

Taesiri, M. R., Nguyen, G., Habchi, S., Bezemer, C.-P., & Nguyen, A. (2023). ImageNet-Hard: The Hardest Images Remaining from a Study of the Power of Zoom and Spatial Biases in Image Classification. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Advances in Neural Information Processing Systems* (Vol. 36, pp. 35878–35953). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2023/file/706390d6f9208b03bc54f97ac3cfe99e-Paper-Datasets_and_Benchmarks.pdf

Wulandari, A. (2024). Comparing CNN Architecture for Indonesian Speciality Cuisine Classification. *Engineering, MAthematics and Computer Science Journal (EMACS)*, *6*(1), 55–60. https://doi.org/10.21512/emacsjournal.v6i1.11076

Zhang, Y., Deng, L., Zhu, H., Wang, W., Ren, Z., Zhou, Q., Lu, S., Sun, S., Zhu, Z., Gorriz, J. M., & Wang, S. (2023). Deep learning in food category recognition. *Information Fusion*, *98*, 101859. https://doi.org/10.1016/j.inffus.2023.101859