JURNAL EMACS
[Engineering, MAthematics and Computer Science] Vol.6 No.1 January 2024: 55-60
e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v6i1.11076

Comparing CNN Architecture for Indonesian Speciality Cuisine Classification

Ajeng Wulandari
Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
ajeng.wulandari@binus.ac.id

*Correspondence: ajeng.wulandari@binus.ac.id

Abstract – *Indonesia's diverse and flavorful cuisine is a hidden gem, reflecting the nation's rich history and cultural tapestry. However, many of these culinary treasures remain undiscovered by a wider audience despite the popularity of beef rendang. This study represents a fascinating blend of technology and gastronomy, using smart computers to unravel the secrets of Indonesian flavors. This research employs one of the most popular neural networks methods called Convolutional Neural Network (CNN) to shine a light on many citizens' favorite regional specialty cuisine which is Padang cuisine from West Sumatra, Indonesia. Gathering a collection of 993 images from 9 various dishes, the machine is trained to automatically recognize these unique culinary delights. Among several different Convolutional Neural Network models trained and tested, DenseNet-201 emerged as the top performer, showcasing remarkable accuracy, precision, recall and f1-score higher than 0.90. By harnessing the power of advanced neural networks, we not only gain insights into the intricacies of the region's culinary traditions but also pave the way for a deeper appreciation and understanding of the cultural significance embedded in every bite. Beyond this research technological achievements, it also emphasizes the importance of preserving and promoting Indonesia's diverse culinary heritage and rich tapestry of global food heritage.*

Keywords: *Padang Cuisines; Food Recognition; Pre-Trained Model; Convolutional Neural Networks*

I. INTRODUCTION

People's relationship with food is very complex. Not only food is needed for their survival instinct, but it is also tied to their nutrition & dietary, emotional comfort, sensory pleasure, even for social bounding. Food in most countries is also related to their cultures, rituals and traditions. The diversity of food allows people to explore different cuisine and flavors from around the world. Indonesia is a diverse nation, consisting of various tribes, races, languages, religions, and ethnicities due to differences in geography, beliefs, history, etc. It's no surprise that there are many variants of Indonesian food specialties.

According to TasteAtlas2023, Indonesia is in the top10 list of countries with best cuisine around the world in 2023 (Afifa, 2023). Several Indonesian foods have been included as the best food in the world, namely beef rendang, nasi goreng, and satay (Cheung, 2017). These signature recipes have been passed down from generation to generation. In addition, Indonesian cuisine has started to attract many tourists which will also contribute to Indonesia tourism (Yuniar et al., 2022). However, the popularity of Indonesian cuisine is less than its ASEAN's neighbors such as Vietnam, Philippines and Thailand (Pepinsky, 2015) (The Economist, 2022). Whereas it can also help promote Indonesian heritage to the world.

Many technologies have been implemented in food computing these past few years, such as perception, recognition, retrieval, recommendation, prediction and monitoring, etc (Min et al., 2019). Recognizing food from images is extremely useful for several purposes like diet food tracking and promoting the food industry. Even though the name of the food is the same, the appearance of the food can be different because it depends on who is serving it. This is the biggest challenge in food introduction that makes the data have large variations in food shape, volume, texture, color, and compositions. Image processing and machine learning approaches have been implemented to solve this food recognition problem like GLCM, Random Forest, CNN. Several food datasets that have been used as benchmarks are Food-101 (Bossard et al., 2014), UPMC Food-101 (Gallo et al., 2020), VireoFood-172 (Chen & Ngo, 2016), Food524DB (Ciocca et al., 2017), Food-101N (Lee et al., 2018), UNIMIB2015 (Ciocca et al., 2015), UECFood-256, and UECFood-100 (Matsuda et al., 2012).

Researchers are also doing some food images classification for their countries's traditional cuisine. They use dataset from the public, web crawling, or building their own dataset. (Sahoo et al., 2019) has developed FoodAI to classify commonly Singaporean daily food for diet journals with combination of SENet and ResNeXt giving a top-1 accuracy of 80.86%. (Konstantakopoulos et al., 2021) collected Mediterranean and Greek Cuisine images from the web and the classification process achieved 97.8% top-5 accuracy with EfficientNet-B2. (Termritthikun & Kanprachar, 2018) utilized and improved the Residual network name NU-ResNet to classify Thailand food collected from search engines and achieved 97.04% top-5 accuracy. (Tasnim et al., 2020) used InceptionV3 for recognizing Bangladeshi Cuisine with an average accuracy of 95.2% approximately. (Fahira et al., 2019) classify Sumatera traditional food that they collected and got the highest 100% accuracy and f1-score after applying histogram color, gabor filter texture and Random Forest classifier. (Fahira et al., 2020) also achieved 0.99 accuracy for Javanese traditional food using Haralick method, Hu Moment Invariance, and Histogram information to do feature extraction and Random Forest to do the classification.

Although there is some research related to Indonesian food classification, it is still limited. This paper endeavors to delve into the heart of Indonesian culinary, with a particular focus on the image classification of its diverse cuisine, using Padang speciality food as a representative case study. Padang is chosen due to its food popularity amongst many people (Ibrahim, 2022). This study explores and compares various CNN architectures aiming to uncover the most effective approach for the classification of Indonesian cuisine. By leveraging the capabilities of advanced neural networks, this research not only discerns the subtleties within the culinary traditions of the region but also hopes to establish a foundation for a more profound appreciation and understanding of the cultural significance inherent in each gastronomic experience.

II. METHODS

The representation of this research proposed model can be seen on the following Figure 1.

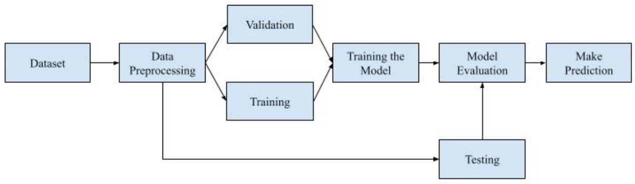


Figure 1. Proposed Model Process

2.1 Data

This research is using Padang Cuisine (Indonesian Food Image Dataset) from Kaggle (Afrinanto, 2022) as a dataset. Padang cuisine is food from the Minangkabau area, West Sumatra, Indonesian. In this image dataset, there will consists of 9 classes in Table I, namely Beef Rendang, Chicken Pop, Fried Chicken, Dendeng Batokok, Fish Curry, Tambusu Curry, Tunjang Curry, Balado Egg, and Padang Omelette with a total of 993 images which you can see it in Figure 2.

Table I. Class Image Distribution

Class	Total Images
Tambusu Curry (*Gulai Tambusu*)	103
Dendeng Batokok	109
Chicken Pop (*Ayam Pop*)	113
Beef Rendang (*Daging Rendang*)	104
Balado Egg (*Telur Balado*)	111
Padang Omelette (*Telur Dadar*)	116
Fried Chicken (*Ayam Goreng*)	107
Fish Curry (*Gulai Ikan*)	111
Tunjang Curry (*Gulai Tunjang*)	119

Figure 2. Example of dataset in several class

2.2 Image Transformation

Image processing has an important role in the process and analysis step before building the model. This research image dataset will be resize into 224*224 pixels and rescale after that. Later, the dataset will be split into three different categories: training, validation and testing.

2.3 Convolutional Neural Networks

Due to its ability to recognize patterns, Convolutional Neural Networks (CNN) has been widely used for several cases especially for image recognition and processing. Yann Lecun is the pioneer of this groundbreaking work in the 1990s for the success of LeNet-5 architecture (LeCun et al., 1998) in classifying handwritten digits. It has 3 main layers, which are convolutional layer, pooling layer and fully-connected layer.

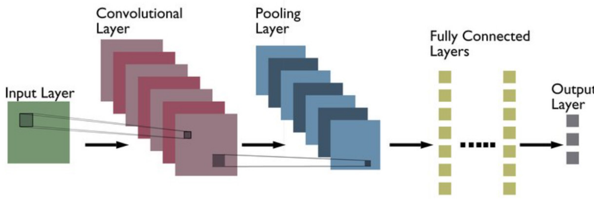


Figure 3. Basic CNN Architecture (Kumar, 2023)

In convolutional layers, it performs convolution operations which are a sort of matrix multiplication to the input data and outputting a feature map. In pooling layers, it will reduce the spatial dimension which helps in reducing the number of parameters or weights. In fully connected layers, it is used in the end of CNN with goals to make prediction based on the learned features from convolutional and pooling layers

2.3.1 VGG

VGGNet (Simonyan & Zisserman, 2015) is 16 layer CNN with up to 95 millions parameters and trained on over one billion images with 1000 classes. This model is computationally efficient so that it is usually used as the baseline in many computer vision applications.

2.3.2 Inception

InceptionV3 Model (Szegedy et al., 2015) is based on the GoogLeNet (InceptionV1) CNN architecture that successfully reduced error rate in ILSVRC2014. It has 42 layers with a lower error rate than its previous 2 versions with several techniques including factorized convolutions, regularization, dimension reduction, and parallelized computations.

2.3.3 MobileNet

MobileNet (Howard et al., 2017) initially developed to fit in mobile devices to classify images or detect objects. This model is a very small CNN that makes it easy to run in real-time.

2.3.4 DenseNet

DenseNet (Huang et al., 2018) is designed by concatenating the feature maps of all previous layers, a dense block allows each layer to access the features of all preceding levels. This model works on the idea that convolutional networks can be substantially deeper, more accurate, and efficient to train if they have shorter connections between layers close to the input and those close to the output.

2.4 Evaluation

Performance of the model will depend on the correctness of mapping the data into their respective categories. Accuracy, Precision, Recall and F1-Score will be used as the evaluation metrics. Confusion matrix approach is also used to assess performance of the model. TP (True Positive) is a condition where the model *correctly* predicts the *positive* class. TN (True Negative) is a condition where the model *correctly* predicts the *negative* class. FP (False Positive) is a condition where the model *incorrectly* predicts the *positive* class. FN (False Negative) is a condition where the model *incorrectly* predicts the *negative* class.

Accuracy can be defined as the percentage of correct predictions made by the classification model. However, accuracy is not a reliable metric for datasets having class imbalance.

$$Accuracy = \frac{TP+FN}{(TP+FP+FN+TN)} * 100\% \quad (1)$$

Precision indicates out of all positive predictions, how many are actually positive.

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

Recall indicates out of all actually positive values, how many are predicted positive.

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

When avoiding both false positives and false negatives are equally important for solving the problem, it needs a trade-off between precision and recall. This is when we use the f1-score as a metric to define the harmonic mean of precision and recall.

$$F1score = \frac{(2*(Precision*Recall))}{Precision+Recall} \quad (4)$$

A confusion matrix is a technique for summarizing the performance of a classification algorithm.

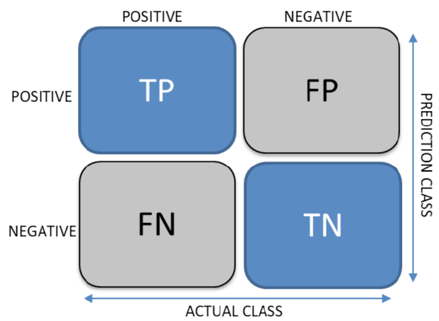


Figure 4. Confusion Matrix

III. RESULTS AND DISCUSSION

All experiments in this research have been done using Kaggle Notebooks and their free GPU NVIDIA Tesla P100. The process has been done by using Python and Tensorflow. After all images in the dataset are transformed, the dataset is split into 80% training process and 20% testing process so that there will be 636 images for training, 159 images for validation and 199 images for testing.

The researcher made comparisons with several pre-trained CNN models namely VGG-16, InceptionV3, MobileNetV3 and DenseNet-201, to find the suitable model to classify this Indonesian Specialty Cuisine. Hyperparameter used in this experiment are batch size (32), input shape (224x224), output layer - softmax (9), optimizer (adam), learning rate (0.00001) and epochs (100) with early stopping if the loss doesn't decrease after 3 epochs.

All CNN models in this experiment have their own training, validation and testing accuracy and the results are represented in Table II. In the training steps, MobileNetV3 has the highest accuracy among the others with achievement 94.03% in training and 87.42% in validation. However, MobileNetV3 accuracy in testing has decreased to 87.44% which is lower than its training accuracy. Meanwhile, DenseNet-201 has the highest testing accuracy at 90.45%.

Table II. CNN Model Accuracy

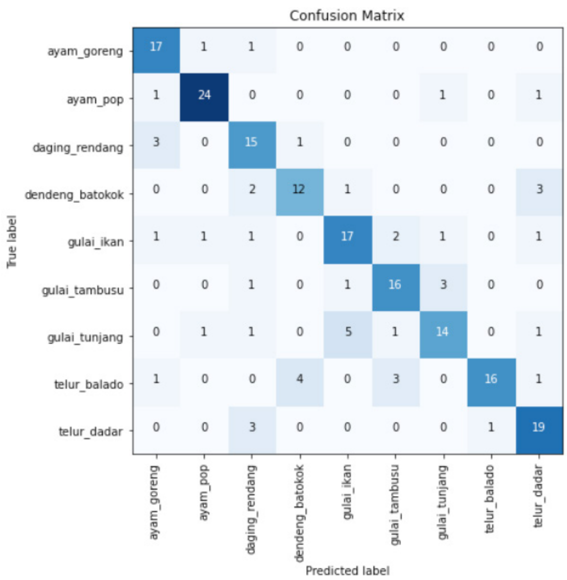
	Training	Validation	Testing
VGG-16	68.87%	71.70%	75.38%
InceptionV3	88.21%	76.10%	85.93%
MobileNetV3	**94.03%**	**87.42%**	87.44%
DenseNet-201	90.88%	86.16%	**90.45%**

Besides accuracy, this research also evaluates the model with precision, recall and f1-score. From those 4 CNN architectures, DenseNet-201 received the highest score among the others with score 0.91 in precision, 0.90 in recall and 0.90 in f1-score which detail can be seen in Table III.

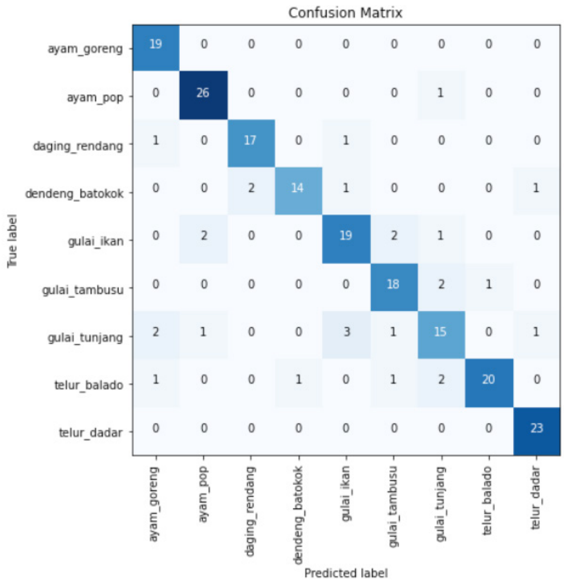
Table III. Model Evaluation

	Precision	Recall	F1-Score
VGG-16	0.76	0.75	0.75
InceptionV3	0.86	0.86	0.86
MobileNetV3	0.88	0.87	0.87
DenseNet-201	**0.91**	**0.90**	**0.90**

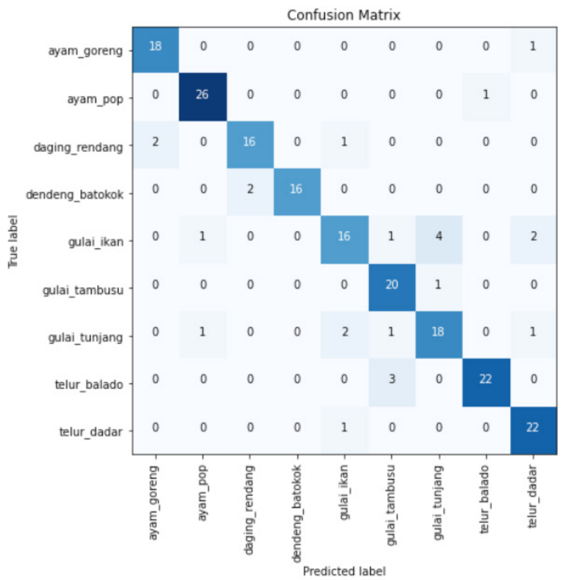
Based on the confusion matrix in Figure 5, it is also known that chicken pop (*ayam pop*) class is the most correctly predicted class with all CNN models architecture.



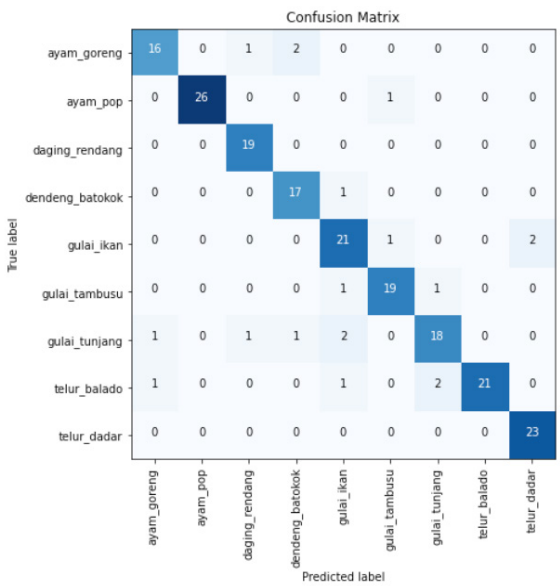
(a)



(b)



(c)



(d)

Figure 5. Confusion Matrix (a)VGG-16, (b) InceptionV3, (c) MobileNetV3, (d) DenseNet-201

This research chose 25 random images in test data and made a prediction using the DenseNet-201 model which has highest accuracy in the testing phase. If the predicted label is true then it will be labeled in green color, otherwise it will be red, shown in Figure 6.

Figure 6. Prediction on Test Data with DenseNet201 Model

IV. CONCLUSION

This research focused on automatic food recognition for Indonesian Specialty Cuisine to increase awareness about Indonesia and its cuisine. Although Beef rendang (Padang Cuisine) has been named the most delicious food in the world for several years, many people have little knowledge about Indonesian Cuisine. This experiment used a Padang cuisine dataset consisting 993 images from 9 classes for the training and testing phase. Several pre-trained CNN models namely VGG-16, InceptionV3, MobileNetV3 and DenseNet-201 compared to find out the best model for this experiment. The results shows DenseNet-201 achieved the highest score in accuracy, precision, recall and f1-score. It means that DenseNet-201 is the most suitable model in this automation food recognition. Beyond that, this research also hopes it can help preserve and promote Indonesia's diverse culinary heritage. In the future, we can gather more data about another Indonesian region's specialty cuisine and implement this CNN model to automatically recognize it.

REFERENCES

Afifa, L. (2023, December 13). Indonesia Ranks 6th Among Countries With Best Cuisine; TasteAtlas 2023. Tempo. https://en.tempo.co/read/1809003/indonesia-ranks-6th-among-countries-with-best-cuisine-tasteatlas-2023

Afrinanto, F. F. (2022). Padang Cuisine (Indonesian Food Image Dataset) (Version 1) [Image]. https://doi.org/10.34740/KAGGLE/DSV/4053613

Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101 – Mining Discriminative Components with Random Forests. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), Computer Vision – ECCV 2014 (Vol. 8694, pp. 446–461). Springer International Publishing. https://doi.org/10.1007/978-3-319-10599-4_29

Chen, J., & Ngo, C. (2016). Deep-based Ingredient Recognition for Cooking Recipe Retrieval. Proceedings of the 24th ACM International Conference on Multimedia, 32–41. https://doi.org/10.1145/2964284.2964315

Cheung, T. (2017, July 12). World's 50 best foods: Reader's choice | CNN. Your Pick: World's 50 Best Foods. https://edition.cnn.com/travel/article/world-best-foods-readers-choice/index.html

Ciocca, G., Napoletano, P., & Schettini, R. (2015). Food Recognition and Leftover Estimation for Daily Diet Monitoring. In V. Murino, E. Puppo, D. Sona, M. Cristani, & C. Sansone (Eds.), New Trends in Image Analysis and Processing—ICIAP 2015 Workshops (Vol. 9281, pp. 334–341). Springer International Publishing. https://doi.org/10.1007/978-3-319-23222-5_41

Ciocca, G., Napoletano, P., & Schettini, R. (2017). Learning CNN-based Features for Retrieval of Food Images. In S. Battiato, G. M. Farinella, M. Leo, & G. Gallo (Eds.), New Trends in Image Analysis and Processing – ICIAP 2017 (Vol. 10590, pp. 426–434). Springer International Publishing. https://doi.org/10.1007/978-3-319-70742-6_41

Fahira, P. K., Rahmadhani, Z. P., Mursanto, P., Wibisono, A., & Wisesa, H. A. (2020). Classical Machine Learning Classification for Javanese Traditional Food Image. 2020 4th International Conference on Informatics and Computational Sciences (ICICoS), 1–5. https://doi.org/10.1109/ICICoS51170.2020.9299039

Fahira, P. K., Wibisono, A., Wisesa, H. A., Rahmadhani, Z. P., Mursanto, P., & Nurhadiyatna, A. (2019). Sumatra Traditional Food Image Classification Using Classical Machine Learning. 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), 1–5. https://doi.org/10.1109/ICICoS48119.2019.8982447

Gallo, I., Ria, G., Landro, N., & Grassa, R. L. (2020). Image and Text fusion for UPMC Food-101 using BERT and CNNs. 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ), 1–6. https://doi.org/10.1109/IVCNZ51579.2020.9290622

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (arXiv:1704.04861). arXiv. http://arxiv.org/abs/1704.04861

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2018). Densely Connected Convolutional Networks (arXiv:1608.06993). arXiv. http://arxiv.org/abs/1608.06993

Ibrahim, R. (2022, June 24). Twenties: Essential Minangkabau cuisine. The Jakarta Post. https://www.thejakartapost.com/culture/2022/06/24/twenties-essential-minangkabau-cuisine.html

Konstantakopoulos, F. S., Georga, E. I., & Fotiadis, D. I. (2021). Mediterranean Food Image Recognition Using Deep Convolutional Networks. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 1740–1743. https://doi.org/10.1109/EMBC46164.2021.9630481

Kumar, A. (2023, December 4). Different Types of CNN Architectures Explained: Examples. Analytics Yogi. https://vitalflux.com/different-types-of-cnn-architectures-explained-examples/

LeCun, Y., Bottou, L., Bengio, Y., & Ha, P. (1998). Gradient-Based Learning Applied to Document Recognition.

Lee, K.-H., He, X., Zhang, L., & Yang, L. (2018). CleanNet: Transfer Learning for Scalable Image Classifier Training with Label Noise (arXiv:1711.07131). arXiv. http://arxiv.org/abs/1711.07131

Matsuda, Y., Hoashi, H., & Yanai, K. (2012). Recognition of Multiple-Food Images by Detecting Candidate Regions. 2012 IEEE International Conference on Multimedia and Expo, 25–30. https://doi.org/10.1109/ICME.2012.157

Min, W., Jiang, S., Liu, L., Rui, Y., & Jain, R. (2019). A Survey on Food Computing (arXiv:1808.07202). arXiv. http://arxiv.org/abs/1808.07202

Pepinsky, T. (2015, October 5). Why Hasn't Indonesian Food Caught On in the US? https://tompepinsky.com/2015/10/05/why-hasnt-indonesian-food-caught-on-in-the-us/

Sahoo, D., Hao, W., Ke, S., Xiongwei, W., Le, H., Achananuparp, P., Lim, E.-P., & Hoi, S. C. H. (2019). FoodAI: Food Image Recognition via Deep Learning for Smart Food Logging. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2260–2268. https://doi.org/10.1145/3292500.3330734

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition (arXiv:1409.1556). arXiv. http://arxiv.org/abs/1409.1556

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision (arXiv:1512.00567). arXiv. http://arxiv.org/abs/1512.00567

Tasnim, N., Romyull Islam, Md., & Shuvo, S. B. (2020). A Convolution Neural Network Based Classification Approach for Recognizing Traditional Foods of Bangladesh from Food Images. In A. Abraham, A. K. Cherukuri, P. Melin, & N. Gandhi (Eds.), Intelligent Systems Design and Applications (Vol. 940, pp. 844–852). Springer International Publishing. https://doi.org/10.1007/978-3-030-16657-1_79

Termritthikun, C., & Kanprachar, S. (2018). NU-ResNet: Deep Residual Networks for Thai Food Image Recognition. 10(1).

The Economist. (2022, September 1). Indonesian food is delicious but little-known. The Economist. https://www.economist.com/culture/2022/09/01/indonesian-food-is-delicious-but-little-known

Yuniar, N., Yumna, M., & Suharto. (2022, October 28). Cuisine to be gateway to tourism: Ministry. Antara News. https://en.antaranews.com/news/257465/cuisine-to-be-gateway-to-tourism-ministry