Grenze International Journal of Engineering and Technology, June Issue

Meat Dish Image Recognizer: Automatic Categorization of Various Meat Dish Images in Malaysia using Deep Learning Techniques

Putra Sumari[1], Valliappan Raman[2], Prabhavathy[3], Chuah Yung Fei[4], Daniel Wong Sai Meng[5], Kang Hean Heng[6], Koay Wei Sheng[7] and Seow Wen Han[8]
[1]School of Computer Science University Sains Malaysia Penang, Malaysia,
[2-3]Department of AI and DS Coimbatore Institute of Technology Coimbatore, India,
[4-8]School of Computer Science University Sains Malaysia Penang, Malaysia
Email: valliappan@cit.edu.in, prabhavathy.phd@gmail.com, (yunfeichuah; danielwong2698; kang_heanheng; wskoay; seowwenhan) @student.usm.my

Abstract*: Malaysia is a gourmet treat, offering a diverse range of meals and delicacies. Here, we are describing the major types of Malaysia meat dish cuisine food. Beef is the most prevalent meat in Malaysian cuisine and it can be prepared in various methods. Depending on the dish, the beef dish can be cooked with different methods, and ingredients and comes with a different flavour. Automatic image recognition is a convenient option for labelling and categorizing dishes. In this paper, the five types of beef dishes classification approaches are presented with a dataset that contains approximately 1843 images captured with different positions, viewpoints, lighting, colours, scale, etc. The proposed model classifies the type of beef dishes from the input images containing daging kunyit, masak merah, rendang, sate and sup gearbox. For this project, Convolutional Neural Network (CNN) algorithms, a deep learning approach widely utilised in image recognition, were used. Additionally, transfer learning approaches in the categorization of meat dishes were examined. By using the parameters of a pretrained dish classification model as the new beginning parameters of the proposed neural network, transfer learning was achieved. The transfer learning model outperforms CNN in terms of accuracy compared with the same model when trained with random initial weights. The final classification accuracy of the transfer learning model by ResNet50 was 89.40% which is higher compared to the proposed CNN deep learning architecture at 54.35% accuracy. The accuracy level of 89.40%, demonstrates the feasibility of this approach in the classification of meat dishes.

***Keywords*: Lauk daging, kunyit, masak merah, rendang, sate, sup gearbox, convolutional neural network, deep learning, transfer learning**

I. INTRODUCTION

As a result of the dynamic advancement of our human civilization, an increased amount of focus has been placed on the excellence of our way of life, in particular the food that we consume. Rice, noodles, and bread are Malaysia's staple food for carbohydrate intake. For protein, beef is eaten most often in Malaysia, it is most frequently seen in meals such as stews, roasts, and curries and is even sometimes served with noodles. If there is one thing on which all Malaysians can agree, it is their undying love for all forms of delicious cuisine. Despite their diverse ethnic,

Grenze ID: 01.GIJET.9.2.401_1

cultural, and religious origins, most Malaysians enjoy eating. West Malaysia has always played a major role in the maritime spice trade. This is partly attributable to the nation's strategic location in Southeast Asia.

Over the last several years, computer vision has become more prevalent in food identification. Deep Neural Networks (DNN), which are employed in image identification and classification, are used to recognise food from their pictured representations. DNN outperforms other machine learning algorithms in terms of performance. One technique that falls under the category of deep learning is known as a convolutional neural network (CNN) [1]. CNN's are the artificial neural networks (ANN) used most often in deep learning. It is now being used in various visual recognition analyses, including video and image identification, face recognition, handwritten digit recognition, food recognition, and other similar applications [1]. Accuracy levels in several areas, including fruit or dishes recognition with CNN, have achieved a degree of perfection comparable to that of humans [1].

Malaysia has several well-known culinary specialties, ranging from appetisers and snacks to meat and poultry, noodle and rice meals, desserts, and beverages. Dish identification in multimedia is considered to be a challenging subject because of the variety of ways in which food might look, both in terms of its form and its colour, as a result of the many techniques of cooking and chopping. The convolutional neural network (CNN) is currently the most advanced approach for doing image classification for the dishes since it enables the system to discover the relevant characteristics on its own and is thus considered the state-of-the-art method. The capacity of CNNs to supervise learning invariant representations of pictures, with the assistance of spatial filters, is largely responsible for their success. Hence, this project mainly focuses on classifying 5 types of beef dishes which are “Daging Kunyit”, “Daging Masak Merah”, “Daging Rendang”, “Sate” and “Sup Gearbox”.

II. Literature Survey

In recent years, deep learning algorithm, such as CNN has been widely used in many real-world applications. CNN is the main Deep Learning architecture for image classification and object detection due to its ability to learn robust representations from images. There are many different types of CNN architectures and each of the architectures has its characteristics. For example, LeNet is a simple yet powerful model that has been used for various tasks such as handwritten digit recognition, traffic sign recognition and face detection. [7] Apart from LeNet, AlexNet is also a CNN architecture, which is very similar to LeNet in terms of architecture. The AlexNet architecture was designed to be used with large-scale image datasets and it achieved state-of-the-art results during its publication. [7]

In 2020, Palakodati et al. [8] present an approach, which uses CNN and Transfer Learning for fresh and rotten fruits classification. Three types of fruits, such as apples, bananas, and oranges are selected as the input images. CNN algorithm is used for extracting the features from input fruits images, and the Softmax activation function is used to classify the images into fresh and rotten fruits. The proposed CNN model achieved an accuracy rate of 97.82%, which is the highest, compared to VGG16, VGG19, MobileNet, and Xception, in classifying fresh and rotten fruits.

Similarly, Sakib et al. [9] explore the possibility of a CNN algorithm for fruit recognition. There has been an existing problem with the stacked fruits on a weighing scale because of complexity and similarity. A dataset consisting of 17,823 images from 25 categories has been used in the experiment. The result shows that the best test accuracy of 100% is achieved from 11 to 15 epochs and the best training accuracy of 99.79% is achieved at epoch 15. Besides, the lowest performance loss is approximately 0.0032 in the presence of conv1, pool1, conv2, and ppol2 with dropout. This low loss will help CNN to perform better in fruit recognition.

It is also worth mentioning the work of Soliman et. al [10], which applies ConvNet, one type of CNN algorithm, to mango classification problems. There are several hundred mango varieties, which are different in size, shape, sweetness, skin colour, and flesh colour. For the approach, a 1200 images dataset is being used. The result of the approach proved that CNN can distinguish types of mango efficiently. The result shows that both the training and validation accuracy of the model achieved 100%.

III. Proposed deep learning for dishes recognition

In this project, we have proposed two CNN architectures, and two transfer learning models to classify the five Malaysian daging dishes namely Daging Kunyit, Daging Masak Merah, Daging Rendang, Daging Sate, and Sup Gearbox. The details of our proposed models are described in the subsequent sections.

CNN is currently one of the most popular deep learning networks, and it has the advantage over its predecessors where it can detect critical features automatically without supervision from humans [11]. We proposed two CNN models for image classification of the five daging dishes described earlier.

A. CNN (No Augmentation)

Our first CNN model is a Sequential model that is comprised of 3 convolutional blocks. Before the first convolutional block, the model begins with the first layer normalizing the RGB channel values of the input images in the [0, 255] range to the [0,1] range to make it suitable for use for a CNN model. In each of the convolutional blocks, we have a convolutional layer with padding fixed to 'same' and activation function fixed to 'ReLU' respectively. A max pooling layer with default parameters is also present in each of the convolutional blocks. In the final convolutional block, a flattening layer is used to flatten the input from the previous layers before feeding it to the fully connected (FC) layers. The first FC layer is set with 128 units and activated by a ReLU activation function. The final FC layer which is the output layer is then set to 5 classes.

B. CNN (With Augmentation)

For our second CNN model, we built upon our first CNN model, by keeping all parameters the same, but added a data augmentation layer and a dropout layer. By adding these layers, we hope to prevent overfitting and study the impact of these layers on the original model. Transfer learning is the application of a learnt model to a new problem after solving one existing problem. This is also known as 'fine-tuning' or 'knowledge transfer' [8].

Transfer learning is a machine learning method in which a model generated for one task serves as the foundation for a model developed for another task. This strategy works well if the dataset is small [12].

The goal of transfer learning is to efficiently extract the high-level characteristics of the images, thereby accelerating the target network's learning process. The performance of the model can be improved and converged much faster by transferring the first model's weights or parameters [13].

Transfer learning removes a few of the top layers of a fixed model base and replaces them with new classifier layers and the base model's final levels. Random weights in the final layers of classification are being maintained. This fine-tuning of high-level feature representations in the base model makes it applicable for the specific task [8].

Pre-trained models are used in transfer learning. VGG16 and ResNet50 transfer learning models were used and the results of accuracy were compared with the deployed CNN model. VGG16

VGG16 was created by Simonyan and Zisserman for the ILSVRC 2014 competition. It has 16 convolutional layers and just 3x3 kernels. The authors' approach is similar to AlexNet in that the number of features map or convolution rises as the network depth increases. The network has 138 million parameters [8]. VGG16 is object identification and classification method that can accurately identify 1000 photos of 1000 distinct categories. It is a common image classification technique that works well with transfer learning [14].

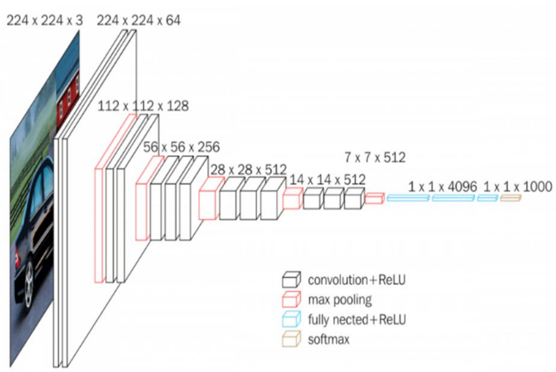

Figure 1 Architecture of VGG16 [15]

The VGG16 pre-trained model is loaded. VGG16 was trained with transfer learning using weights from a model trained with the ImageNet dataset and the last layer is discarded. The discarded last layer of 1000 neurons acts as a classifier for the prior feature map. The remainders of the layer were set to be not trainable after discarding the last layer. This helps to stop the weights from the particular layer from being changed.

C. ResNet50

ResNet-50 is a 50-layer deep convolutional neural network. It has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer [16]. A Residual Neural Network (ResNet) is a type of Artificial Neural Network

(ANN) that builds a network by piling residual blocks on top of one another. ResNet50 used a stack of 3 layers instead of the earlier 2 blocks in ResNet34. This results in much higher accuracy than ResNet34. The 50-layer ResNet achieves a performance of 3.8 x 10^9 Floating point operations [17].

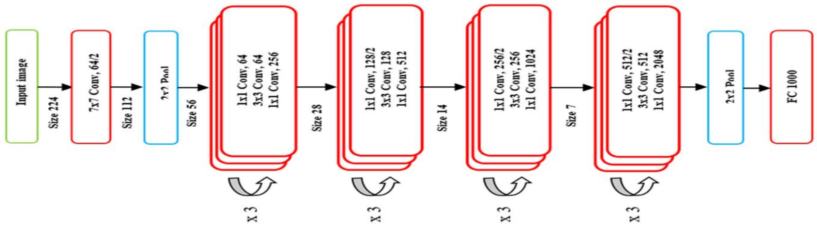

Figure 2 Architecture of ResNet50 [18]

The ResNet50 pre-trained model is loaded. ResNet50 was trained with transfer learning using weights from a model trained with the ImageNet dataset. Similarly, with VGG16, the last layer is discarded. The final layer was set to be not trainable after discarding the last layer.

D. Performance Results

CNN models are 2 out of the 4 models we are discussing in this experiment. The CNN models were built similarly by having the same layers with only 1 major difference – the introduction of the data augmentation layer. A data augmentation layer is added to the original CNN model we have built so we could study the difference in how the model behaves differently after the extra layer. We have also added a Dropout layer to further reduce the possibility of neural network overfitting on the dataset. The outcome of the effect has been recorded and visualized by using Python’s matplotlib library showcased in the latter part of this document

E. Effect of optimizers

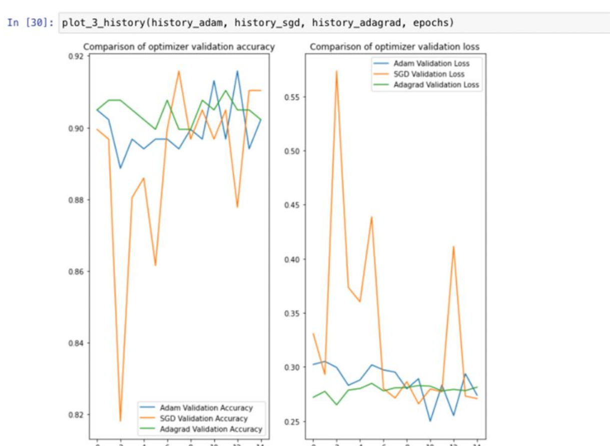

Figure 3 Effect of different optimizers on ResNet50

After selecting the best-performing neural network model, we tried to further optimize the model by using different optimizers and observe on the performance over 15 epochs.

An optimizer is a function or algorithm that modifies the attributes such as weights and learning rate in the neural network to help reduce the overall loss and improve the accuracy [28].

F. Effect of Dense layer

Dense layer is the implementation of the fully-connected layer in Keras [29]. It is the most basic type of neural network layer we use very often while building a neural network.

In general terms, a neural network is considered “deep” when it has more layers of the neural network, while a Dense layer is usually the most common building block of a deep neural network.

G. Effect of filter number

The filters, also known as kernels, are the learnable parameters of CNN. The kernels weight the connections between the neurons (or nodes) in a multi-layer perceptron are the learnable parameters of that model (or feed-forward neural network). The purpose of each layer of filters is to pick up different patterns. Patterns such as edges, corners, and dots, among other things, are captured, for instance, by the first layer of filters. The subsequent layers mix the patterns from the previous ones to create larger designs (like combining edges to make squares, circles, etc.).

H. Effect of number of epochs

Epochs are one of the hyperparameters that determine how many times the learning algorithm will run over the whole training dataset. Each epoch consists of one or more batches in which the neural network is trained using a portion of the dataset. Typically, there are several epochs—hundreds or thousands—which allow the learning method to continue until the model's error has been suitably reduced [31]. The number of epochs a model should run to train depends on a variety of factors relating to both the data and the model's objective [32].

I. Accuracy comparison

Our methodology for this experiment is to first identify the neural network model with the best performance by comparing the models with the same configuration and hyperparameter.

After identifying the best model, we will then adjust the hyperparameters to not only further improve the accuracy, but at the same time, we study the effects of tuning the hyperparameters.

Table1 Comparison of accuracy and loss between training and validation at the 10th epoch.

TABLE I. COMPARISON OF ACCURACY AND LOSS

Models	**Training Accuracy**	**Validation Accuracy**	**Training Loss**	**Validation Loss**
CNN (no augmentation)	0.9193	0.6957	0.2449	1.0247
CNN (with augmentation)	0.6685	0.5435	0.8910	1.6324
ResNet50	0.8861	0.8940	0.3319	0.3039
VGG16	0.7471	0.8071	1.0556	0.7375

For CNN without the data augmentation layer, it can achieve the highest training accuracy at a whopping 0.9193. It also has the lowest training loss of 0.2249. On the training dataset, CNN without the data augmentation layer performs the best by vastly ahead of CNN (with augmentation) and our version of transferred model VGG16. We suspect this is due to the CNN model overfitting the training dataset. An overfit model tends to have high variance as well, it also usually has bad performance on a new dataset that has not been learned. We can see this in the performance of accuracy and loss at the validation set where it is far behind the transfer learning models like ResNet50 and VGG16.

Our CNN model does not seem not to stand out in performance among the neural network models we have studied. A notable difference between this model compared to the CNN model without a data augmentation layer is, that this model does not seem to suffer as much from overfitting. This can be observed by comparing the gap between training accuracy and validation accuracy, the gap is not as wide as the same gap in CNN without a data augmentation layer. From this, we can say the data augmentation layer possibly reduced overfitting and gives the CNN model a better capability of generalizing by adapting better to new unseen data. Although that, the gap between training loss and validation loss is still very large.

IV. CONCLUSION

The classification of Malaysian meat dish cuisine is very important in improving our tourism industry. In our work, several CNN models and Transfer Learning models are introduced in our classification task of several Malaysian meat dish cuisine. Transfer learning models such as VGG16 and ResNet 50 are used in this problem and the accuracy rate are compared with the proposed CNN models. The effect of different hyperparameters, such as optimizer, modification of different layers and learning rate are being investigated in this work. The result showed that ResNet, transfer learning model proposed achieved an accuracy rate of 89.40%, which is higher than proposed CNN deep learning architecture, which only achieved 54.35% in accuracy. Thus, ResNet model is feasible in classifying the Malaysian meat dish cuisine into their respective categories.

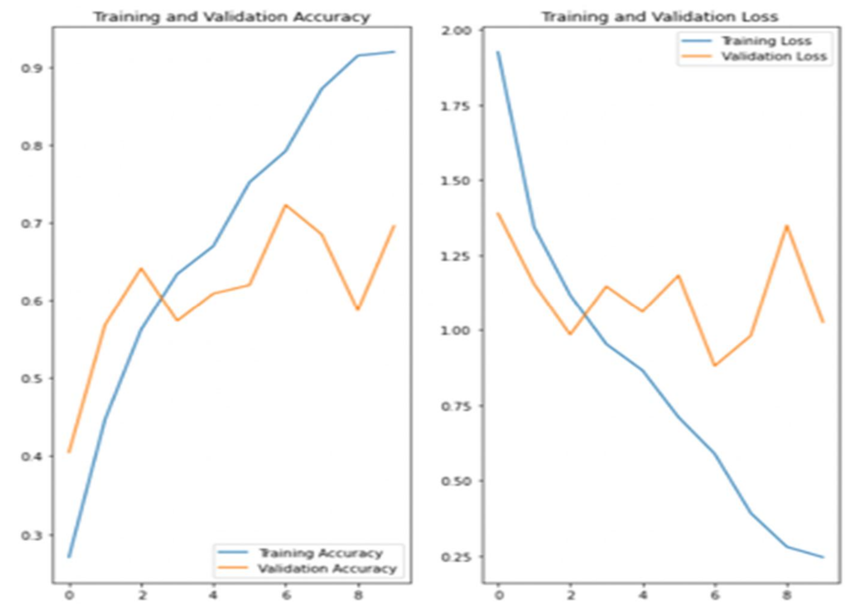

Figure 4 Performance comparison for CNN (no augmentation)

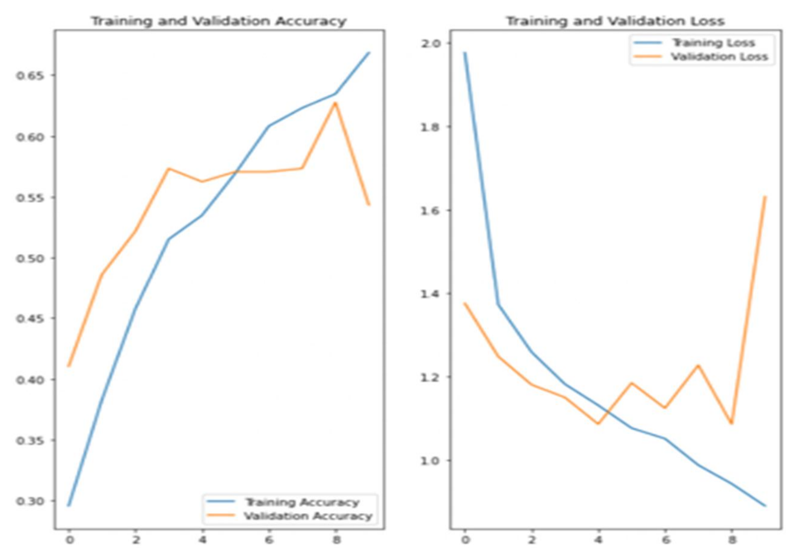

Figure 5 Performance comparison for CNN (with augmentation)

REFERENCES

[1] Z. A. M. A. B. S. Shadman Sakib1, "Implementation of Fruits Recognition Classifier using Convolutional Neural Network Algorithm for Observation of Accuracies for Various Hidden Layers," ArXiv e-Journal , 2019.

[2] A. Kumar, "Different Types of CNN Architectures Explained: Examples," 12 April 2022. [Online]. Available: https://vitalflux.com/different-types-of-cnn-architectures-explained-examples/. [Accessed 10 April 2022].

[3] V. R. C. Y. D. S. B. Sai Sudha Sonali Palakodati, "Fresh and Rotten Fruits Classification Using CNN and Transfer Learning," Rev. d'Intelligence Artif. 34, vol. 34, pp. 617-622, 2020.

[4] Shadman Sakib, Zahidun Ashrafi, Md. Abu Bakr Sidique, "Implementation of Fruits Recognition Classifier," ArXiv e-Journal, vol. 1904.00783, 2019.

[5] Alaa Soliman Abu Mettleq, Ibtesam M. Dheir, Abeer A. Elsharif, Samy S. Abu-Naser, "Mango Classification Using Deep Learning," International Journal of Academic Engineering Research, vol. 3, no. 12, pp. 22-29, 2019.

[6] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," Journal of Big Data, vol. 8, no. 1, pp. 1-74, 2021.
[7] J. I. N.-V. J. C. C.-T. G. R.-R. Jose Luis Rojas-Aranda, "Fruit Classification for Retail Stores Using Deep Learning," SN Applied Sciences, vol. 4, no. 1, p. 74, 2020.
[8] D. E. A.-P. A. P.-S. H. R. G. A. P.-V. O. I. Álvarez-Canchila, "Colombian fruit and vegetables recognition using convolutional neural networks and transfer learning," Journal of Physics: Conference Series, vol. 1547, no. 1, pp. 1-6, 2020.
[9] R. G, "Everything you need to know about VGG16," 23 September 2021. [Online]. Available: https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918. [Accessed 16 July 2022].
[10] A. Kaushik, "Understanding ResNet50 architecture," 2022. [Online]. Available: https://iq.opengenus.org/resnet50-architecture/. [Accessed 16 July 2022].
[11] G. Boesch, "Deep Residual Networks (ResNet, ResNet50)," 2022. [Online]. Available: https://viso.ai/deep-learning/resnet-residual-neural-network/. [Accessed 16 July 2022].
[12] A. Gupta, "Analytics Vidhya," 2021 October 7. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-deep-learning-optimizers/. [Accessed 2022 July 17].
[13] Keras, "Dense layer," [Online]. Available: https://keras.io/api/layers/core_layers/dense/. [Accessed 17 July 2022].