IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.17, No.1, January 2023, pp. 67~78
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.80049

Fishku Apps: Fishes Freshness Detection Using CNN With MobilenetV2

Muthia Farah Hanifa[1], Anugrah Tri Ramadhan*[2], Ni'Matul Husna[3], Nabila Apriliana Widiyono[4], Rhamdan Syahrul Mubarak[5], Adisti Anjani Putri[6], Sigit Priyanta[7]
[1]Universitas Islam Negeri Syarif Hidayatullah, Jakarta, Indonesia
[2]Universitas Dian Nuswantoro, Semarang, Indonesia
[3,4,5,7]Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia
[6]Universitas Udayana, Bali, Indonesia
e-mail: [1]muthiafarah04@gmail.com, *[2]**ramareksotinoyo@gmail.com**,
[3]nimatul.husna@mail.ugm.ac.id, [4]nabilaapriliana@mail.ugm.ac.id,
[5]rhamss794@mail.ugm.ac.id, [6]adistianjaniputri90@gmail.com, [7]seagatejogja@ugm.ac.id

Abstrak

Ikan laut merupakan salah satu komoditas ekonomi yang sangat menjanjikan bagi perekonomian Indonesia. Ikan laut akan menurun kadar proteinnya seiring dengan menurunnya tingkat kesegaran ikan yang akan dikonsumsi. Penggolongan ikan segar dan tidak segar masih banyak orang yang tidak mengetahui akan hal itu, sehingga diperlukan sebuah sistem yang dapat mengklasifikasikan mana ikan segar dan mana ikan yang tidak segar. Penelitian sebelumnya telah berhasil mengklasifikasikan ikan tongkol menggunakan algoritma convolutional neural network (CNN) dengan akurasi sebesar 90%, pada tahap preprocessing pada penelitian ini, dilakukan segmentasi untuk memisahkan objek yang akan diteliti dan background gambarnya, selanjutnya dilanjutkan dengan ekstraksi fitur menggunakan color moment yang bertujuan untuk mendapatkan nilai dari objek yang akan diteliti. Penelitian ini dilakukan untuk meningkatkan nilai akurasi pada pada klasifikasi kesegaran ikan tongkol, dan juga menambahkan beberapa ikan untuk dideteksi kesegarannya seperti bandeng dan kembung menggunakan base model MobilenetV2. hasilnya mampu menghasilkan akurasi sebesar 97%, 94%, dan 93% pada masing-masing ikan. Deteksi kesegaran pada penelitian ini telah diimplementasikan pada aplikasi berbasis mobile Fishku.

Kata kunci— *Ikan segar, CNN, MobilenetV2*

Abstract

Marine fish are one of the most promising economic commodities for the Indonesian economy. Marine fish will decrease in protein content along with the decreasing level of freshness of the fish that will be consumed. There are still many people who do not know about the classification of fresh and unfresh fish, so we need a system that can classify which fish are fresh and which are not. Previous studies have succeeded in classifying tuna using a convolutional neural network (CNN) algorithm with an accuracy of 90%. In the preprocessing stage of this research, segmentation is carried out, which aims to separate the object to be studied and the background image, then feature extraction is carried out using a color moment, which aims to get the value of the object to be studied. This research was conducted to increase the accuracy value in the freshness classification of tuna and also to add some fish for freshness detection, such as mackerel and milkfish, using the MobilenetV2. The results were able to produce accuracy of 97%, 94%, and 93% for each fish. The freshness detection method in this study has been implemented in the Fishku mobile-based application.

Keywords— *Fresh fish, CNN, MobilenetV2*

Received December 12th,2022; Revised January 31th, 2023; Accepted January 31th, 2023

1. INTRODUCTION

Fish is one of the best sources of protein. Indonesia, which is a maritime country, has a wealth of marine fish that are rich in protein. Unfortunately, this has not been utilized optimally because fish consumption in Indonesia is still relatively low, especially for millennials[1]. The classification of fresh and non-fresh fish, which many people do not know, is also one of the reasons for low fish consumption**.**

Fresh fish can be selected manually using the characteristics described in SNI 01-2729.1-2006, which state that there are several characteristics of fresh fish that can be used to determine the freshness of the fish to be consumed, such as changes in color, shape, and smell of fish. In this study, identification of fish freshness was carried out based on color changes that occurred in unfresh fish and their shape. The fish we tested were milkfish, mackerel, and tuna.

Several previous studies that classified fish freshness used only one machine learning model or only one fish in one application that could be classified as freshness.Another problem that we found in previous research and which we want to fix in our research is the large size of the model, so it will be difficult in the deployment process later, and also in this study, we want to increase the accuracy of the previous research.

Research conducted by [2] tested the freshness of milkfish by eye. The dataset used in this study from each class is not up to 100 images, which could be one of the causes of the model not being able to generalize the problem at the test stage properly. Tests were carried out with several base models, such as Xception, MobilenetV1, ResNet50, and VGG16. The final test results show that the model with an architecture that has a convolution network depth such as VGG16 [3] is able to achieve good results, compared to other base models, with an accuracy of 97%. Another research that has been carried out is by [4] which classifies the freshness of tuna based on eye color extracted manually using image segmentation. This study obtained an accuracy of 90% at 10 epochs. The next stage in this research is the development process using Matlab. This study [5] uses a different method from the research that has been discussed, namely the K-Nearest Neighbor (KNN) algorithm, tested using K values of 1, 3, 5, and 7. The best accuracy is obtained when K = 1, with an accuracy of 93%. It is not known what kinds of fish are classified. Apart from being tested on fish eye objects, the CNN algorithm can also be applied to the detection of diseases in the human eye, especially eyes affected by cataracts as has been done in research[6].

2. METHODS

2.1 Fresh Fish

Fish is a source of protein for the human body. But along with the reduced level of freshness of the fish, the protein content of the fish also decreases. Consumption of fish needs to be considered for the level of freshness, so that fish consumers can also benefit from the nutritional benefits of fish consumed as described in [7]. Even fresh fish has a higher nutritional content than processed fish commonly found in the market. Fresh fish have strong physical characteristics, when compared to fish that are not fresh, such as physical appearance on the eyes, flesh, texture, and on the gills. So it is the physical characteristics of fresh fish and not what can be recognized by the convolutional neural network algorithm in this study.

2.2 Digital Image

An image is a representation of the visual appearance of an object. Another definition of an image is a 2-dimensional function or matrix f(x,y), where x and y are spatial coordinates for each value, and f is a magnitude at each coordinate (x,y), which is called intensity or degree. gray in the 0-255 range Each pixel's discrete value is represented by a picture element (pixel), which has an index (x, y) or row and column in the image. Each smallest element is referred to as a pixel, and its value is a numeric value that serves as the basic information of an image.

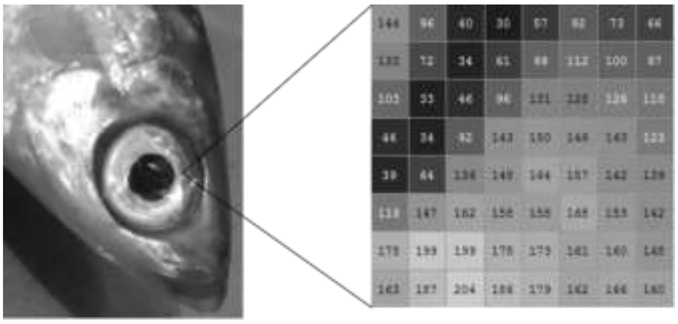


Figure 1 8x8 Pixel Representation of the Milkfish Image

2.3 Augmentation

Data augmentation is a technique for multiplying training data when the model is being trained from datasets that already exist or will be used in research because deep learning requires large datasets to get good results. Because the data in the research to be carried out is classified as being too small for deep learning, data augmentation is carried out to obtain more data in order to achieve the desired accuracy and also to avoid overfitting problems. Here are some instances of data augmentation used in this study.

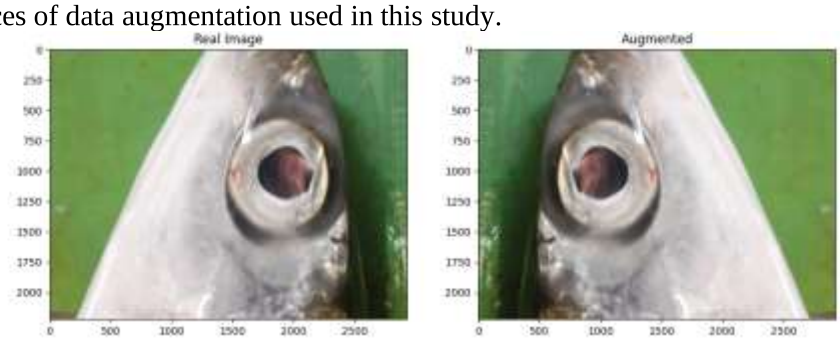


Figure 2 Application of Horizontal Flip Augmentation Data

2.4 Convolutional Neural Network

Convolutional Neural Network (CNN) is one of the algorithms of Deep Learning, which is also one of the developments of the neural network algorithm. The CNN layer consists of 2 types, the first is the Convolutional Layer which is useful for recognizing features by producing a feature map that can recognize special features in an image.

2. 4.1 Convolution

Convolution is a way to recognize a special pattern in an image by generating several feature maps that have been trained. Mathematically, convolution is the product of two matrices that are usually of different sizes; the kernel or filter size is smaller than the size of the image to be convoluted. The equation of convolution is as follows: For initialization, the initial value on the filter is a random value; later, as the model is trained, the parameters in the convolution matrix will be able to recognize patterns based on the model that has been trained with certain data.

$$g(x,y) = h(x,y) * f(x,y) \tag{1}$$

In the mathematical equation above, g is the feature map or convoluted images, next, h is the original image atau image to be convoluted, and f is the kernel or filter of convolution. Meanwhile, all x and y of each function are discrete values of each 2-dimensional image.

2. 4.2 Stride

Stride is a parameter used to determine how many steps to shift the filter in the convolution process. If you use 1 stride, then the filter shift is 1 pixel; if you use 2, the filter shift is 2 pixels, and so on. The filter will then move horizontally and then vertically. In this study, we use stride 1 in the hope that the results will provide more detailed information about specific features in the image and prevent the unnecessary accumulation of information on each pixel in the image. But it requires large computational costs and takes a relatively long time to compute.

The use of stride can also affect how many map features the model generates. The output dimensions of the padding and stride effect matrices can be calculated using the following equation.

$$n_{out} = \frac{n_{in}+2p-k}{s}+1 \quad (2)$$

Where:
nout = The numbers of output features
nin = The numbers of input features
P = Size of padding
S = Size of stride

2. 4.3 Pooling

Pooling is a technique to reduce the size of the image, so that the computational costs will not be too large and the training time will be shorter. In our research, the pooling used is max pooling, which will take the largest value for each part of the image to produce a new output matrix. The size of the convolution output is smaller than the initial image size.

2.5 Fully Connected Layer

The Fully-Connected Layer is part of the CNN that works like a normal neural network. The input from the convolutional-layerLayer is a high-dimensional matrix or tensor which will then be converted into a 1xN dimensional vector, this process is called flattening.

2. 5.1 Activation Function

The activation function is used to calculate non-linear data in the above model. The activation functions used in the above architecture are ReLU and sigmoid which are non-linear activation functions.

1. Rectified Linear Units (ReLU)

   The ReLU activation function [8] is used at each convolution layer, where the ReLU activation function itself has several advantages, including not activating all neurons simultaneously so that it can save time during the training process, and its computational simplicity because it only uses the maximum function so that it can reduce computational costs. the following is the equation of sigmoid.

   $$y = \max(0, x) \quad (3)$$

   The above equation consists of y which is the class to be searched, the max(0, x) function can be interpreted as a function that is useful for determining if the data is a number less than 0, it will be changed to 0, otherwise, data does not change at all.
2. Sigmoid

   The sigmoid activation function is a non-linear activation function. The non-linear activation function is used to determine and separate non-linear data in order for the model to generalize problems with non-linear data variations. Sigmoid has an output from 0 to 1, so it is suitable for predicting a target that has two or binary possibilities.

$$S(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

The above equation can be interpreted as an exponential value raised to the power of the input, which is then multiplied by negative 1. Later, the output of the sigmoid activation function is a value between 0 and 1. So in this study, if the value is more than 0.5, it is a class of fresh fish, whereas if the value is less than 0.5, it is a class of fish that is not fresh.

2. 5.2 Loss Function

The loss function, or what is often referred to as the cost function, is a function that is used to measure how far the predictions obtained from the predicted model are from the dataset label, or "ground truth." By using the error value obtained from the loss function, it is then used to update the parameter weights in the model so that, ideally, along with the more "intelligent" parameters, the error value will also decrease. The loss function that will be used in this research is binary cross-entropy because it is a binary classification.

$$\text{Binary Cross Entropy Loss} = -p \ln q - (1 - p) \ln\ (1 - q) \tag{5}$$

The explanation of the mathematical equation of this function is that p is a label or class that is actually contained in the data, or more commonly known as "ground truth," and q is a class that has been predicted by the label. Because it only requires one neuron at the end of the output layer, binary type loss is expected to be more efficient and save more memory.

2. 5.3 Optimizer

The optimizer serves to minimize the error value obtained from the cost function. The optimization function works convergently, meaning that each iteration will decrease and lead to a value of 0. The optimizer used in this study is adaptive moment estimation (Adam), which is a combination of the gradient descent with momentum optimization function and RMSProp [9] . The proof of Adam's optimization function has been demonstrated in research methods for local convergence analysis in batch mode for deterministic fixed training sets [10].

Notation Algorithm of Adam Optimizer

Require: α
Require: $\beta_1, \beta_2 \in (0, 1)$
Require: w_0
Require: $f(w)$
1: $m_0 = 0, v_0 = 0, t_0 = 0$
2: **While** w not converged **do**
3: $t = t + 1$
3: $m_{t+1} = \beta_1 m_t + (1 - \beta_1)\nabla_w f(w_t)$
4: $v_{t+1} = \beta_2 v_t + (1 - \beta_2)\nabla_w f(w_t) \otimes \nabla_w f(w_t)$
5: $w_{t+1} = w_t - \alpha \frac{\sqrt{1-\beta_2^{t+1}}}{(1-\beta_1^{t+1})} m_t \oslash \sqrt{v_{t+1 \otimes \epsilon}}$
6:**End while**
7:**Return** w_t

Figure 3 Notation Algorithm of Adam Optimizer

The Adam optimization function, which is widely used in machine learning research and deep learning in particular, is depicted in Fig. 3. In our work, the first step to do is to initialize the first moment vector and second moment vector and also initialize timestep and the rest is if parameters did not converge, always update the parameters with respect to the error value, and then result in the desired parameters.

2.6 Transfer Learning with Mobilenetv2

Transfer learning is the implementation of a pre-trained CNN model that has been trained on ImageNet and is then used to deal with different problems, so the model does not need to be trained from scratch again. In this study, the base model used for transfer learning is MobilenetV2 [11]. MobilenetV2 implements inverted residual connection, which is described in this study the results reported on the connecting bottleneck shortcuts perform better than shortcuts connecting the expanded layers. In addition to implementing the residual layer, MobilenetV2 also implements a linear bottleneck that can improve the performance of the model, especially on spurious correlations that have been studied in [12], by anticipating the non-linearity activation function, which often corrupts information in low-dimensional space. Another reason MobilenetV2 is one of the most efficient base models is because the number of parameters tends to be lower when compared to other base models such as VGG.

2.7 Research Flowchart

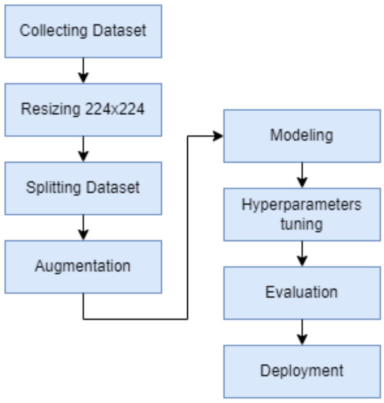


Figure 4 Research Flow Diagram

A research flowchart shows how the research for this study was conducted. As seen in fig. 4, beginning with gathering datasets by capturing pictures of both fresh and non-fresh fish, the size of the data was changed from various sizes to 224x224. Distribution of the training and test data should then be done, followed by augmentation, which tries to increase the training set's data. Next comes modeling, which uses the CNN algorithm to perform the categorization. To determine the appropriate hyperparameters for the case study examples in this research, the following step is hyperparameter tuning. The process of evaluation, which is the measuring stage to determine whether the model that has been created is good or not, is continued after acquiring the best hyperparameters. Then is the deployment to apply the CNN algorithm to cases in the real world through the Fishku application.

3. RESULT AND DISCUSSION

3. 1 Research Condition

The research was conducted using 3 machine learning models, each of which will detect the freshness of tuna, milkfish, and mackerel. In previous studies, fish freshness detection was only able to detect the freshness of one fish per application. Each model was tuned using different hyperparameters, particularly the learning rate and the number of neurons in the fully connected layer, in order to achieve the best results reported in this study.

3. 2 Preprocessing

In the labeling process, the data that has been collected is given a name to be recognized or grouped according to their respective classes. This process is to ensure that the pictures have been entered into each class to be classified, of which in this study there were 2 classes, namely classes with categories of fresh and non-fresh fish, for each type of fish.

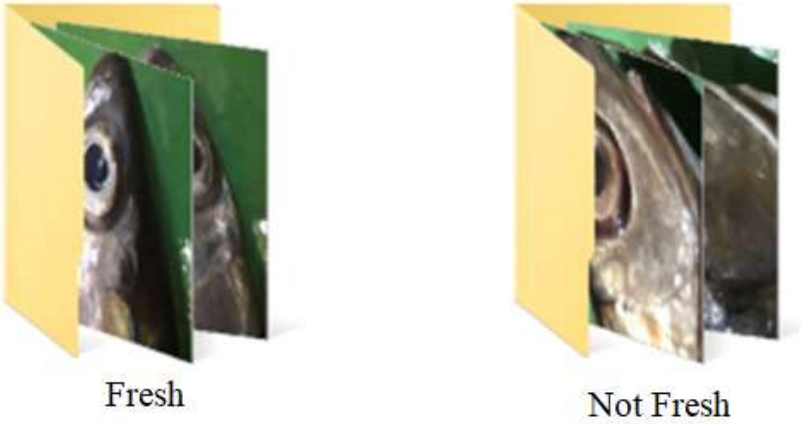


Figure 5 Labeling Dataset for Each Classes

The next stage in preprocessing is to equalize the pixel size between one image and another as in Fig.5, commonly called resizing. The purpose of the resizing process is to reduce the image value and make it easier during the training process, because, in deep learning, the training process requires relatively expensive computational costs and also takes a lot of time. The size of each image is reduced to 224x224 pixels to match the standard input image from the base model that will be used concurrently, making the computation process lighter.

3. 3 Augmentation

The augmentation carried out in this research study has proven to be very helpful in adding datasets and improving the machine learning models that have been created. Some of the augmentations carried out in this study include.

A. Horizontal Flip

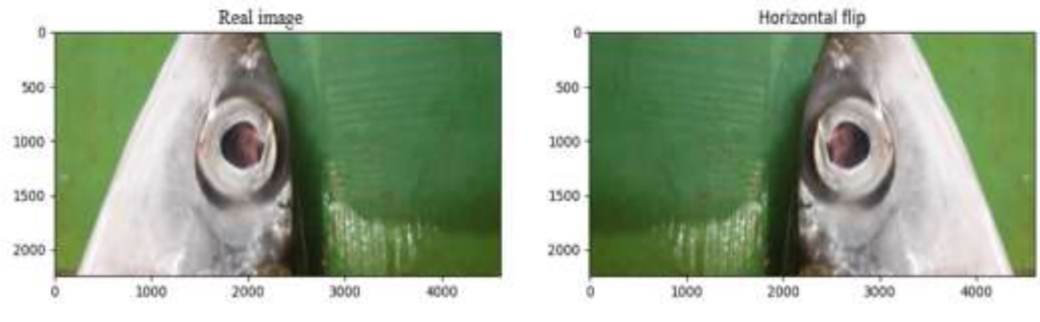


Figure 6 Comparison Between Real Images and Horizontal Flip

Horizontal flip is one of the augmentation techniques used to flip half of the image randomly and horizontal. This is used so that at the test stage there will be images with fish positions, such as horizontal flips, that can be correctly predicted, as has been done in Fig. 6.

B. Vertical Flip

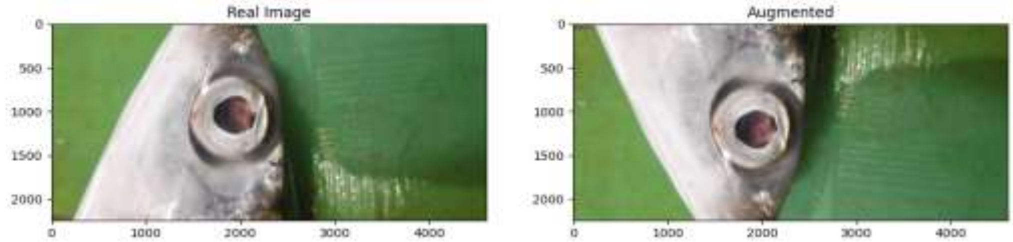


Figure 7 Comparison Between Real Images and Vertical flip

Vertical flip is one of the augmentation techniques used to flip half of the image randomly and vertically. This is used so that at the test stage there will be images with fish positions, such as horizontal flips, that can be correctly predicted, as has been done in Fig. 7.

C. Zoom Range

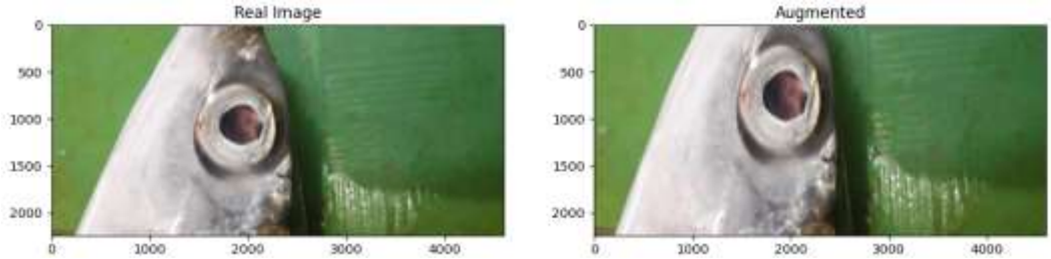


Figure 8 Comparison Between Real Images and Zoom Range

The zoom range is also an augmentation technique that is useful for enlarging the image size of the object to be classified. In this study Fig. 8, the maximum zoom value is 20%, this is done so that the object is not damaged when it is enlarged too much.

D. Rotation Range

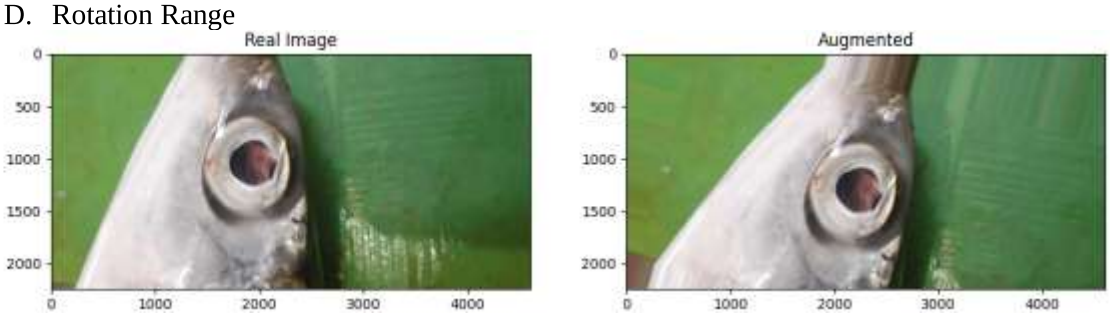


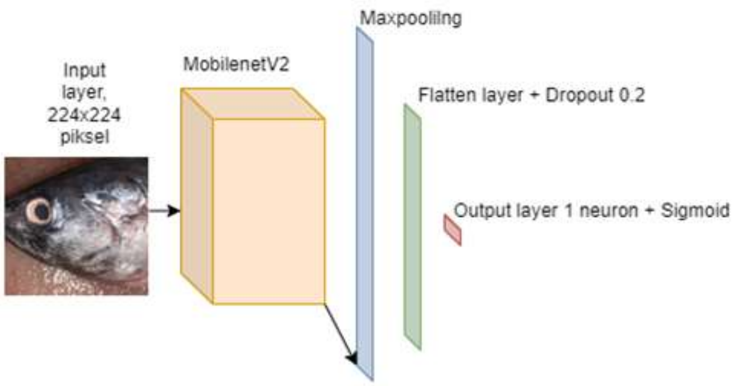
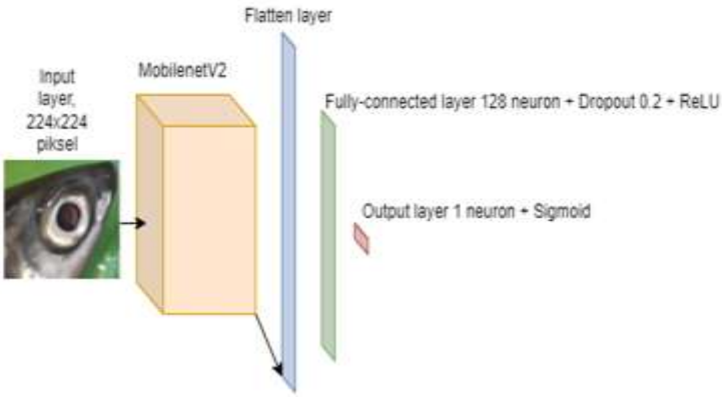
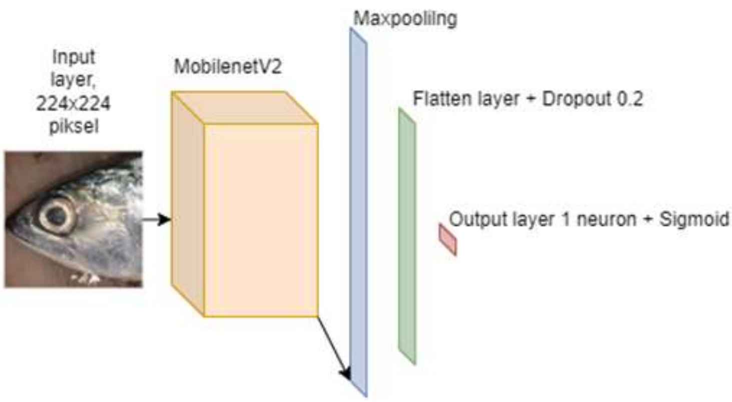
Figure 9 Comparison Between Real Images and Rotation Range

The rotation range that has been done in Fig. 9 adds new data by rotating the image with a percentage of 40% in this study. The choice of a number of 40% was made on the grounds that the images would not have too few rotations.

3. 4 Architecture

We use the MobilenetV2 base mode because the dataset in this study is relatively small. By using the base model, it is hoped that the model can recognize the special patterns found in fresh and non-fresh fish. In the convolution layer that uses MobilenetV2, the parameters contained in the model are in "freeze," so that later during training, the parameters that have been trained on MobilenetV2 are not updated during the backward propagation process. Besides that, this can also save on computational costs, making training more efficient. lighter and shorter training time.

Table 1 Accuracy results obtained from each model and architecture

Fish	Learning Rate	Model	Accuracy
Tuna	1-e2		0.97%
Milkfish	3-e4		0.94%
Mackerel	1-e2		0.93%

3. 5 Test Result

Table 2 Random Test Result

No.	Kind of fish	Image	Prediction	Ground Truth
1.	Milkfish		Fresh	Fresh
			Fresh	Fresh
			Fresh	Non-Fresh
2.	Mackerel		Fresh	Fresh
			Non Fresh	Non Fresh
			Non Fresh	Non Fresh
3.	Tuna		Non Fresh	Non Fresh
			Fresh	Fresh
			Fresh	Fresh

The tests shown above were randomly selected from the test data set. Several models have been able to recognize the categories of fresh and non-fresh fish from each fish used to test the model. Table 2 shows that the model can predict the freshness of some fish correctly, but in test 3, when classifying the freshness of milkfish, the model is less able to recognize images that are slightly tilted. This means that the model can only recognize images where the object's eyes are straight and close to the camera.

3. 6 Deployment

At the deployment stage, the devices used are Android and iOS devices to run machine learning models. The model that has been trained will then be deployed on a cloud service provider's system, namely Google Cloud Service. Cloud services are used on the grounds that there are three models and a large size, so it is not possible to deploy them directly on smart phone devices.

Figure 6 How Machine Learning Works on Fishku App

In the illustration above, after the model is deployed, it is then made so that the model can be called an API. Later, the mobile device only needs to send an image of the fish eye, which will be predicted to be fresh or not, and then just send a request, which will then be returned as a response from the backend side. Here is what it looks like on a mobile device.

4. CONCLUSION

In this study, it has succeeded in increasing the accuracy of tuna from 90% to 93% by implementing MobilenetV2 on the classification problems in this study. On the application side, the three machine learning models have been deployed well, so they can be used by a wider community.

ACKNOWLEDGMENTS

We would like to thank the Bangkit Program, which is supported by Google, Gojek, Traveloka, and DPRTM, for providing extraordinary support and providing financial assistance so that this research can be carried out and completed.

REFERENCES

[1] D. S. Untari, T. A. Wibowo, and R. Anwar, "Minat Konsumen Millenial Terhadap Konsumsi Ikan Air Laut dan Ikan Air Tawar," *J. FishtecH*, vol. 11, no. 1, pp. 30–38, 2022.

[2] M. Sholihin, "Identifikasi Kesegaran Ikan Berdasarkan Citra Insang dengan Metode Convolution Neural Network," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 3, pp. 1352–1360, 2021, doi: 10.35957/jatisi.v8i3.939.

[3] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.1556

[4] F. F. Romadhoni, "Klasifikasi Kesegaran Ikan Tongkol Berdasarkan Warna Mata Menggunakan Metode Backpropagation," vol. 2, no. 2, pp. 405–414, 2022.

[5] S. Saputra, A. Yudhana, and R. Umar, "Identifikasi Kesegaran Ikan Menggunakan Algoritma KNN Berbasis Citra Digital," *Krea-TIF J. Tek. Inform.*, vol. 10, no. 1, pp. 1–9, 2022, doi: 10.32832/kreatif.v10i1.6845.

[6] I. Weni, P. E. P. Utomo, B. F. Hutabarat, and M. Alfalah, "Detection of Cataract Based on Image Features Using Convolutional Neural Networks," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 1, p. 75, 2021, doi: 10.22146/ijccs.61882.

[7] L. J. Damongilala, "Kandungan Gizi Pangan Ikan," *Patma Media Graf. Bandung*, pp. 1–60, 2021.

[8] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," Mar. 2018, [Online]. Available: http://arxiv.org/abs/1803.08375

[9] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980

[10] S. Bock and M. Weis, "A Proof of Local Convergence for the Adam Optimizer," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2019-July, no. July, pp. 1–8, 2019, doi: 10.1109/IJCNN.2019.8852239.

[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[12] P. W. Koh *et al.*, "Concept Bottleneck Models," *37th Int. Conf. Mach. Learn. ICML 2020*, vol. PartF16814, pp. 5294–5304, 2020.