Journal of Computing and Applied Informatics (JoCAI) Vol. 2, No. 1, 2021 | 60 – 70

DATA SCIENCE
Journal of Computing and Applied Informatics

MobileNets-V1 Architecture for Web Based Fish Image Classification

Herlambang Dwi Prasetyo[1], Pandu Ananto Hogantara[2], and Ika Nurlaili Isnainiyah[3]

[1,2,3]Department of Computer Science, Faculty Of Computer Science, University Of Pembangunan Nasional Veteran Jakarta, Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, 12450, Jakarta, Indonesia

Abstract. Recently, the research study about fish identification become a very challenging to researchers. Climate and environmental changes have a major impact on fish species and their environment. To identify fish using manual process is time consuming and need effort to gather samples in different environment. The identification of fish species is performed by using feature extraction and a series of features. Generally, the characteristic is divided into two groups namely general characteristics and anatomical features. General characteristics is characteristic that can be seen directly without the aid of tools. The characteristics include color, texture, and fiber direction. Although, manual is performed by expert but is possible that identification is not accurate. Therefore, to overcome the problem, we create a web-based application for identifying fish by using image as input. We use 10 class data with 300 images for each class. Then, we split into training and testing with 80:20 ratio. The application was developed by using the MobileNets-V1 model. The proposed method has accuracy on 89 %, that obtain from training score is 91.04%, validation is 88,96%. This score is higher than other methods that used in this application. Total time for computation process is about 127 minutes.

Keyword: Deep Learning, Epoch, Fish Image Classification, MobileNet-V1, Transfer Learning,

Received 28 May 2021 | Revised 25 July 2021 | Accepted 28 July 2021

1 Introduction

Observing the diversity of behavior and life patterns of fish is very important so that humans can gain insight and knowledge about marine ecological systems and ecosystems [1]. Climate and environmental changes have a major impact on fish species and fish habitats. The manual method is usually time consuming and requires a lot of effort to obtain samples in different environments [2]. Through structured observations about the behavior of various fish species through

*Corresponding author at: Department of Computer Science, Faculty Of Computer Science, University Of Pembangunan Nasional Veteran Jakarta, Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, Jakarta, Indonesia

E-mail address: herlambangdwi.prasetyo@gmail.com (Herlambang Dwi Prasetyo), panduanantoh@gmail.com (Pandu Ananto Hogantara), nurlailika@upnvj.ac.id (Ika Nurlaili Isnainiyah)

ISSN: 2580-6769 | e-ISSN: 2580-829X | https://doi.org/10.32734/jocai.v5.i2-6291
Journal Homepage: https://jocai.usu.ac.id

calculation and distribution of ecosystems, it can provide important information about the ecological conditions of the sea which are commonly used as parameters for observing changes in the marine environment [1]. Fish identification can be done with digital image processing techniques, the technological developments by using a digital camera and computer vision to make easy for humans to identify various types of fish visually. Also, it can help researchers to observe their movement patterns is affected by various condition and provide whole activity pattern based on the deeper knowledge [3].

In recent years the topic of research to identify and recognize fish species has become a very challenging challenge for researchers [4]. The identification of fish species is carried out using feature extraction and is based on a series of features. Generally, the characteristic is divided into two groups namely general characteristics and anatomical features. General characteristics is characteristic that can be seen directly without the aid of tools. The characteristics include color, texture, and fiber direction. this technique can only be done for those who are experienced and it is possible that this method often experiences misidentification. Although, manual is performed by expert but is possible that identification result is not accurate. Therefore, to overcome the problem, we create a web-based application for identifying fish by using image as input. This application was built using the MobileNets-V1 model. MobileNets-V1 is a Convolutional Neural Network (CNN) architecture which address computation with a huge amount of data. Convolutional Neural Network is a deep learning method that is able to recognize an object in a digital image with high computation and need a huge amount of data. CNN's capability is claimed to be the best method in terms of object detection and object recognition [5]. The fundamental difference between the MobileNets-V1 architecture and CNN architecture is the using of a convolutional layer or filter thickness layer that matches the thickness of the input image. MobileNets-V1 divides convolution into depth wise convolution and pointwise convolution [5]. MobileNets-V1 has been proved to show excellent results in many fields of studies. Research conducted by Rajbongshi et al. (2020) that applied MobileNets-V1 architecture to classify rose's diseases has successfully obtained an accuracy of 95.63% [6]. Then, research conducted by Suharto et al.,(2020) which also applied MobileNets-V1 architecture, to classify freshwater fish species has successfully obtained an accuracy of 90.00% [7]. MobileNets-V1 has also been used to detect diseases, which is done by Ansar et al. to detect breast cancer with an accuracy score of 86.8% [8]. Therefore, in this research we apply MobileNets-V1 to classify fish.

2 Research Methodology

2.1 Related Works

There are many researches about fish identification has proposed. Pornpanomchai, et al.(2013) using Artificial Neural Networks and Euclidean Distance Method (EDM) algorithm to fish image recognition by using dataset of 300 images of testing and 600 images of training that obtain accuracy about 81,67%v[9]. Research conducted by Hernandez-Serna, et al. with the title

Automatic Identification of Species with Neural Networks using an Artificial Neural Network with a dataset of 697 fish images yields an accuracy of 91.65% [10]. Research conducted by Spampinato, et al. with the title 'Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos using the Moving Average Detection Algorithm and Adaptive Gaussian Mixture Model algorithm with a dataset of 20 underwater videos to detect, track and count fishes produces an accuracy of 85% [11].

2.2 MobileNets-V1 Architecture

In this paper we use MobileNets-V1 which has architecture as shown as Table 1.

Table 1. MobileNets-V1 Architecture

Type/Stride	Filter Shape	Input Size
Conv / s2	3 x 3 x 3 x 32	224 x 224 x 3
Conv dw / s1	3 x 3 x 32 dw	112 x 112 x 32
Conv / s1	1 x 1 x 32 x 64	112 x 112 x 32
Conv dw / s2	3 x 3 x 64 dw	112 x 112 x 64
Cont / s1	1 x 1 x 64 x 128	56 x 56 x 64
Conv dw / s1	3 x 3 x 128 dw	56 x 56 x 128
Conv / s1	1 x 1 x 128 x 128	56 x 56 x 128
Conv dw / s2	3 x 3 x 128 dw	56 x 56 x 128
Conv / s1	1 x 1 x 128 x 256	28 x 28 x 128
Conv dw / s1	3 x 3 x 256 dw	28 x 28 x 256
Conv / s1	1 x 1 x 256 x 256	28 x 28 x 256
Conv dw / s2	3 x 3 x 256 dw	28 x 28 x 256
Conv / s1	1 x 1 x 256 x 512	14 x 14 x 256
5 x Conv dw / s1	3 x 3 x 512 dw	14 x 14 x 512
5 x Conv / s1	1 x 1 x 512 x 512	14 x 14 x 512
Conv dw / s2	3 x 3 x 512 dw	14 x 14 x 512
Conv / s1	1 x 1 x 512 x 1024	7 x 7 x 512
Conv dw / s2	3 x 3 x 1024 dw	7 x 7 x 1024
Conv / s1	1 x 1 x 1024 x 1024	7 x 7 x 1024
Avg Pool / s1	Pool 7 x 7	7 x 7 x 1024
FC / s1	1024 x 1000	1 x 1 x 1024
Softmax / s1	Classifier	1 x 1 x 1000

MobileNets is one of many CNN architectures is available for mobile applications. The advantages of this CNN architecture are the thickness of the convolution filter shown in the picture. In this way, size of produced model and the bottleneck layer contained in the input and output is saved, making training process more accurate and faster. By using the MobileNets architecture on CNN will reduce the need for redundant calculations, so it is suitable for use on mobile devices/smartphones. MobileNets divides two twitch layers into deep convolution and point convolution. Then, use a process to change the result of the feature map pool layer to a vector shape. This process is called the fully connected layer [12].

2.3 Research Design

This research classifies 10 genus groups, namely *lutjanus, macropharyngodon, oxycheilinus, pervagor, plectropomus, pseudanthias, pseudocheilinus, scolopsis, thalassoma,* and *wetmorella*. by using the Convolutional Neural Network (CNN) algorithm and MobileNets architecture. The main process is training data. This process aims to generate a model that will be used for testing data. The parameter to measure the success rate of the model is the accuracy. The accuracy of the model is determined by evaluating against testing data. The best accuracy of the model will be used and deployed to website-based applications.

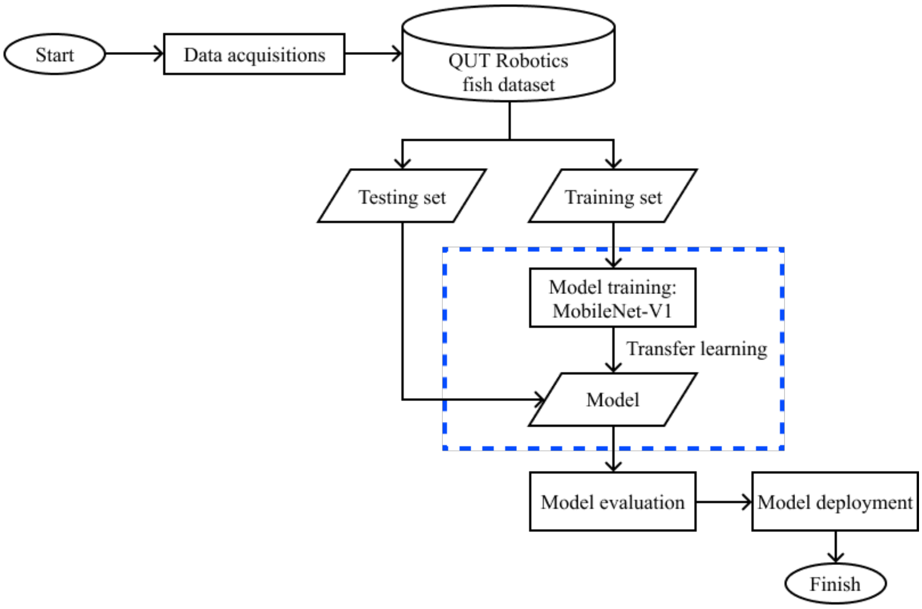

Figure 1. Research Design

This research begins with collecting data. Then, the data that has been collected will be divided into two parts, namely training and testing data. The training data will be used to train the model while the testing data will be used to evaluate the performance of the model. In model training phase, training data are divided into validation data which is used to estimate the model's performance and define whether the model is overfitting or underfitting. The model training was carried out in three scenarios based on number of epochs. We train model with transfer learning on top of the MobileNets-V1 architecture. After the final model is obtained, the model is evaluated to reach performance measurement which shown as confusion matrix in order to calculate accuracy score. In the final stage, we deploy model to design web-based application.

3 Result and Discussion

3.1 Data Acquisitions

The dataset used in this research are the QUT Robotics fish dataset [13] that we obtain from Kaggle [14]. The dataset consists of 3960 images collected from 468 fish species. We group the fish based on its genus, for example, *lutjanus adetii* and *lutjanus argentimaculatus* will be grouped as *lutjanus*. In total, there are 194 genus groups and we only take 10 genus groups, namely *lutjanus, macropharyngodon, oxycheilinus, pervagor, plectropomus, pseudanthias, pseudocheilinus, scolopsis, thalassoma,* and *wetmorella*. The sample data for each class is highly imbalanced, so in order to balance the dataset we also scrap fish images from the internet until the total number of sample data for each class reaches 300 images. Therefore, there are 3000 images in total that were used in this research as illustrated on Figure 2.

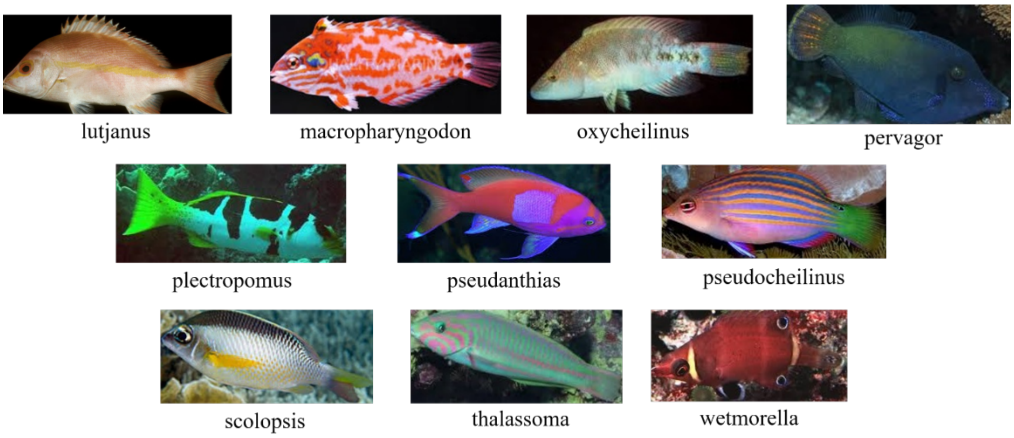

Figure 2. Fish Image Sample

3.2 Data Pre-Processing

The fish images in dataset which we obtain have various size, so that we crop the images manually without make the images is different to original image.Next, we divide the dataset into a training set and testing set with 80:20 ratio. We also divide the training set into a validation set with the ratio of the validation set is 20%. Deep learning models generally work better with huge amounts of data, but it is not an easy task to collect huge amounts of data as we must pre-process it manually. For resolving this issue, image augmentation is done on this stage. The image augmentation techniques are rotating, zooming, shearing, flipping, brightness adjustment, and normalizing the pixel value by dividing it with 255. We apply all the image augmentation techniques on training set to increase the amount of data and to enhance the overall dataset. In testing and validation set only pixel normalization is applied. To apply transfer learning, we also should match image resolution with size of input layer in MobileNets-V1 model which is 224x224.

3.3 Model Training

Firstly, we would like to explain about the experimental environment used in this research as follows: Intel Core i7-4510U@2.00Ghz, NVIDIA GeForce 940M/2GB/DDR3, and 12GB/DDR3. In this research, we propose MobileNets-V1 pre-trained by ImageNet as the base model with transfer learning techniques. There are 86 layers in the base model and since we have a very limited resource of computation, we only unfreeze eight layers to make the eight first layers of MobileNets-V1 trainable. The purpose of unfreezing selected layer is to force the weights to be tuned from generic feature maps to features associated to our dataset [15]. On top of the MobileNets model we add a global average pooling (GAP) layer which must be added before we can proceed to the classification phase. This GAP layer function is to down-sample the size of the feature map simply by taking the average of the whole feature map and as compared to the traditional data flattening layer, GAP layer is capable of improving the robustness of the model significantly [16]. Next, we add a fully connected layer with the size of 512 which is activated by the ReLu activation function so that the model will be able to learn more complex patterns. The last layer is the classification layer which is a fully connected layer with the size of 10 which is activated by the softmax activation function. Table 2 shows the summary of this new model.

Table 2. Summary of Proposed Model

Layer (type)	Output Shape	Parameters
Mobilenet_1.00_224	(7 x 7 x 1024)	3,228,864
Global average pooling (2d)	1024	0
Fully connected layer	512	524,800
Fully connected layer	10	5130

3.4 Model Evaluation

We have five training scenarios based on the number of epochs which is {30, 40, 50, 60, 70} The optimizer we use is Adam optimizer with learning rate of 1e-5. Table 3 shows the training results for each epoch scenario.

Table 3. Training Result on Different Epochs

Epoch	Training Accuracy	Validation Accuracy	Duration (minute)
30	82,27%	83,75%	48
40	85,56%	84,38%	63
50	88,01%	87,71%	80
60	89,03%	90,21%	95
70	91,04%	88,96%	127

We choose the model that is trained on 70 epochs as our final model as it is producing the best accuracy. Since this research is about classification, we evaluate the model using a confusion matrix towards the testing set which is shown in Table 3 and we also plot the training history that is shown on Figure 3.

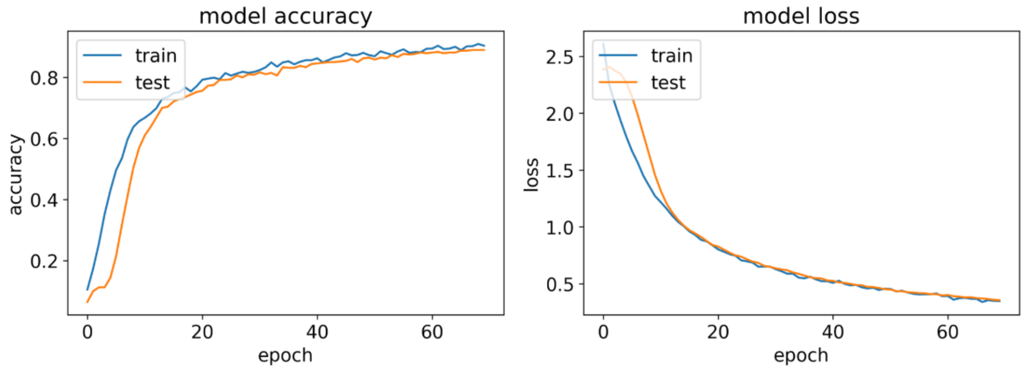

Figure 3. Model Training History

Figure 3 show the model use 70 epochs, the MobileNets model produce the good result because there is no overfit between the accuracy and loss model.

Based on the confusion matrix shown in Table 4, the accuracy generated by the model that is trained with 70 epochs and learning rate of 1e-5 is 89,00% that can be calculated as follows [17]:

$$Accuracy = \frac{1}{B_T}\sum_{i=1}^{N_c} M_i$$

$$= \frac{52+58+49+54+53+53+58+54+50+53}{600} \times 100\%$$

$$= 89{,}00\%$$

Table 4. Confusion Matrix

Confusion Matrix		Predicted Class									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Actual Class	**0**	52	0	1	0	3	2	1	1	0	0
	1	0	58	0	0	0	0	0	0	1	1
	2	1	0	49	2	3	0	2	1	3	0
	3	1	4	0	54	3	0	0	0	1	0
	4	0	1	1	2	53	0	0	3	0	0
	5	2	0	0	1	1	53	2	1	0	0
	6	0	0	1	1	0	0	58	0	0	0
	7	5	0	1	0	0	1	0	54	1	0
	8	1	2	3	0	2	0	1	1	50	0
	9	0	1	1	2	0	0	3	0	0	53

Class label: 0 (*lutjanus*), 1 (*macropharyngodon*), 2 (*oxycheilinus*), 3 (*pervagor*), 4 (*plectropomus*), 5 (*pseudanthias*), 6 (*pseudocheilinus*), 7 (*scolopsis*), 8 (*thalassoma*), 9 (*wetmorella*)

Further, we also compare our proposed model with similar research that uses the same dataset as ours. Although our model accuracy is lower, but our model still produces similar accuracy which can still be compared. Table 5 shows model comparison with similar research.

Table 5. Comparison of Proposed Model with Similar Research

Research	Validation Accuracy	Testing Accuracy
Khalifa, et al. [4]	97,10%	85,59%
Iqbal, et al. [2]	98,20%	90,48%
Proposed model	88,96%	89,00%

In Table 5 our model has lower accuracy than the others. Our model received a testing accuracy of 89.0%, which is better than the model created by [4]. The model formed by [4] is a modification of the AlexNet model which gets an accuracy of 85.59%. Then, when our model is compared to [2], the accuracy is lower by 1.48%. The model by [2] is a modified and reduced version of AlexNet with an additional dropout layer before the classification layer. This model managed to obtain testing accuracy of 90.48%. In this research only the eight first layers of the MobileNets-V1 are unfreeze since we have very limited computation power with low memory of GPU. Therefore, pattern on the data cannot be fully recognized by the model, resulting in low accuracy. If the layers on the model are fully unfreeze, the model will produce much higher accuracy.

3.5 Model Deployment

At the deployment model stage, the model deployed on the application system is a model that has been trained on 70 epochs with training accuracy of 91,04% and testing accuracy of 89.00%. The model with this evaluation value is deployed in a website-based application using Python Flask and Bootstrap, the screen display of the application is available in Figure 4 and Figure 5. On the main page of the application shown in Figure 4, user input a fish image. After image is uploaded, the user press prediction button to get the prediction result what the kind of species as shown as prediction result page on Figure 5.

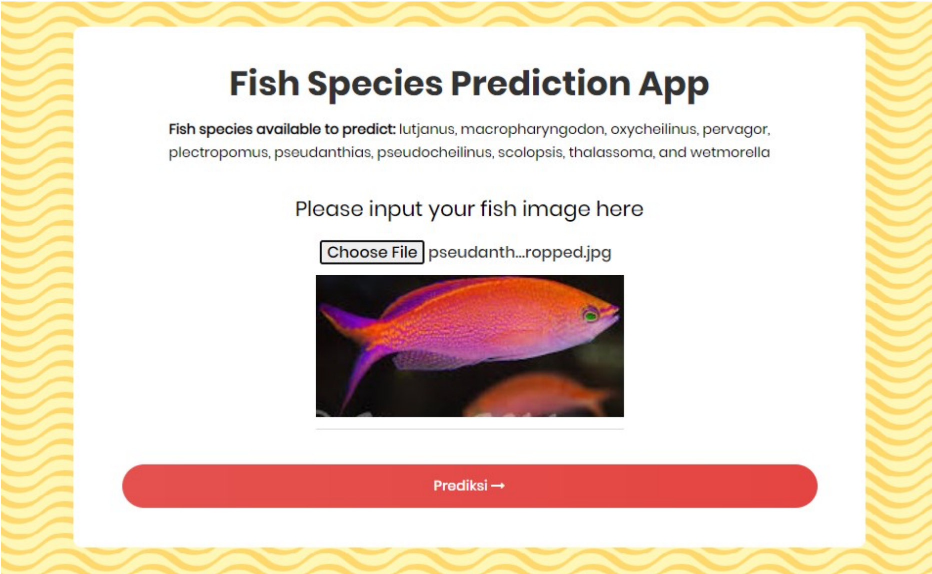

Figure 5. Home Page

Figure 4. Prediction Result

4 Conclusion

The proposed method of classification using MobileNets-V1 of fish species gives an accuracy of 89.00%, train score is 91.04%, validation is 88.96% which is excellent and comparable with the other current implemented methods used for this application. In research, we use 10 class data with 300 images for each class. Then, we split the dataset into a training and testing set with 80:20 ratio and we also split the training data into a validation set. Hence the proposed approach can certainly be used for real time applications as the computation time is 127 minutes. The method couldn't achieve 100% accuracy as some images couldn't be classified accurately due to limited time and resources of GPU memory. Further, we will improve our algorithm by implementing image enhancement techniques to counter for the lost features in the images.

REFERENCES

[1] D. Rathi, S. Jain, and S. Indu, "Underwater Fish Species Classification Using Convolutional Neural Network and Deep Learning," *arXiv*, pp. 1–6, 2018.

[2] M. A. Iqbal, Z. Wang, Z. A. Ali, and S. Riaz, "Automatic Fish Species Classification Using Deep Convolutional Neural Networks," *Wirel. Pers. Commun.*, vol. 116, no. 2, pp. 1043–1053, 2021, doi: 10.1007/s11277-019-06634-1.

[3] S. Fauzi, P. Eosina, and G. F. Laxmi, "Implementasi Convolutional Neural Network Untuk Identifikasi Ikan Air Tawar," *Semin. Nas. Teknol. Inf.*, pp. 163–167, 2019.

[4] N. E. M. Khalifa, M. H. N. Taha, and A. E. Hassanien, *Aquarium Family Fish Species Identification System Using Deep Neural Networks*, vol. 845. Springer International Publishing, 2019.

[5] Hendriyana and Y. H. Maulana, "Identifikasi Jenis Kayu menggunakan Convolutional Neural Network dengan Arsitektur Mobilenet," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 70–76, 2020.

[6] A. Rajbongshi, T. Sarker, M. M. Ahamad, and M. M. Rahman, "Rose Diseases Recognition using MobileNet," *4th Int. Symp. Multidiscip. Stud. Innov. Technol. ISMSIT 2020 - Proc.*, 2020, doi: 10.1109/ISMSIT50672.2020.9254420.

[7] E. Suharto, Suhartono, A. P. Widodo, and E. A. Sarwoko, "The use of mobilenet v1 for identifying various types of freshwater fish," *J. Phys. Conf. Ser.*, vol. 1524, no. 1, pp. 1–6, 2020, doi: 10.1088/1742-6596/1524/1/012105.

[8] W. Ansar, A. R. Shahid, B. Raza, and A. H. Dar, "Breast cancer detection and localization using mobilenet based transfer learning for mammograms," *Commun. Comput. Inf. Sci.*, vol. 1187 CCIS, pp. 11–21, 2020, doi: 10.1007/978-3-030-43364-2_2.

[9] C. Pornpanomchai, B. Lurstwut, P. Leerasakultham, and W. Kitiyanan, "Shape-and Texture-Based Fish Image Recognition System," 2013.

[10] A. Hernández-Serna and L. F. Jiménez-Segura, "Automatic identification of species with neural networks," *PeerJ*, vol. 2014, no. 1, 2014, doi: 10.7717/peerj.563.

[11] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. B. Fisher, "Detecting ,Tracking And Counting Fish In Low Quality Unconstrained Underwater Videos." [Online]. Available: http://ecogrid.nchc.org.tw.

[12] N. Purnama, "Herbal Plant Detection Based On Leaves Image Using Convolutional Neural Network With Mobile Net Architecture," vol. 6, no. 1, 2020, doi: 10.33480/jitk.v6i1.1400.

[13] K. Anantharajah *et al.*, "Local inter-session variability modelling for object classification," *2014 IEEE Winter Conf. Appl. Comput. Vision, WACV 2014*, pp. 309–316, 2014, doi: 10.1109/WACV.2014.6836084.

[14] S. Sripaad, "Fish Species Image Data," 2020. https://www.kaggle.com/sripaadsrinivasan/fish-species-image-data/ (accessed May 01, 2021).

[15] Q. Xiang, G. Zhang, X. Wang, J. Lai, R. Li, and Q. Hu, "Fruit image classification based

on Mobilenetv2 with transfer learning technique," Oct. 2019, doi: 10.1145/3331453.3361658.

[16] S. I. Saedi and H. Khosravi, "A deep neural network approach towards real-time on-branch fruit recognition for precision horticulture," *Expert Syst. Appl.*, vol. 159, p. 113594, 2020, doi: 10.1016/j.eswa.2020.113594.