Identification of yellowfin tuna (*Thunnus albacares*), mackerel tuna (*Euthynnus affinis*), and skipjack tuna (*Katsuwonus pelamis*) using deep learning

I Ayuningtias, I Jaya* and M Iqbal

Department of Marine Science and Technology, Faculty of Fisheries and Marine Science, IPB University, Dramaga, Bogor 16680, West Java, Indonesia

*E-mail: indrajaya@apps.ipb.ac.id

Abstract. Yellowfin tuna (*Thunnus albacares*), mackerel tuna (*Euthynnus affinis*), and skipjack tuna (*Katsuwonus pelamis*) have important economic values for the capture fisheries in Indonesia. Activities of identifying these fish and other types of tuna have been done manually, which can lead to errors and ultimately affect statistics, stock estimates, or traceability. The aim of this research is to use deep learning methods in identifying three species of tuna, specifically yellowfin tuna, mackerel tuna, and skipjack tuna. YOLO's newest model, YOLOv5, was used to identify the fish. The number of epochs that produces the optimum accuracy value for use in the YOLOv5 model is 400. The values for training loss, accuracy, precision, recall and F1-Score when the model is learning with a total of 400 epochs are 0.000253, 95%, 98.1%, 93.9%, and 96%. Based on these results, the three species of tuna can be identified with high accuracy.

Keywords: deep learning, epoch, traceability, training loss, YOLOv5

1. Introduction

The fisheries sector has an important contribution to economic growth in Indonesia. Marine fish that have great resources and potential are pelagic fish groups such as yellowfin tuna (*T. albacares*), mackerel (*E. affinis*), and skipjack (*K. pelamis*). These fish have important economic value for the capture fisheries sector in Indonesia [1]. The economic value found in yellowfin tuna, mackerel, and skipjack tuna are the main commodities of the fisheries sub-sector for local and global fish consumption. The morphology of yellowfin tuna, mackerel, and skipjack have almost the same characteristics with each other. The morphological forms of fish that are almost the same cause frequent errors in the identification process of these fish. Fish identification activities by fishermen, the community or certain parties are still done manually. The manual identification process can lead to identification errors that can affect statistical calculations, stock estimates, traceability of yellowfin tuna, mackerel, and skipjack [2].

The number of fisheries personnel who are tasked to identify fish is very limited. The difficulty in identifying fish is due to the similarities between one type of fish to another, so it is necessary to have an easier way but accurate to identify similar fishes. The fish identification and recognition system that will be built is expected to make it easier for the community, or related parties to identify fish without having to depend on or meet with experts directly [3].
Identification problems that usually occur in the field are when fish sorting is done manually by fish auction officers and the community for processing, auctioning or selling. Processing of fishery products must be carried out with proper species identification processes, errors in identification can affect the traceability of the processed products made. Fish auction activities carried out at TPI (Fish Auction Place) begin by sorting fish by type. Fish identification by auction officers is still made manually by looking at the morphological characteristics of the fish. The time required to sort fish, if done manually, takes quite a long time. Based on this, a system is needed that can help auction officers so that they can shorten the time in the fish sorting process. There is a method that can make it easier for the public or fish auction officers to identify fish without the need to know their characteristics directly. The method is called deep learning which is based on an Artificial Neural Network (ANN).

Deep learning is a field of machine learning that utilizes many layers of nonlinear information processing to perform feature extraction, pattern recognition, and classification [4]. Deep Learning learns to classify directly from images or sounds [5]. Deep Learning is a learning method that utilizes a multi-layered ANN. The ANN is made to resemble the human brain and the neurons are connected to each other to form a very complex network of neurons [6].

2. Materials and Methods
The proposed approach for \textit{T. albacares}, \textit{E. affinis}, and \textit{K. pelamis} detection and segmentation is performed in five steps: data collection, data labelling, training, and finally, evaluation of the model.

2.1 Data Collection
The dataset is a data from verified source, hence it can be used for research as a valid data source. The dataset used for this study is an image of yellowfin tuna (\textit{T. albacares}), mackerel (\textit{E. affinis}), and skipjack tuna (\textit{K. pelamis}). The dataset that has been collected is divided into 2 parts, namely the training set and the validation set with a ratio of 80\%: 20\%. The training set and validation set must be varied in order to get training results with good accuracy. A training set is a data set that is used to train or build a model. A development set or validation set is a data set used for optimization when training a model. The model is trained using a training set and in general, performance during training is tested with a validation set. This is useful for generalization (so that the model is able to recognize patterns generically) [7].

2.2 Data Labeling
Image labelling is an important step in this research. The purpose of image labelling is to label the image according to the fish species so that it can be recognized in the training process. Image labelling using labelling software. The image labelling process begins by selecting the YOLO format and then making a ground truth box. The box is made to mark the object to be identified. After making the box, the next step is to label it according to the species name of the object in the picture. The label names used for the training process are \textit{Thunnus_albacares}, \textit{Euthynnus_affinis}, and \textit{Katsuwonus_pelamis}. The result of the data from the labelling process is in the form of a *.txt file that can be used for the training process.

2.3 Training
Training or learning process is a stage where the YOLOv5 model [8] is trained to produce high accuracy values. The training process is carried out to recognize objects by training the models that have been created. There are two stages in the training process, namely feedforward and backpropagation. The feedforward process begins by preparing the number and size of the layer to be formed, the size of the subsampling, the vector image. The feedforward process is the first stage in the training process that produces a weighted value. This value is used to evaluate the YOLOv5 model. The feedforward process will produce several layers to classify image data using the updated weights and biases from the backpropagation process. This stage will also be reused during the testing process. The backpropagation process is the second stage of the training process. The results of the feedforward process will then trace the error from the first layer. A sign that the data has been traced will result in a new weight and bias value [9].
2.4 Validation test

The next process is to test the validation of the training that has been done. The validation of the training results is carried out to determine the ability of the training results to recognize the dataset. The training test process is carried out using weights or bias weight values obtained from the results of the training that has been carried out. All images included in the test set will be fed forward with the weight obtained from the training process. The input data used in the form of images containing 3 species of fish *T. albacares*, *K. pelamis*, and *E. affinis* used in this study. Furthermore, the output produced in this process is an image with prediction results in the form of a box and description of the label and confidence score [10]. The next stage after automatic identification is to check the results obtained to ensure the identification results are in accordance with the actual fish species.

2.5 Model evaluation

The evaluation of the YOLOv5 deep learning model can use a confusion matrix. The confusion matrix is used to measure the performance of the YOLOv5 model which has previously been trained. The confusion matrix is a matrix that displays the predictions of the actual classification and classification [11]. The confusion matrix is used as a reference for the model to perform well or not. Calculation of model performance can be done by calculating the values of accuracy, precision, recall, and F1-Score.

Recall and precision are a ratio of predictions made to the positive class, which serves as a measure of how precise and complete the classification has been. The recall is the number of correct predictions in the positive class divided by the number of true positives in the study. Precision is the number of correct predictions in the positive class divided by the number classified as positive. Accuracy is the ratio of correct predictions (positive and negative) to the overall data. Accuracy can answer correct predictions from all available data [12]. F1-Score is the harmonic mean of precision and recall [13]. Precision, recall, and F1-Score values need to be known to see the tendency of the model to make predictions [14]. The formulas used to calculate recall, precision, accuracy, F1-Score are:

\[
Recall = \frac{TP}{TP+FN} \quad (1)
\]

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

\[
Accuracy = \frac{TP}{TP+FP+FN} \quad (3)
\]

\[
F1-Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)
\]

3. Results and Discussion

3.1 Dataset

The dataset used for training in this study were 550 images consisting of 3 species of fish *T. albacares*, *K. pelamis*, and *E. affinis*. The number of images for each fish species can be seen in Table 1.

| Species       | Images | Number of Images |
|---------------|--------|-----------------|
| *T. albacares* | ![Image](image1.png) | 188             |
| *K. pelamis*   | ![Image](image2.png) | 202             |
| *E. affinis*   | ![Image](image3.png) | 160             |
The dataset is divided into training set and validation set. The dataset used in the Train set process is 440 images while the dataset used for the validation set process uses 110 images.

3.2 Training result

The training or learning process in this study was carried out to obtain the value of the weight used for the identification of yellowfin tuna, mackerel, and skipjack tuna. The learning parameters used in this study are a learning rate of 0.01, momentum of 0.937, image size of 416 pixels, batch 16, and the number of epochs 200, 300, and 400. One of the parameters used for the learning process is epochs. Epoch is when the entire dataset has gone through the training process on the ANN until it is returned to the beginning in one round [15]. One epoch used for the training process means the machine learning algorithm has 'learn' from the training data as a whole. The learning process is carried out repeatedly depending on the number of epochs used to achieve the convergence of the weighted values [16].

The appropriate epoch value for the learning/training process cannot be known with certainty. Several experiments are needed so that the epoch value can produce optimum accuracy. In this study, the epoch values were 200, 300, and 400 times. The use of a varying number of epochs is carried out to determine the optimum accuracy value that can be used to detect objects. The number of epochs that will be used also affects the time during the training process. Due to the increasing number of epochs, the time that will be needed for the learning process will be longer.

The learning process in the YOLOv5 model produces a training loss graph that shows the process of improving the error value during the dataset learning process. Training loss is commonly used to maximize the performance of machine learning algorithms and is based on how well the model is created in the training and model validation stages. Training loss can indicate how bad or how good a model is after iterations of model optimization [17]. The training loss values generated during the learning process for the number of epochs of 200, 300, and 400 can be seen in Table 2.

| Epoch | Training loss |
|-------|---------------|
| 200   | 0.000372      |
| 300   | 0.000291      |
| 400   | 0.000253      |

Table 2 shows the value of training loss from the learning process using the number of epochs 200, 300, and 400. Epochs with a number of 200 produce a higher error value than the number of epochs 300 and 400. The value of training loss generated when the number of epochs is 200 is 0.000372 and the value of training loss when the number of epochs is 300 is 0.000291. The number of epochs that produce the lowest training loss is epoch 400 with the resulting value of 0.000291. Based on the results of the training loss from each epoch, it can be seen that the value of the training loss decreased with the increase in the number of epochs. The larger the epoch used, the lower the loss value generated in the training data. Based on this, it can be concluded that to reduce the loss value obtained and it can be done by increasing the number of epochs in the training process.

3.3 Validation Test

The validation test process or training results test is carried out to check the accuracy and precision of the model that has been carried out in the learning process. The validation test is carried out using the weight values obtained from the learning process. When performing validation tests, the dataset only performs a feedforward process. The dataset used for the validation test process uses 133 images. The distribution of image compositions for T. albacares, K. pelamis, E. affinis amounted to 36, 55, and 42 images, respectively. In binary classification, data is separated into positive (object) and negative (non-object). The results of the validation test must measure the level of accuracy of the model in detecting objects. The method used to measure the accuracy of the model can use the confusion matrix. The parts that will be calculated to determine the level of model accuracy are precision, recall, accuracy, and F1-Score. Accuracy is the ratio of correct predictions of the entire data. Recall or sensitivity is the proportion of true positive predictions compared to the overall data.
that are true positive. Precision is the proportion of positive correct predictions compared to the overall positive predicted results [18].

Evaluation of the model with the confusion matrix uses the results of the YOLOv5 model from the training process that has been carried out. The epochs used for the learning process are 200, 300, and 400 times. The results of the evaluation of the confusion matrix for each epoch can be seen in Tables 3, 4, and 5.

**Table 3.** The results of the evaluation of the confusion matrix during the learning process use the number of epochs 200.

|     | Predicted          |          |          |          |
|-----|--------------------|----------|----------|----------|
|     | *K. pelamis*       | *T. albacares* | *E. affinis* | *Not identified* |
| Actual | 188               | 6        | 2        | 9        |
|       | *T. albacares*   | 6        | 154      | 0        | 6        |
|       | *E. affinis*     | 1        | 1        | 42       | 2        |

**Table 4.** The results of the evaluation of the confusion matrix during the learning process use the number of epochs 300.

|     | Predicted          |          |          |          |
|-----|--------------------|----------|----------|----------|
|     | *K. pelamis*       | *T. albacares* | *E. affinis* | *Not identified* |
| Actual | 198               | 0        | 0        | 7        |
|       | *T. albacares*   | 10       | 154      | 0        | 2        |
|       | *E. affinis*     | 2        | 0        | 42       | 2        |

**Table 5.** The results of the evaluation of the confusion matrix during the learning process use the number of epochs 400.

|     | Predicted          |          |          |          |
|-----|--------------------|----------|----------|----------|
|     | *K. pelamis*       | *T. albacares* | *E. affinis* | *Not identified* |
| Actual | 200               | 0        | 0        | 5        |
|       | *T. albacares*   | 9        | 154      | 0        | 3        |
|       | *E. affinis*     | 0        | 2        | 42       | 2        |

The data in Tables 3, 4, and 5 show the results of the evaluation of the confusion matrix from the weight results when using the number of epochs of 200, 300, and 400 when the learning process is carried out. The results of the confusion matrix evaluation for each epoch are quite diverse. Objects are correctly identified increase with the increase in the number of epochs made during the learning process. Especially in the species *K. pelamis*, the objects identified were correct for the number of epochs of 200, 300, and 400, namely 188, 198, and 200 objects. Meanwhile, the objects were correctly identified for the species *T. albacares* and *E. affinis* resulted in fixed object detection for each epoch.

Objects that are not identified and identified incorrectly in the evaluation results of the confusion matrix decrease with the increase in the number of epochs. Objects that are not identified and identified are incorrect when epoch 200 gets a higher value than the number of other epochs. The number of objects that were not identified and identified incorrectly during epoch 200 were 17 and 16 objects. When the learning process uses the number of epochs of 400, objects that are not identified and identified are incorrectly decreased compared to the number of previous epochs. The number of objects that were not identified and predicted incorrectly during epoch 400 were 10 and 11 objects.

Based on the results of the confusion matrix evaluation from Tables 3, 4, and 5, the number of correctly identified objects has increased while the number of incorrectly identified and unidentified objects has decreased. When the number of epochs increases, the resulting loss value will be lower which causes an increase in the accuracy value for the YOLOv5 model. The higher the accuracy
value, the more accurate the identification of the displayed image will be. However, sometimes a large number of epochs does not guarantee that the accuracy obtained will be high, it is influenced by the large number of datasets used for research [19]. The performance of the YOLOv5 model can be seen by calculating the values of accuracy, precision, recall, and f1 score. The calculation results from the evaluation of the confusion matrix for each epoch can be seen in Tables 6, 7, and 8.

**Table 6.** The calculation results from the confusion matrix during the learning process use the number of epochs 200.

| Label Name             | Precision | Recall | F1 Score | Accuracy |
|------------------------|-----------|--------|----------|----------|
| *Katsuwonus_pelamis*   | 0.964     | 0.917  | 0.940    | 0.921    |
| *Thunnus_albacares*    | 0.957     | 0.928  | 0.942    | 0.921    |
| *Euthynnus_affinis*    | 0.955     | 0.913  | 0.933    | 0.921    |
| Average                | 0.958     | 0.919  | 0.938    |          |

**Table 7.** The calculation results from the confusion matrix during the learning process use the number of epochs 300.

| Label Name             | Precision | Recall | F1 Score | Accuracy |
|------------------------|-----------|--------|----------|----------|
| *Katsuwonus_pelamis*   | 0.943     | 0.966  | 0.954    | 0.945    |
| *Thunnus_albacares*    | 1         | 0.928  | 0.963    | 0.945    |
| *Euthynnus_affinis*    | 1         | 0.913  | 0.955    | 0.945    |
| Average                | 0.981     | 0.936  | 0.958    |          |

**Table 8.** The calculation results from the confusion matrix during the learning process use the number of epochs 400.

| Label Name             | Precision | Recall | F1 Score | Accuracy |
|------------------------|-----------|--------|----------|----------|
| *Katsuwonus_pelamis*   | 0.957     | 0.976  | 0.966    | 0.95     |
| *Thunnus_albacares*    | 0.987     | 0.928  | 0.957    | 0.95     |
| *Euthynnus_affinis*    | 1         | 0.913  | 0.955    | 0.95     |
| Average                | 0.981     | 0.939  | 0.96     |          |

Based on Tables 6, 7, and 8 the accuracy values generated when the model carried out the learning process with a number of epochs of 200 got the lowest accuracy value, namely 0.921 or 92.1%. The highest accuracy value of 0.95 or 95% was obtained when the model carried out the learning process with the number of epochs of 400. The results of the accuracy for the number of epochs of 300 were 0.945 or 94.5%. Accuracy describes how accurately the model can classify objects correctly. In the field of machine learning, especially in classification, the model is said to have high accuracy if it can predict the output of a number of inputs correctly [14]. Study to detect reef fish of the Chaetodontidae family using the deep learning method (YOLOv1) obtained 85.87% accuracy [15].

The F1 scores for each of the epochs of 200, 300, and 400 are 93.8%, 95.8%, and 96%. The value produced by the F1 score continues to increase with the increase in the number of epochs used during the learning process. The precision value generated when using the number of epochs 200 is 95.8%, while the precision value for the number of epochs 300 and 400 is the same value, which is 98.1%. The ability of the model to distinguish objects of the same species can be seen in the precision value. The highest precision value is obtained when the dataset is trained with 400 epochs, which is 98.1%. The recall values generated for the use of the epochs of 200, 300, and 400 are 91.1%, 93.6%, and 93.9%. The ability of the model to distinguish objects with different species can be seen in the recall value obtained. The highest recall value was obtained when the dataset was trained with a total of 400 epochs, which was 93.9%.

Based on the recall, precision, F1-Score, and accuracy values for each epoch, the number of epochs that produces the optimum accuracy value for use in the identification of yellowfin, mackerel, and skipjack tuna is the number of epochs of 400. The results of the calculation of recall, precision,
F1-Score, and accuracy is obtained from 90%. It can be concluded that the YOLOv5 model can identify yellowfin tuna, mackerel, and skipjack tuna as well. An example of identification using the YOLOv5 model can be seen in Figure 1.

![Identification Results of K. pelamis, E. affinis, and T. albacares.](image)

**Figure 1.** Identification Results of *K. pelamis*, *E. affinis*, and *T. albacares*.

Figures 1 is example of the detection results of *K. pelamis*, *E. affinis*, and *T. albacares* species during the validation test process. Based on the figure, the model that has been carried out in the learning process can detect species of *K. pelamis*, *T. albacares*, and *E. affinis* well and produce a high confidence score.

**4. Conclusion**

Identification of yellowfin tuna (*T. albacares*), mackerel (*E. affinis*), and skipjack tuna (*K. pelamis*) using the YOLOv5 model can detect these fish well. The number of epochs that are good for use in the YOLOv5 model is 400, it produces a low training loss value. The values for training loss, accuracy, precision, recall and F1-Score when the model is learning with a total of 400 epochs are 0.000253, 95%, 98.1%, 93.9%, and 96%.

**References**

[1] Firdaus M 2018 The profile of tuna and cakalang fishery in Indonesia *Buletin Ilmiah Marina Sosial Ekonomi Kelautan dan Perikanan* 4(1) 23–32 (in Bahasa Indonesia)

[2] Sibagariang O P, Fauziyah and Agustriani F 2011 Analisis potensi lestari sumberdaya perikanan tuna longline di Kabupaten Cilacap, Jawa Tengah *Maspasi Journal* 03 24–29

[3] Prasmatio R M, Rahmat B and Yuniar I 2020 Deteksi dan pengenalan ikan menggunakan algoritma Convolutional Neural Network *JIFoSI* 1(2) 510–521

[4] Deng L and Yu D 2013 *Deep Learning: Methods and Applications* vol 7 ed Y Eldar et al. (Hanover: Foundations and Trends in Signal Processing) pp 197–387

[5] Ilahiyah S and Nilogiri A 2018 Implementasi Deep Learning pada identifikasi jenis tumbuhan berdasarkan citra daun menggunakan Convolutional Neural Network *JUSTINDO* 3(2) 49–56

[6] Nugroho P A, Fenriana I and Arijanto R 2020 Implementasi Deep Learning menggunakan Convolutional Neural Network (CNN) pada ekspresi manusia *Jurnal Algor* 2(1) 12–21

[7] Putra J W G 2020 Pengenalan konsep pembelajaran mesin dan deep learning editions 1.4 Dwnloaded 5 April 2021 [Internet]. https://wireagotama.github.io/resources/ebook/intro-to-ml-secured.pdf

[8] YOLOv5 2021 https://github.com/ultralytics/yolov5
[9] Putra W S E, Wijaya Yudhi A and Soelaiman R 2016 Klasifikasi citra menggunakan Convolutional Neural Network (Cnn) pada Caltech 101 Journal Teknik ITS 5(1) 65–69
[10] Santoso S A 2020 Design and Implementation of Chaetodontidae Fish Identification Algorithm with Deep Learning Method Bachelor Thesis IPB University, Bogor
[11] Mahardhika A A, Saptono R and Anggrainingsih R 2016 Sistem klasifikasi feedback pelanggan dan rekomendasi solusi atas keluhan di UPT Puskom UNS dengan algoritma Naïve Bayes Classifier dan Cosine Similarity Journal ITSMART 4(1) 36-42
[12] Anggraini W 2019 Deep Learning for Face Detection Using Hijab Using the Convolutional Neural Network (CNN) Algorithm with Tensorflow Bachelor Thesis Universitas Islam Negeri Ar-Raniry, Banda Aceh
[13] Hidayatullah A F, Fadila A A, Juwairi K P and Nayoan R A 2019 Identifikasi konten kasar pada tweet Bahasa Indonesia Jurnal Linguistik Komputasional 2(1) 1–5
[14] Natan O, Gunawan A I and Dewantara B S B 2019 Gird SVM: aplikasi Machine Learning dalam pengolahan data akuakultur Jurnal Rekayasa Elektrika 15(36) 7–17
[15] Thohari A and Hertantyo G B 2018 Implementasi Convolutional Neural Network untuk klasifikasi pembalap MotoGP berbasis GPU Proc. National Conference on Electrical Engineering, Telematics, Industrial Technology, and Creative Media 2018 (Purwokerto, 11 Agustus; 2018) (Purwokerto: Institut Teknologi Telkom Purwokerto) pp 50–55
[16] Wibawa M S 2016 Pengaruh fungsi aktivasi, optimasi dan jumlah epoch terhadap performa jaringan saraf tiruan Jurnal Sistem dan Informatika 11(2) 167–174
[17] Bariyah T, Rasyidi M A and Ngatini N 2021 Convolutional Neural Network untuk metode klasifikasi multi-label pada motif batik Techno Com. 20(1) 155–165
[18] Putri O N 2020 Implementation of the Cnn Method in Mushroom Image Classification in Image Processing Analysis Bachelor Thesis Universitas Islam Indonesia, Yogyakarta
[19] Yusup I M 2020 Automatic Recognition of Chaetodontidae Family Reef Fish Using Deep Learning Bachelor Thesis IPB University, Bogor