XLIII CILAMCE
21-25, November, 2022
FOZ DO IGUAÇU – PR – BRAZIL

Employment of Artificial Intelligence in Fish Fraud Identification

Edyene Cely Amaro Oliveira[1], Carol Conceição Haddad, Ana Júlia Kavalco, Camila Chaves Mariano, Camila Imulia Porto Braz, Daniel Teófilo da Silva, Dayvidson Henrique Moreira da Silva, Eduardo Lucas Ramos da Silva, Gabriel Alexandre do Amaral, Herbert Lucas Gomes da Silva, João Vitor Clavilho dos Santos, Lucas da Rocha Silva, Matheus Henrique Marcelino de Oliveira

[1]*Dept. Centro Universitário Una campus Aimorés*
Rua dos Aimorés, 1451 - Lourdes, Belo Horizonte - MG, 30140-071
edyene.oliveira@prof.una.br

Abstract. Acquiring fish for consumption whether in restaurants, supermarkets or fishmongers can become a frustrating task from the moment it is perceived that the species at hand is not the expected fish. Added to this is the fact that fraud always aims to deceive to make a profit. There are some applications that aim to identify fish species; however, they have no specific focus on certain species. With this, the user applies to all species of fish and ends up disappointed with the inefficiency of the application. This project aims to implement an application that uses artificial intelligence to identify certain species of fish. The main point of the application are fish species and cannot be used for ornamental fish. The method used was to obtain images and configuration of models of artificial neural networks for image recognition. During the tests the implemented model reached 75% accuracy. The model had difficulty recognizing three species accurately, it was the case of Merluza fish that was confused with Salmon and Trout. It is believed that this occurred due to the low number of samples, but in short, the model achieved satisfactory results.

Keywords: Artificial Intelligence, Fraud, Fish, Neural Network

1. Introduction

Archaeological studies indicate that man has been using fish as food for thousands of years. Evidence found by scholars points to almost 2 million years [1]. Today, the amount of fish consumed varies worldwide, in Brazil, for example, the annual consumption per person is 10.5 kg per year, in Japan this value reaches 23.4 kg, while the world average is 20.5 kg/person/year in 2018 [2].

Although the fish is present on the table of people all over the world, there is concern at the time of purchase. This is due to the practice of fish fraud. Fraud has several meanings; however, this study will focus on replacing one species with another, usually with greater availability and lower commercial value. The identification of this fraud is complex since some fish have a very similar appearance. Thus, even for those who have a certain knowledge, identification is not trivial [3].

In the literature review, some projects were found in order to identify fish, such as: [4]. In Brazil, there is supervision carried out by the Ministry of Agriculture, Livestock and Supply (MAPA). The process consists of collecting samples of domestic and imported fish and in the laboratory carry out DNA tests to verify that the genetic information of the samples conforms to the species described on the label.

The aim of this project is to implement an application that uses artificial intelligence to recognize certain species of fish from the image (photography). In this way this project aims to reach corporate and final consumers. With this, from the *smartphone, the buyer* will be able to photograph fish in supermarkets and sales places to identify the species indicated by the seller. The use of the application will have relevance for the consumer who will have confidence that he is paying fair price for the fish purchased and for the state, because it will minimize the spread of fraud. In addition, with the increase in the database certain establishments and entrepreneurs will be able to use it to extract insights for their business.

In financial terms, the application will help the end consumer to buy a fish that he seeks generating savings and satisfaction, since defrauding is always with the aim of taking advantage. Fish traders across the country will be able to use the app to also not be fooled into acquiring lots of fish, as such people mostly buy directly from the fisherman. Restaurants are segments that will also benefit from the application, because they use large amounts of fish to be used in the menus daily.

The general objective of the work is to design and implement an application that allows the identification of fish, in principle, of the species of Merluza, Tilapia, Trout and Salmon fresh or frozen, whole or fillets.

As secondary objectives can be highlighted so far:

1. Generation of a database of fish images with as many photos as possible with the 4 different species of fish (Tilapia, Merluza, Trout and Salmon).
2. Design and implementation of preprocessing algorithms for identification, organization, and identification of images suitable for the Machine learning algorithm.
3. Implementation or identification of an algorithm for embedding generation of images to be used in the artificial neural network.
4. Implementation and identification of a neural network algorithm suitable for image classification
5. Implementation of an application with attractive and friendly interface to receive the photo of the fish sent by the user and classify the species using artificial intelligence.
1. Implementation of a software with attractive and user-friendly interface to receive the phrase and / or key words typed by the user and perform the search of jurisprudence using artificial intelligence.

2. Methodology

The project began with the acquisition of images of the fish: Tilapia, Merluza, Salmon and Trout. These images were obtained from Google search engines. The purpose of the images is specifically for model training and is not available to third parties. The images extracted are of whole fish and fillets.

Pre-processing was performed in the set of images to select those that did not contain data that could impair the identification by the model. Thus, images that had for example, price, stamp, or logo of some supermarket were discarded. The name of each image has been changed to, for example: Tilapia.1; Tilapia.2; Merluza.1; Merluza.2; Merluza.100.

In addition, the images were stored in folders on the disk with their names indicative of the species. Example of data can be seen in Figures 1, 2, 3, and 4.

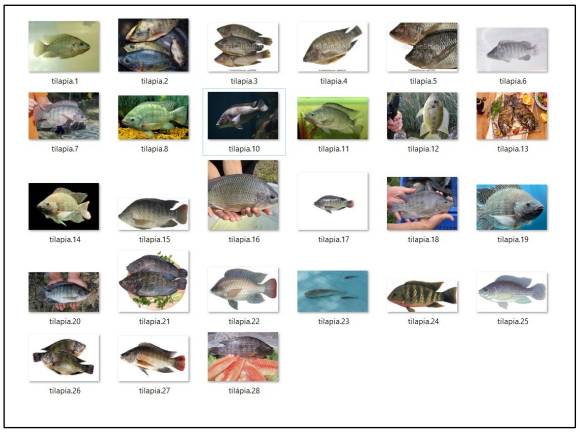


Figure 1 - Tilapia
Source: Public Google Images

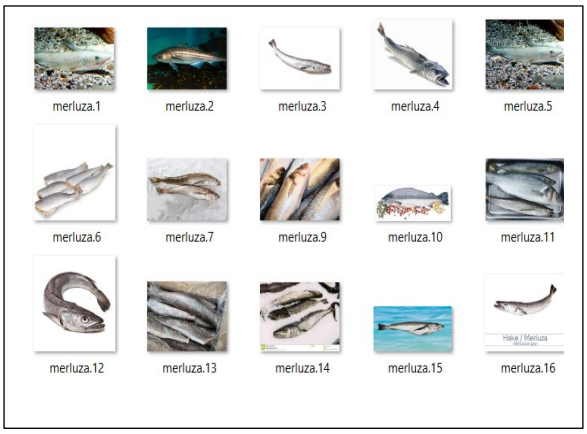


Figure 2 - Merluza
Source: Public Google Images

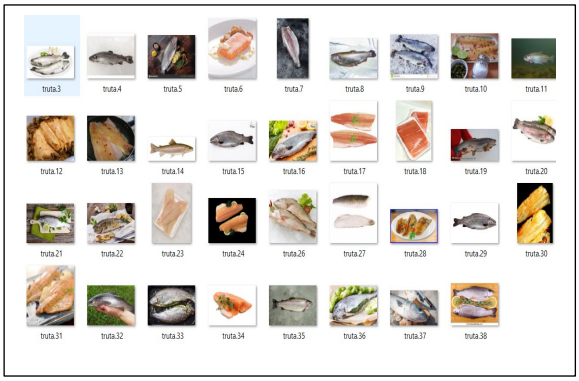


Figure 3 - Trout
Source: Google Images

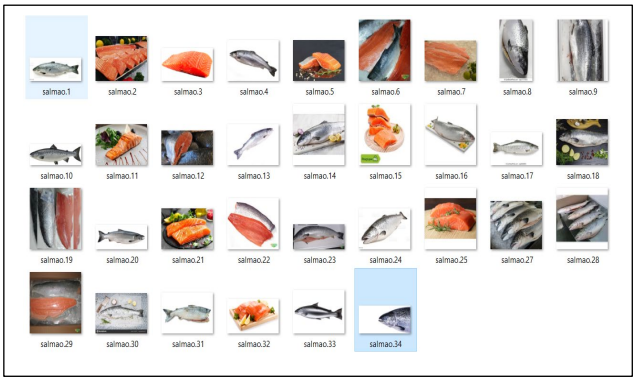


Figure 4 - Salmon
Source: Google Images

The machine learning algorithm used was the CNN Artificial Neural Network – *Convolution Neural Network*. This neural network aims to receive an image and perform classification according to historical data. Figure 5 shows the identification of fish by a consumer.

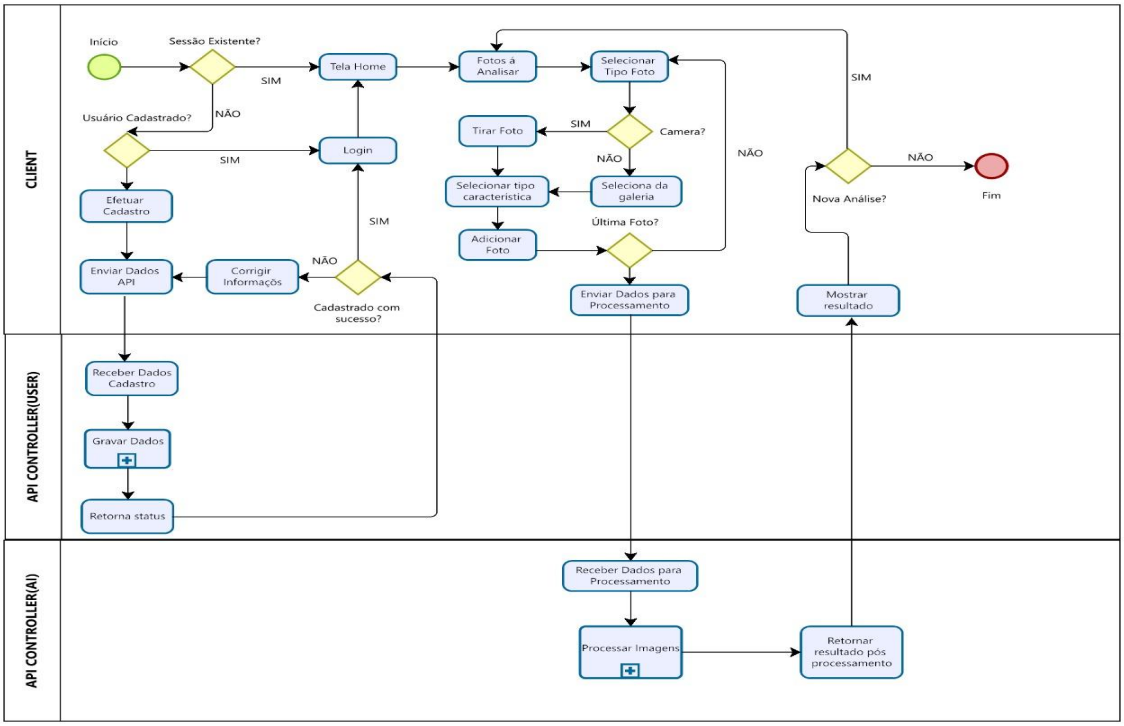


Figure 5 - Fish identification process to be carried out by the user
Source: the authors

3. Results

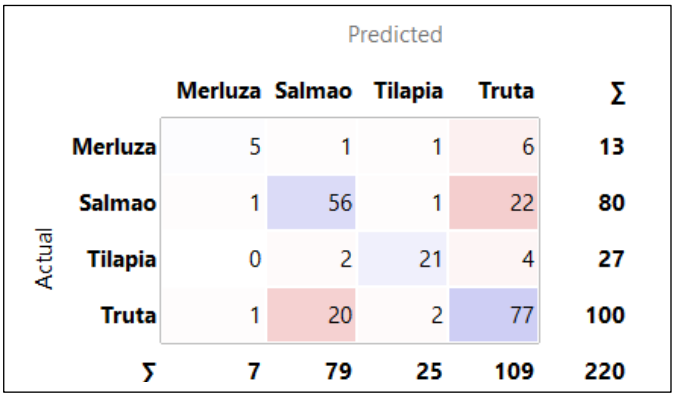
3.1 Model configuration

The hyper parameters used for neural network configuration were: first 2D convolutional layer with 32 neurons and RGB color layers. The second layer was applied MaxPooling 2D with 2x2 matrix and in the third the input vector was generated for the dense neural network of classification. The fourth and fifth layers were dense MLP being the fourth with 128 neurons and the last with 4 outputs. The activation function used was *Relu and* softmax. The learning rate was: 0.0001 and the optimization algorithm was Adam.

3.2 Results and comments

Were performed tests applying validation with manual separation with cutting of 80% for training and 20% for testing. Cross-validation was performed using 5 and 10 folds. In the images below can be seen the results obtained.

a) 5 Folds



Model	AUC	CA	F1	Precision	Recall
Neural Network	0.869	0.723	0.720	0.724	0.723

Figure 6 - Confusion matrix and results of metrics: AUC, Accuracy, F1, Precision and Recall using 5 folds.

The results showed that the system works, however, it is necessary to obtain a greater number of images. In Figure 6 of the 220 images used for testing, the model correctly classified 159, i.e., 72%. The greatest number of errors in relation to the correct answers was with the Merluza fish. Of the 13 samples the model missed 8.

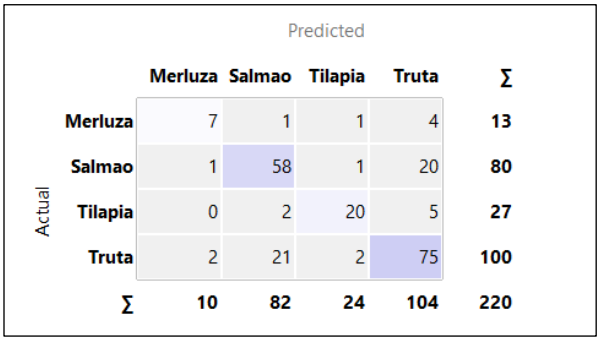
The second highest ranking was salmon fish with 24 wrong classifications. Trout came in third in relation to errors with 23 fish with wrong ratings.

In short, the model using cross-validation with 5 folds confused the fish: Trout with Merluza and salmon. In the confusion matrix, it is perceived that of the 13 Merluzas, 6 the model classified as Trout. 20 Trout were erroneously classified as Being Salmão and at the same time, 22 Salmon fish were classified as Trout.

In the case of Merluzas/Trout the percentage of misconception was 46%, while Trout/Salmon was 20% and Salmon/Trout was 27.5%.

The results metrics point to 72% of correct answers of the model, which is the total calculated. The AUC metric showed a result of 87% for this model. AUC is invariant in scale. It values how well predictions are sorted, rather than their absolute values. AUC is threshold-invariant classification. It values the quality of model predictions, regardless of the classification limit chosen.

b) 10 Folds



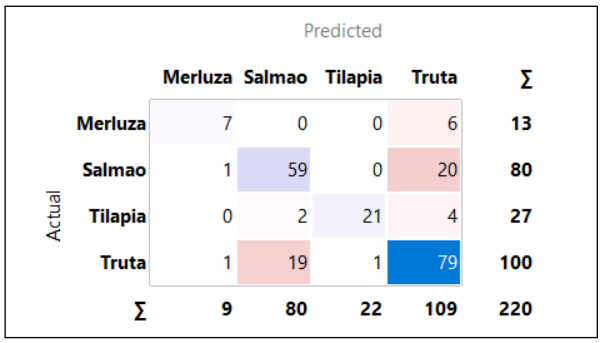
Model	AUC	CA	F1	Precision	Recall
Neural Network	0.875	0.727	0.727	0.729	0.727

Figure 7 - Confusion matrix and results of metrics: AUC, Accuracy, F1, Precision and Recall using 10 folds.

By applying cross-validation using 10 folds, the model improved slightly over errors. From the total of 220 samples, the model correctly classified 160, remaining at 72% as shown in the accuracy graph. Some differences in relation to the previous model of 5 folds were the amount of Merluza errors in relation to trout. In this case the model misclassified only 4. In the Salmon/Trout ratio the model classified 20 erroneously, while in the Trout/Salmon comparison the model missed 21.

The AUC metric also indicated a small improvement in the model, bringing it to 87.5%.

1. 80% data for training and 20% for testing.



Model	AUC	CA	F1	Precision	Recall
Neural Network	0.877	0.755	0.755	0.761	0.755

Figure 8 - Confusion matrix and metric result: AUC, Accuracy, F1, Precision and Recall using manual separation.

In the validation using the manual separation 80% for training and 20% for tests it is noted that the model hit 166 of the 220 samples. This represents 75% of correct ratings, as can be seen in accuracy, F1 score, accuracy, and recall metrics. The AUC metric of the model reached 88% indicating that the model is likely to correctly classify a sample.

4. Conclusions

This fish recognition project using artificial intelligence presented satisfactory results in this first test. Analyzing the results, it is perceived that the model is capable of recognizing fish. Images of whole fish, fillets, fresh and frozen were used. At this stage the images did not go through any treatment to improve the resolution, besides having made no choice of images without details. The set of images was used in the same way they were obtained, that is, there were images with spices and marks.

The images totaled 220 in total, 13 of the species Merluza, 80 images of Salmon, 27 of Tilapia and 100 of Trout. The correct ratings in all configurations were around 72%. It is necessary to obtain a greater number of images to continue the project. In addition to the RNA model, it is necessary to present more samples of each species. In addition, it is important for images to go through a process of selection and resolution treatment.

In the tests it was noticed that the model confused the species Merluza with Salmon and Trout. It may be that by performing improvement in the resolution and selection of images without details the model can achieve greater accuracy. In addition to quantity, these models need a lot of sample data to learn how to differentiate each image. In addition, other algorithms will be tested for optimal model definition.

References

[1] A. Gibbons, “The World’s First Fish Supper”, 2010. https://www.science.org/content/article/worlds-first-fish-supper (acessado maio 26, 2022).

[2] SeaFood, “Especial Sofia 2020: Consumo mundial chega a 20,5 kg per capita/ano”, *Sea Food Brasil*, 2020. https://www.seafoodbrasil.com.br/especial-sofia-2020-consumo-mundial-chega-a-205-kg-per-capita-ano.

[3] L. V. T. Neto, Danilo Alves Pimenta, Denise Aparecida Andrade de Oliveira, “A Fraude em Pescado e o Método de DNA Barcode Para Identificação de Espécies”, Belo Horizonte, 2017.

[4] J. Guay, “O aplicativo ‘Fishazam’ pode acabar com a fraude de pescado identificando filés”, 2017. https://apolitical.co/solution-articles/pt/fishazam-app-end-fish-fraud-identificando-filetes.