


Article

Automated Identification of Morphological Characteristics of Three *Thunnus* Species Based on Different Machine Learning Algorithms

Liguo Ou ¹, Bilin Liu ^{1,2,3,4,*}, Xinjun Chen ^{1,2,3,4} , Qi He ^{5,*}, Weiguo Qian ^{6,*} and Leilei Zou ⁷¹ College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China² The Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China³ National Distant-Water Fisheries Engineering Research Center, Shanghai Ocean University, Shanghai 201306, China⁴ Key Laboratory of Oceanic Fisheries Exploration, Ministry of Agriculture and Rural Affairs, Shanghai 201306, China⁵ College of Information Technology, Shanghai Ocean University, Shanghai 201306, China⁶ School of Fishery, Zhejiang Ocean University, Zhoushan 316022, China⁷ School of Foreign Languages, Shanghai Ocean University, Shanghai 201306, China

* Correspondence: bl-liu@shou.edu.cn (B.L.); qihe@shou.edu.cn (Q.H.); qianweiguo@zjou.edu.cn (W.Q.)

Abstract: Tuna are economically important fish species. The automated identification of tuna species is of importance in fishery production and resource assessment in that it would facilitate the informed monitoring of tuna fishing vessels and the establishment of electronic observer systems. As morphological characteristics are important for tuna identification, this study aims to verify the performance of the automated identification of three *Thunnus* species through morphological characteristics based on different machine learning algorithms. Firstly, morphological outlines were visually analyzed using EFT (elliptic Fourier transform) and CNN (convolutional neural network). Then, the EFT feature data and deep feature data of the tuna outline images were extracted, and principal component analysis of the two different morphological characteristics was performed. Finally, different machine learning algorithms were used to analyze the identification performance of tuna of the same genus and different species. The experimental results showed that EFT features had the highest identification accuracy in KNN (K-nearest neighbor), with 90% for *T. obesus*, 90% for *T. albacores*, and 85% for *T. alalunga*. Deep features had the best identification performance in SVM (support vector machine), with 80% for *T. obesus*, 90% for *T. albacores*, and 100% for *T. alalunga*. Deep features were better than EFT features in identification performance. The biodiversity and intergeneric differences among tuna species can be well analyzed using these two different morphological characteristics. Machine learning algorithms open up the way for rapid near-real-time electronic observer systems in these important international fisheries.

Keywords: *Thunnus*; machine learning; morphometrics; convolutional neural network; morphological visualization; automated identification

Key Contribution: Deep features were better than EFT features in identification performance and the biodiversity and intergeneric differences among tuna species can be well analyzed using these two different morphological characteristics. This method will promote the development of tuna fishery automation and provide a feasible strategy for fish biodiversity research.



Citation: Ou, L.; Liu, B.; Chen, X.; He, Q.; Qian, W.; Zou, L. Automated Identification of Morphological Characteristics of Three *Thunnus* Species Based on Different Machine Learning Algorithms. *Fishes* **2023**, *8*, 182. <https://doi.org/10.3390/fishes8040182>

Academic Editor: Susana Franca

Received: 27 February 2023

Revised: 27 March 2023

Accepted: 28 March 2023

Published: 29 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fishery products provide the basis for ensuring the supply of high-quality protein and play a central role in global food and nutrition security [1]. Among them, tuna are important targets for large-scale international industrial marine fisheries around the world [2]. Many

important species of tuna have significant values, both economically and nutritionally, and play a significant role in the economic development of many countries [3]. In addition, tuna are among the most widely consumed seafood in the world and can be processed into a variety of products, including canned fish and a variety of other products [4]. The principal tuna species are *Thunnus obesus*, *T. albacores*, *T. alalunga*, *T. thynnus*, *T. orientalis*, *T. maccoyii*, and *Katsuwonus pelamis* [5].

Species identification for fishery purposes has been the subject of a major Food and Agriculture Organization (FAO) program since the 1960s [6]. Information on the biology of fish is crucial for planning a sustainable management strategy for fishery resources [7]. The documentation of morphological information is important to validate the taxonomy status and kinship relationship within or between species [8]. At the same time, morphometric characteristics are powerful tools for analyzing discreteness and relationships among fish species. For this reason, the analysis of morphometric characteristics has been widely used by ichthyologists to differentiate species and populations within a species [8–10]. Morphometrics is a common and inexpensive method to delineate taxa. Hence, it is a popular technique in fish taxonomic identification [11]. Numerous studies have shown that the analysis of morphometric data is a suitable approach to validate the taxonomic status of fish [12,13]. However, with the rapid development of artificial intelligence, the identification and classification of fish species have undergone great changes. Although conventional fish classification methods still play a very important role, the combination of artificial intelligence technology and fish biology can produce more efficient and automated identification results.

The development of artificial intelligence technology has effected gradual development of fisheries management towards an automation and informed direction across the globe [14]. Machine learning is at the center of artificial intelligence and is a key technology for realizing automated fish identification. Combined with high-performance computers, machine learning technology can mine high-dimensional features and deep information in data, thereby offering a solution to automated fish identification and fishery monitoring, and introducing the fishery industry into a new era [15,16]. Biologists and related scientists have conducted research on fish species classification based on machine learning techniques. Fish species classification is usually achieved by collecting fish images, extracting fish image features, and constructing classification models [17]. Machine learning technologies have been used to classify fish species by extracting the morphology, color, texture, and other features of fish images, and have made some progress [18]. Ogunlana et al. classified fish species using machine learning techniques based on the shape features of the fish [19]. They analyzed the training data of 76 fish (38 *Ethmalosa fimbriata* and 38 *Scomberomorus tritor*) and the testing data of 74 fish (37 fish for each species). SVM achieved an accuracy rate of 78.6%, KNN had 52.7% accuracy, ANN (artificial neural network) had 60% accuracy, and K-means (K-means clustering) had 51% accuracy. Tharwat et al. extracted the texture and color features of fish images, used linear discriminant analysis to reduce the dimensions of feature vectors, and then constructed a classifier algorithm [20]. They collected a dataset that consists of four different fish species, namely *Argyrosomus regius*, *Sardinella maderensis*, *Scomberomorus commerson*, and *Trachinotus ovatus*. Their experimental results revealed that the classifier had an accuracy of approximately 96.4%. Andayani et al. used a combination of GIM (geometric invariant moments), GLCM (gray-level co-occurrence matrix) texture features, and HSV (hue saturation value) color features to select regions of interest and used a probabilistic neural network classifier algorithm to classify fish, and an accuracy rate of 89.7% was obtained [21]. There were three fish species classified in their research, which were skipjack tuna, tongkol, and tuna with “out-of-water” conditions. Traditional machine learning methods require extraction for the recognition and classification of fish image features [22], but the extraction of fish features is affected by the diversity of fish species, the environment of image collection, and the conditions for the implementation of feature algorithms.

Deep learning is capable of overcoming the deficiencies of conventional image classification approaches [23]. Increasing computing power and data size, along with advanced deep learning research, have contributed to the popularity of deep learning [24]. Deep learning is a field of machine learning, and CNN is the most widely applied deep learning method in which the multiple layers are trained and tested in a robust way [25,26]. CNN fuses both feature extraction and classification blocks into a single and compact learning body, which can result in a significant decrease in operational costs associated with human observers [27]. Deep learning technology has good performance for fish classification (Table 1), which is mainly due to the fact that deep learning methods such as CNN do not need any feature extraction upfront. The identification accuracy is directly output by inputting fish images [26].

Table 1. Deep learning methods for performance of fish classification.

Author	Model	Accuracy
Rathi et al. (2017) [28]	CNN	96.3%
Villon et al. (2018) [29]	CNN	94.9%
Rekha et al. (2020) [30]	CNN	92%
Iqbal et al. (2021) [26]	Original AlexNet	87%
Iqbal et al. (2021) [26]	Improved AlexNet	90%

In addition, deep learning has many advantages as a fish feature extraction method. There are relevant research reports using CNN models as tools for the extraction of fish features. CNN extracts fish features as deep data and carries out identification in combination with traditional machine learning. Tamou et al. extracted features from a fish-image dataset using the pretrained AlexNet network and used a linear SVM classifier. The dataset included 23 species of fish. The experiment results achieved an accuracy of 99.5% [31]. Deep et al. used deep learning technology to identify fish, and the classification accuracy of CNN was 99%, that of CNN-SVM was 98%, and that of CNN-KNN was 99% [32]. The Fish4Knowledge dataset (23 fish species) was used in their study.

As the scale of marine fishing continues to expand, more and more attention has been focused on fish biodiversity. In the future, tuna catches are likely to transform into a new sustainable fishing model with automation and intelligence. Therefore, the automated identification of tuna species will help to protect tuna biodiversity, which is significant for improving the economic efficiency of fishing enterprises, enhancing fishery resource assessment and sustainable management, and ensuring sustainable fishery production in the future. In this study, we propose a method for the automated identification of tuna species through morphological characteristics based on different machine learning algorithms. We used elliptical Fourier transform features (morphometric features) and deep features (CNN-obtained morphological outline features) to study the biological diversity of different tuna species of the same genus and analyze the identification performance of machine learning for different morphological features.

2. Materials and Methods

2.1. Materials

In this study, a total of 300 *Thunnus obesus*, *T. albacores*, and *T. alalunga* were used as research objects—100 of each species (Figure 1) that are among the most important commercial tuna species for global tuna fisheries. The digital images of tuna were collected by observers on board. Images were acquired using a digital camera and a smartphone. Surveys were conducted in the Central Ocean and Western Pacific Ocean from March 2021 to September 2022. The tuna were aligned horizontally and centered in the images. The images were processed to give images 400 pixels high × 800 pixels wide; these were then saved in a JPEG file format.

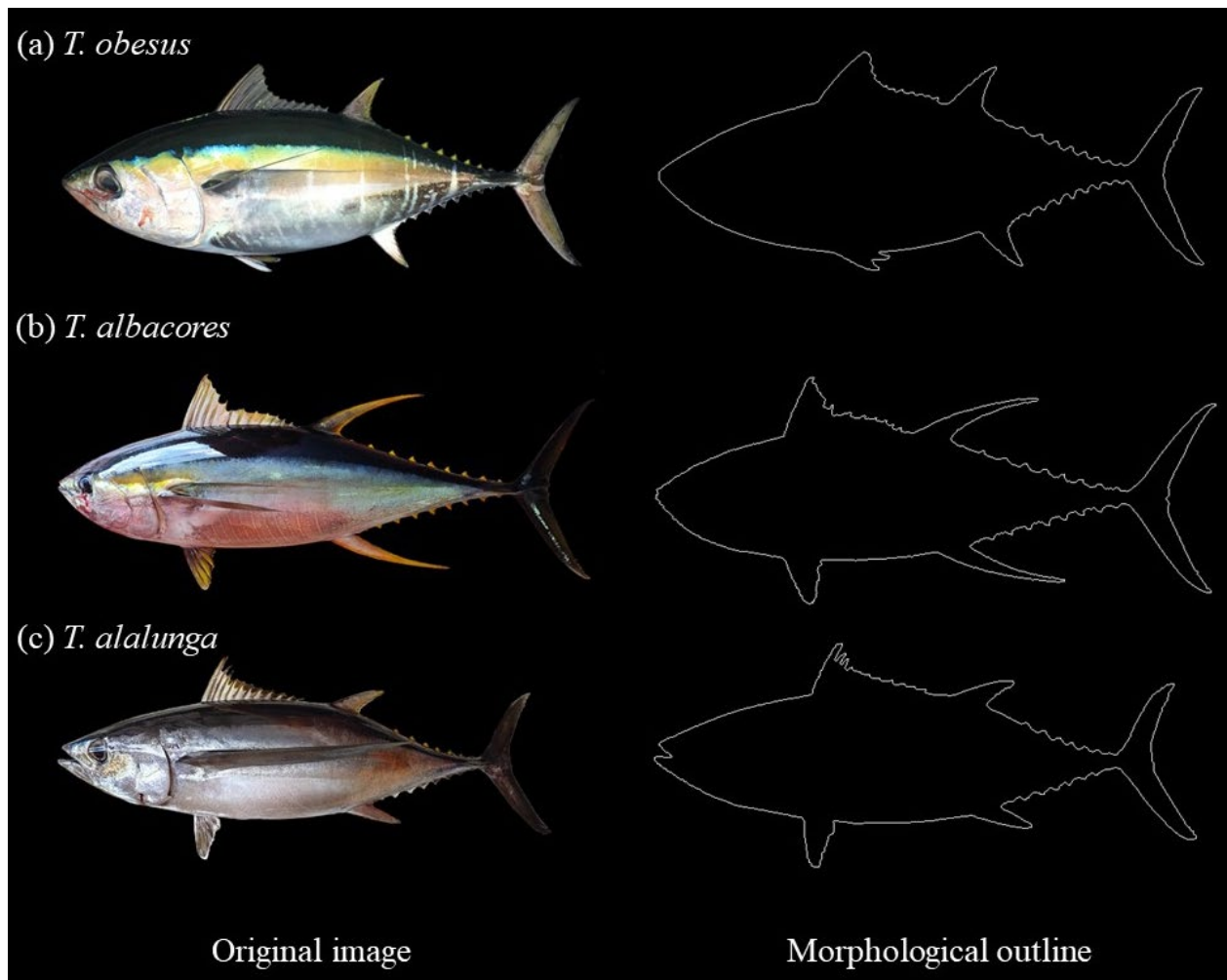


Figure 1. Three *Thunnus* species: the original image is on the left and the morphological outline is on the right.

2.2. Methods

The automated identification of three tuna species consisted of three main steps: (a) Morphological outline images were obtained by preprocessing tuna images; (b) elliptic Fourier transform feature data (EFT feature data) were obtained by using the elliptic Fourier transform, and the deep feature data of the shape contour were obtained by using a convolution neural network; (c) different machine learning algorithms trained two kinds of morphological feature data, obtaining evaluation metrics, ROC (receiver operating characteristic) curves, AUC (area under the curve) values, and the confusion matrix by identifying the test set using the trained model (Figure 2).

2.2.1. Preprocessing of Tuna Images

Images of the three *Thunnus* species were preprocessed using the computer vision library (OpenCV). Image processing techniques such as bilateral filter, gray transformation, image binarization, and contour extraction were applied to obtain a contour image of each tuna (Figure 1). The tuna contour image was processed to obtain two different-sized image sets, with 400 pixels high \times 800 pixels wide and 224 pixels high \times 224 pixels wide.

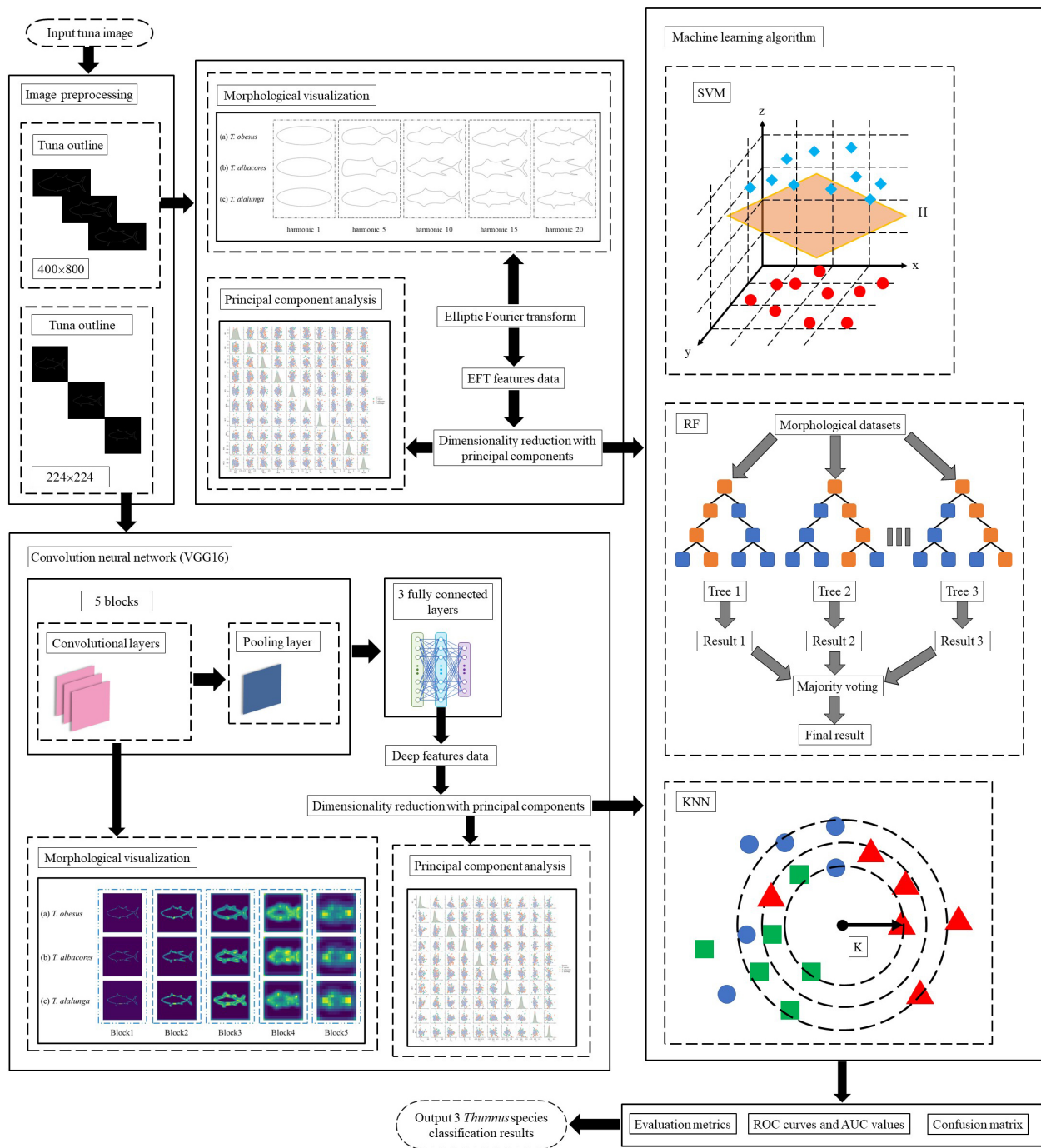


Figure 2. Flowchart of morphological characteristic identification based on machine learning algorithms.

2.2.2. Elliptic Fourier Transform Features and Morphological Reconstruction

An elliptical Fourier function perfectly describes a closed curve with an ordered set of data points in a two-dimensional plane [33]. It uses an orthogonal decomposition of a curve into a sum of harmonically related ellipses. These ellipses can be combined to reconstruct an arbitrary approximation of the closed curve.

The program “pyefd” [34] was used to generate 20 harmonics for the morphological outline of each tuna. It was implemented in Python using “Elliptic Fourier Features of a Closed Contour”. Each harmonic is composed of 4 coefficients resulting in 80 coefficients per tuna outline. Each tuna outline was normalized for size and orientation using this program, which caused the degeneration of the first three coefficients to fixed values: $a_1 = 1$, $b_1 = 1$, and $c_1 = 0$. Therefore, each tuna was represented by 77 coefficients (EFT features)

for the elliptic Fourier transform analysis. The morphological outline of each tuna was reconstructed using “pyefd” with the reconstructed harmonic number from 1 to 20.

2.2.3. Deep Features and Convolution Neural Network Visualization

Deep learning is a field of machine learning, and CNN is the most widely applied deep learning method [26]. In this study, VGG16 (variant of visual geometry group network) [35] was used to extract deep features from tuna outline images. VGG16 has 5 blocks (convolutional layers and pooling layers), and the network structure of VGG16 is composed of 13 convolutional layers, 3 fully connected layers, and soft-max output layers [36]. The convolution kernel of VGGNet is 3×3 , and the ReLU function is used in the activation units of all hidden layers.

Tuna outline images were resized from 400 pixels high \times 800 pixels wide to 224 pixels high \times 224 pixels wide, and outline images were inputted to VGGNet. Visual analysis was performed on the first convolutional layer of each block, and the mean image of the five convolutional layers of VGG16 was outputted. The feature extraction process of VGG16 is a process that proceeds from morphological outlines (biological features) to deep data, that is, the process of the conversion of an outline image into data. Deep features were extracted using VGG16 to extract tuna outline features, and the second layer of the fully connected layer was used as the deep feature data to identify the three tuna species. The deep feature data of each tuna outline image had 4096 features.

2.2.4. Machine Learning Algorithm

Three classification models were used in this study: support vector machine (SVM), random forest (RF), and K-nearest neighbor (KNN). SVM is a widely used machine learning method for classification. It can perform a nonlinear classification using kernel functions, implicitly mapping its inputs into a transformed high-dimensional feature space and finding a hyperplane with the minimum predictive error [37]. The radial basis function (RBF) was used as the kernel function for the parameter optimization of the SVM model. RF associated with multiple decision trees is a brand of ensemble machine learning algorithms. This generalization is commonly used as the error term for a forest convergence when the number of trees in the forest is large, which avoids the overfitting problem [38]. The random forest algorithm has been widely used in various classification problems [16]. KNN is one of the classification algorithms of machine learning. A sample is essentially regarded as most similar to the K-nearest samples in the dataset, and if most of the said k samples belong to a certain category, the sample also belongs to this category [14]. The sample is then classified by measuring the distance between different eigenvalues.

2.2.5. Evaluation Metrics

The performance of each deep convolutional network model presented in this study was measured with precision, recall, and F1-score [39]. Precision is the ratio of true positives to all the positives predicted by the model. Recall is the ratio of true positives to all the positives in the tuna dataset. The F1-score is a composite measure of precision and recall, providing a better indication of the overall performance of the model. These measures are defined as follows:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1\ score = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

Here, true positive (TP) correctly predicts in the positive class, true negative (TN) correctly predicts in the negative class, false positive (FP) incorrectly predicts in the positive class, and false negative (FN) incorrectly predicts in the negative class.

2.2.6. ROC Curves and AUC Values

In this study, receiver operating characteristic (ROC) curves were analyzed to estimate the specificity and sensitivity of the established different convolutional network model for the discrimination of tuna species. In the ROC curve, the x -axis indicates a false-positive rate (FPR), and the y -axis indicates a true-positive rate (TPR). AUC is the area below the ROC curve. The area under the curve (AUC) has a value between 0 and 1 [37,40].

$$TPR = \frac{TP}{TP + FN} \quad (4)$$

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

$$AUC = \frac{1 + TPR - FPR}{2} \quad (6)$$

2.2.7. Confusion Matrix

The confusion matrix is a matrix representation of the prediction results. It was used to describe the performance of the classifier on a set of tuna test data. In this study, the datasets of the three tuna species were drawn in a matrix form based on the criteria of the actual tuna category and the tuna category predicted using the classification. The numbers of TP, TN, FP, and FN obtained for each dataset were used to draw the confusion matrix for each species.

2.2.8. Data Processing

To compare and analyze the three species, we used principal component analysis to reduce the dimensions of two different feature data, and they retained 10 principal components for data analysis. The principal component analysis diagram was drawn. Different morphological feature datasets were divided into an 80% training set and a 20% test set. In addition, the parameter of SVM was set to $C = 60$, $\gamma = 0.5$. RF parameters were set to estimators = 100, criterion = "gini". The parameter of KNN was set to $K = 1$.

The experimental environment included computer processor Intel(R) Core (TM) i7-7700 HQ CPU @ 2.80 GHz; mainboard model LNVNB161216; primary hard drive NVMe SAMSUNG MZVLW128 (119 GB); graphics card Intel (R) HD Graphics 630 (1024 MB) and NVIDIA GeForce GTX 1060 (6144 MB); and Python 3.6.6 for image data processing.

3. Results

3.1. Visualization of Tuna Morphology

We found that the elliptic Fourier transform and deep convolution neural network could effectively characterize tuna morphological information. Tuna morphology was visually reconstructed using the elliptic Fourier transform (Figure 3). The transforms of harmonic 1 were all ellipses. When the harmonic number was five, the morphological reconstruction of the three *Thunnus* species showed preliminary fish species characteristics. When the harmonic number was 20, their morphological reconstruction was basically consistent with the original tuna morphology. Through the first convolution layer of each block (Figure 4), it was found that the deep convolution neural network could directly obtain the tuna morphological information. The morphological specificity of tuna could be observed during the convolution process from block 1 to block 4. At block 5, the tuna morphological data were converted into deep data.

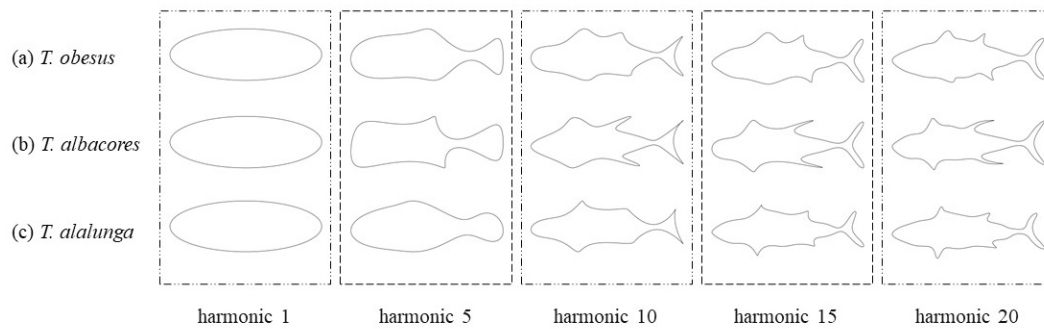


Figure 3. Morphological reconstruction of three *Thunnus* species from the number of Fourier harmonics.

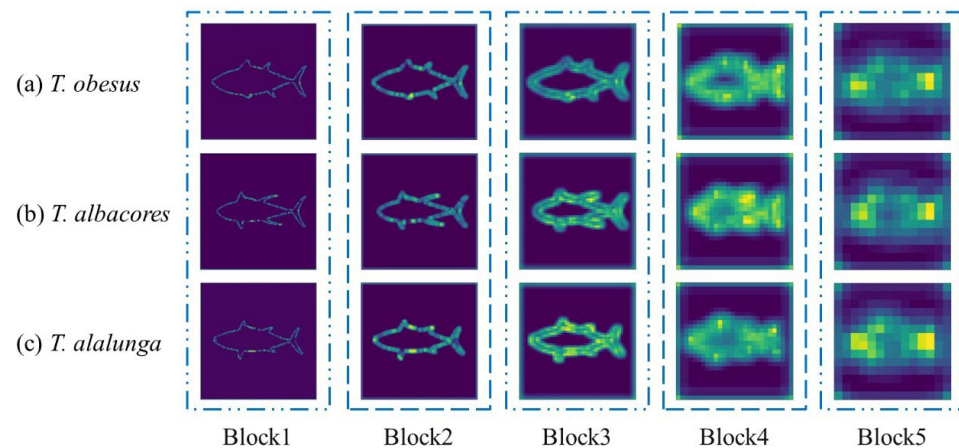


Figure 4. Visualization of three *Thunnus* species with deep convolutional neural networks.

3.2. Principal Component Analysis of *Thunnus* Species

The results of principal component analyses of the two morphological characteristics of tuna showed that they could well extract the morphological information of the three *Thunnus* species. From the first PC (principal component) to the tenth PC, the cumulative contribution rate of EFT feature data was 83%, and the cumulative contribution rate of deep feature data was 82% (Table 2). Among them, the PC1 of the EFT feature data contributed 27%, and deep features contributed 42%. As shown in Figure 5, the principal component analysis of EFT feature data revealed that there were subtle differences among the tuna species between different principal components, and each two principal components could not well distinguish the three tuna species. However, from the single principal component analysis of diagonal lines, it could be seen that there were differences among the three tuna species. The results of the principal component analysis for the deep feature data were similar to those of EFT feature data (Figure 6).

Table 2. Contribution rates of different principal components.

Principal Components	EFT Features	Deep Features
PC1	27%	42%
PC2	16%	15%
PC3	10%	9%
PC4	8%	5%
PC5	5%	3%
PC6	5%	3%
PC7	4%	2%
PC8	3%	1%
PC9	3%	1%
PC10	2%	1%
Cumulative contribution rate	83%	82%

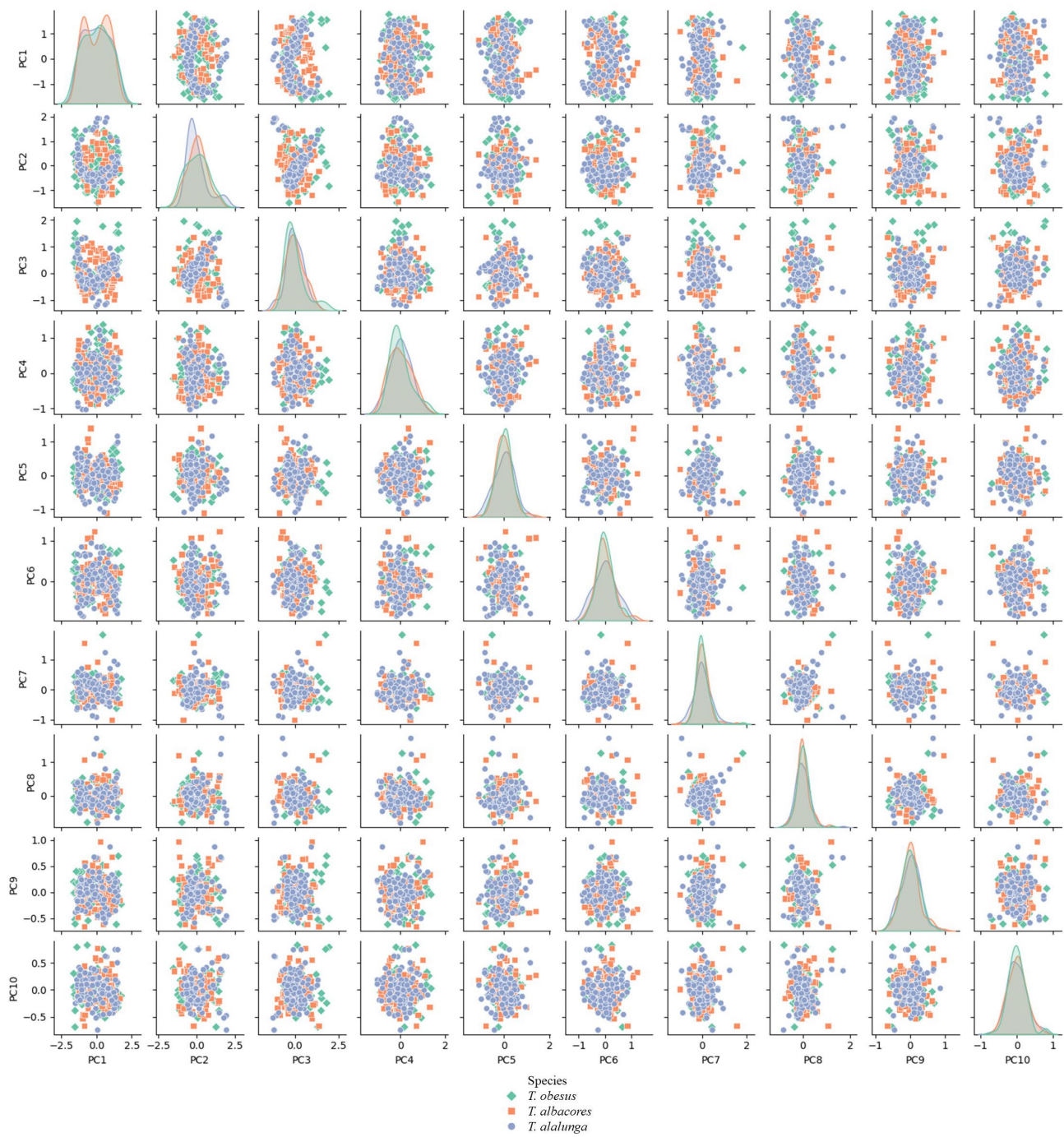


Figure 5. Principal component analysis of EFT features: the principal component had 100 tuna for each species.

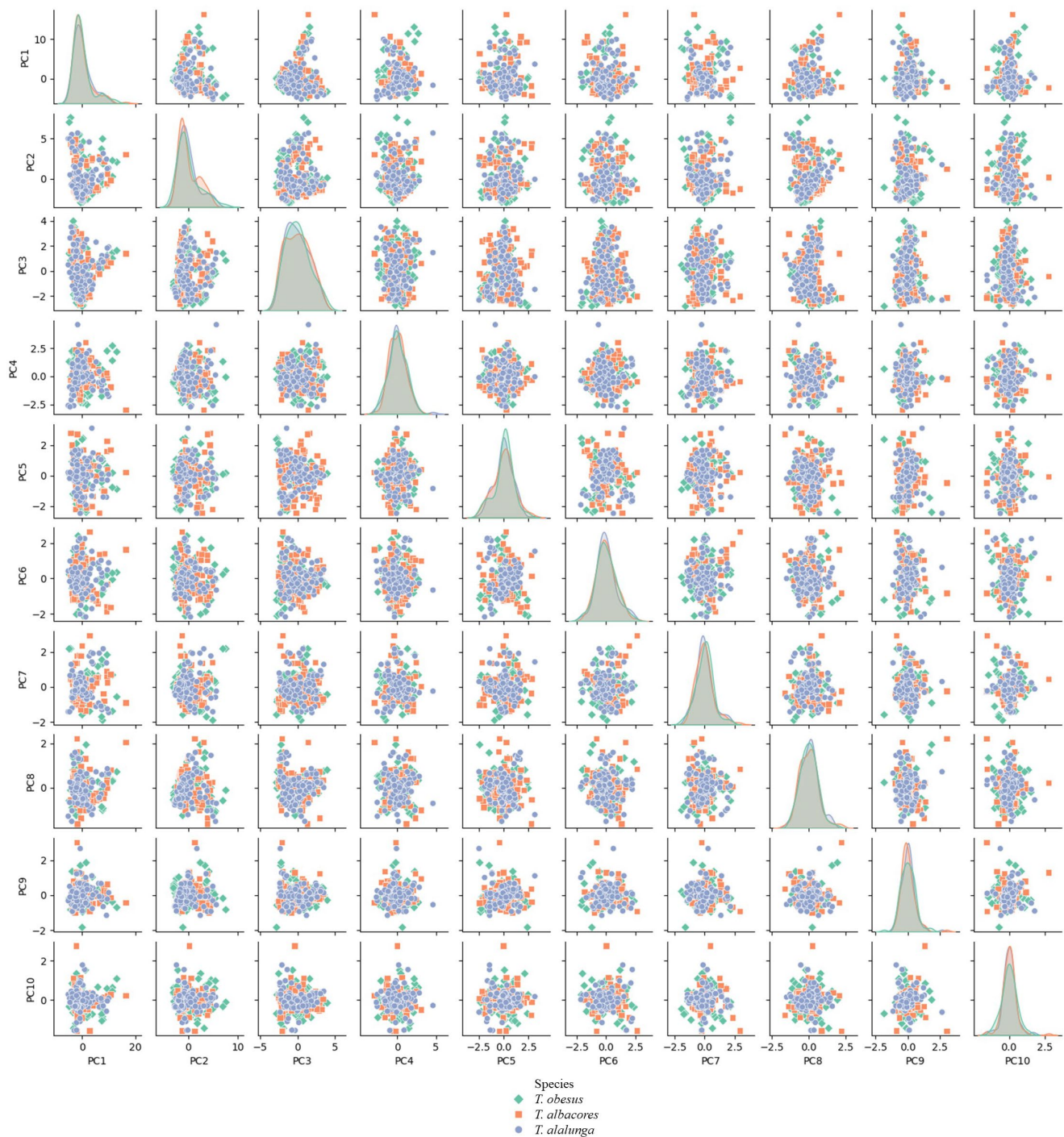


Figure 6. Principal component analysis of deep features: the principal component had 100 tuna for each species.

3.3. Evaluation Metrics of Different Machine Learning Algorithms

The evaluation metrics for the analysis of the three *Thunnus* species showed that different machine learning algorithms had different levels of identification performance across different morphological characteristics and therefore different species. According to the evaluation metric analysis of EFT features, the average performance of different machine learning algorithms was 82% in precision, recall, and F1-score (Table 3). Among them, KNN had the highest average performance with precision, recall, and F1-score, averaging 89%, 88%, and 88%, respectively. The average performance of the three machine

learning algorithms for deep features was 87% precision, 86% recall, and 86% F1-score (Table 4). Among them, SVM was the best, with 91% precision, 90% recall, and 90% F1-score.

Table 3. Values of performance evaluation metrics obtained using EFT features.

Algorithm	Species	Precision	Recall	F1-Score
SVM	<i>T. obesus</i>	84%	80%	82%
	<i>T. albacores</i>	94%	85%	89%
	<i>T. alalunga</i>	78%	90%	84%
	mean	85%	85%	85%
RF	<i>T. obesus</i>	78%	70%	74%
	<i>T. albacores</i>	62%	75%	68%
	<i>T. alalunga</i>	78%	70%	74%
	mean	73%	72%	72%
KNN	<i>T. obesus</i>	86%	90%	88%
	<i>T. albacores</i>	86%	90%	88%
	<i>T. alalunga</i>	94%	85%	89%
	mean	89%	88%	88%

Table 4. Values of performance evaluation metrics obtained using deep features.

Algorithm	Species	Precision	Recall	F1-Score
SVM	<i>T. obesus</i>	100%	80%	89%
	<i>T. albacores</i>	82%	90%	86%
	<i>T. alalunga</i>	91%	100%	95%
	mean	91%	90%	90%
RF	<i>T. obesus</i>	77%	85%	81%
	<i>T. albacores</i>	83%	75%	79%
	<i>T. alalunga</i>	85%	85%	85%
	mean	82%	82%	82%
KNN	<i>T. obesus</i>	94%	85%	89%
	<i>T. albacores</i>	89%	80%	84%
	<i>T. alalunga</i>	79%	95%	86%
	mean	87%	87%	86%

3.4. ROC Curves and AUC Values of *Thunnus* Species

ROC curves were analyzed to evaluate the performance of different machine algorithms for the identification of *Thunnus* species. For the average AUC values of the two morphological features of tuna, KNN had the best performance of 0.869, followed by SVM with 0.8453 and RF with 0.844 (Figure 7). The average performance of EFT features on three machine learning algorithms was 0.8292, and that of deep features was 0.8763. Therefore, the AUC value showed that there was a difference in the performance of EFT features and deep features for the identification of morphological information, and EFT features and deep features had their own advantages within different algorithms.

3.5. Comparison of EFT Features and Deep Features Using Confusion Matrix

EFT features and deep features were used for predicted number analyses. Based on the obtained results of EFT features, KNN had the lowest number of prediction errors, followed by SVM and RF (Figure 8). KNN had two, two, and three errors in the prediction of *T. obesus*, *T. albacores*, and *T. alalunga*, respectively. The prediction results of deep features were different from those of EFT features. SVM had the fewest prediction errors, followed by RF and KNN. The number of prediction errors of SVM for *T. obesus*, *T. albacores*, and *T. alalunga* was four, two, and zero, respectively.

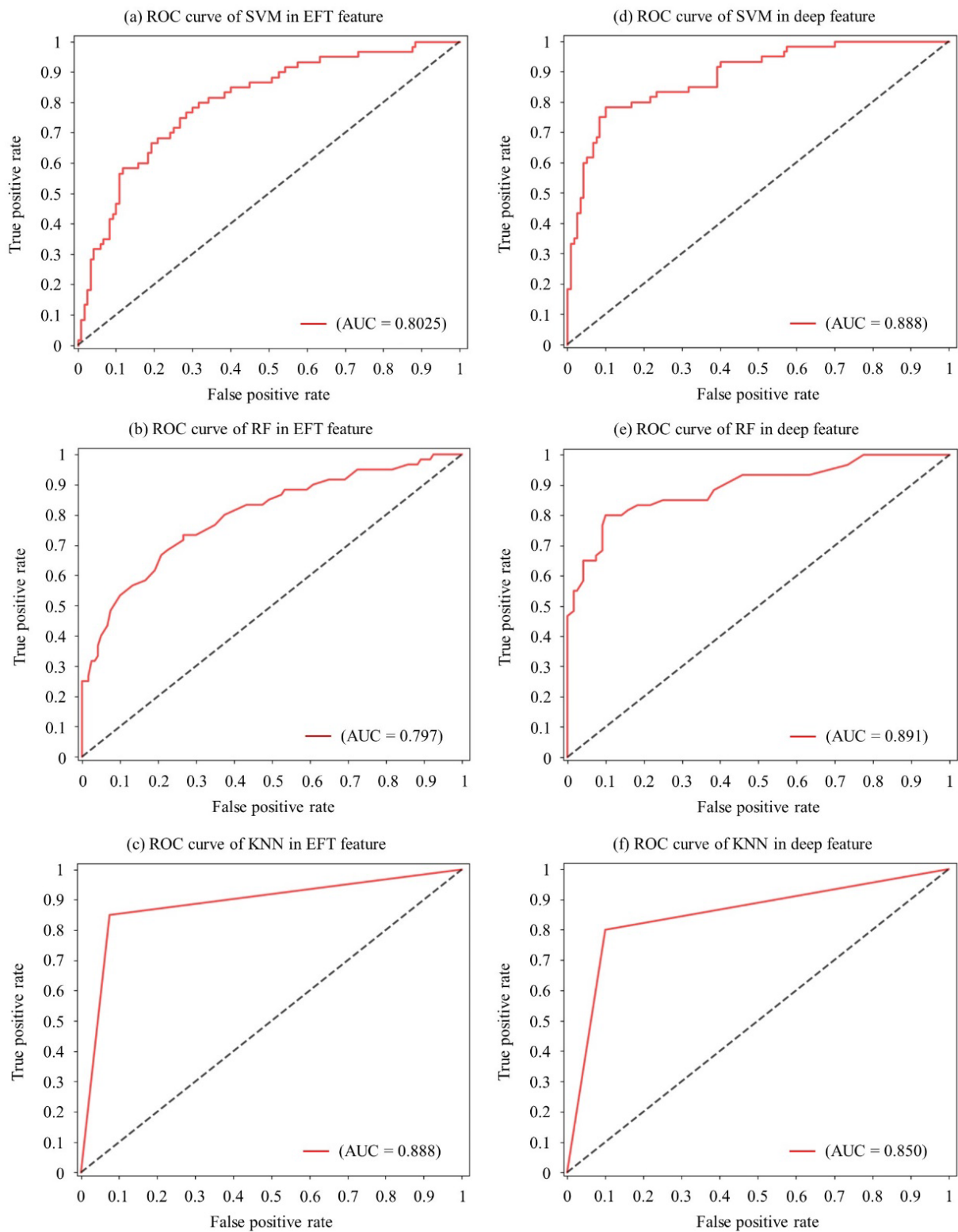


Figure 7. ROC curves and AUC values for different machine learning algorithms: from (a) to (c) is EFT feature; from (d) to (f) is deep features; (a,d) represent SVM; (b,e) represent RF; (c,f) represent KNN. The red curve represents the average ROC curve of the three tuna species.

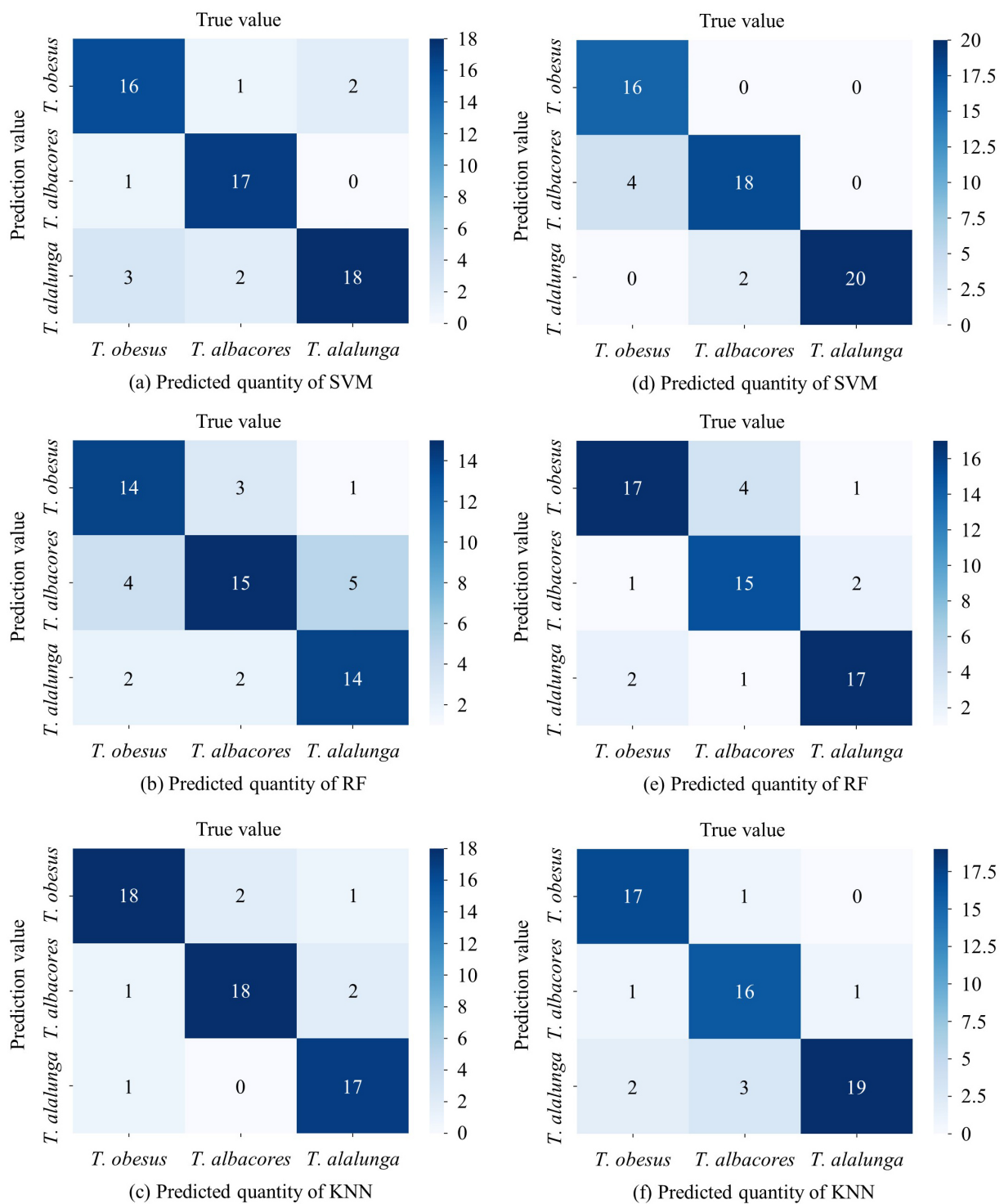


Figure 8. Predicted number of different machine learning algorithms in *Thunnus* species: from (a) to (c) is the confusion matrix of the EFT features; from (d) to (f) is the confusion matrix of the deep features; (a,d) represent SVM; (b,e) represent RF; (c,f) represent KNN.

Through the analysis of identification accuracy using EFT features and deep features, the results showed that both features could identify tuna. In different machine learning algorithms, the average identification performance of EFT features was 82%, while that of deep features was 86% (Figure 9). The identification performance of EFT features in the KNN algorithm was the highest. The identification accuracies of *T. obesus*, *T. albacores*, and *T. alalunga* were 90%, 90%, and 85%, respectively. Deep features showed the highest

identification performance among the SVM algorithms, with 80%, 90%, and 100% accuracy for *T. obesus*, *T. albacores*, and *T. alalunga*, respectively.

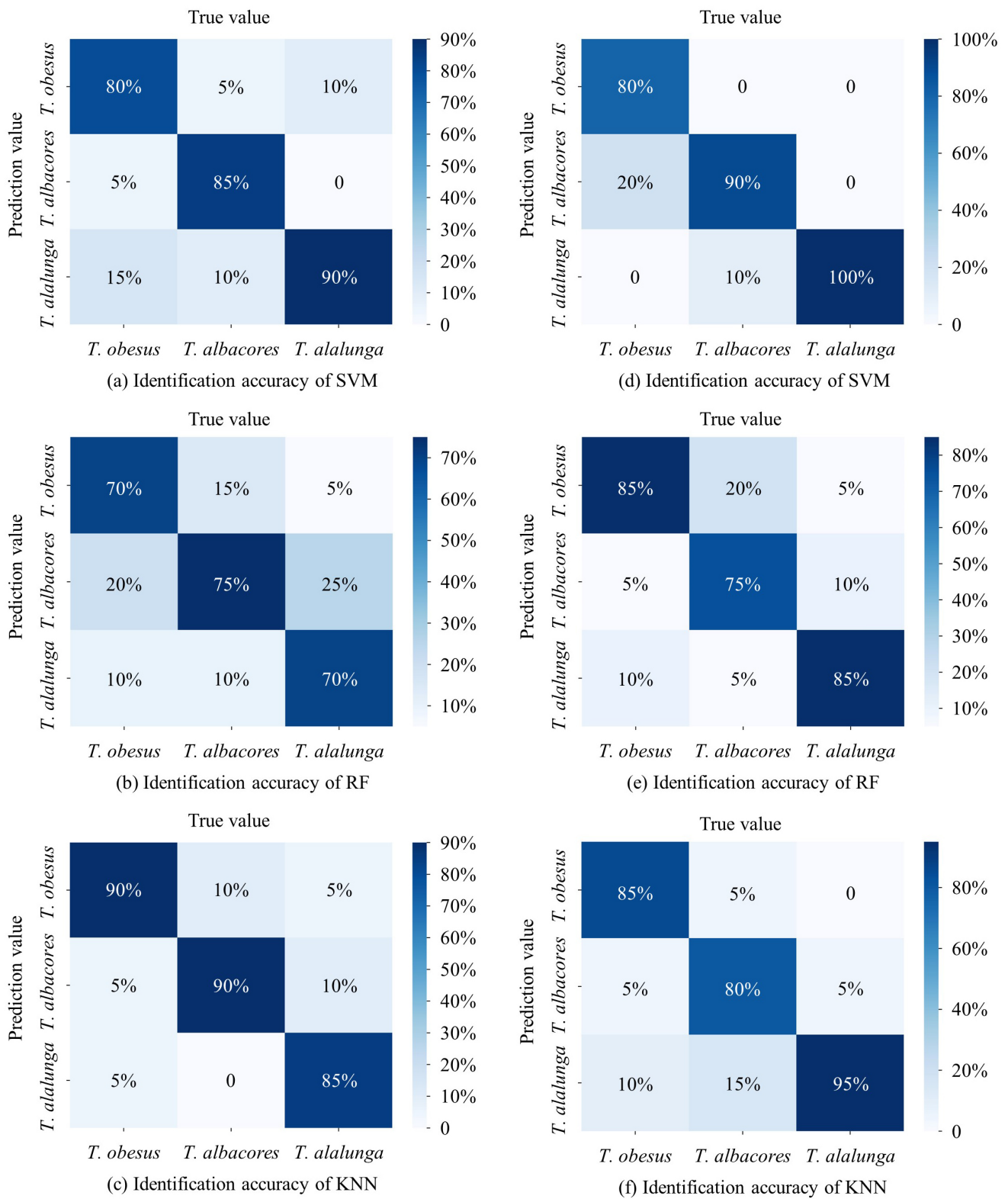


Figure 9. Identification accuracy of different machine learning algorithms in *Thunnus* species: from (a) to (c) is the confusion matrix of the EFT features; from (d) to (f) is the confusion matrix of the deep features; (a,d) represent SVM; (b,e) represent RF; (c,f) represent KNN.

4. Discussion

4.1. Visual Analysis of Morphology of Genus *Thunnus* Species

Morphological characteristics are an important feature commonly used in automated fish analyses [41]. On the one hand, morphological characteristics are determined by the nature of fish. Many fish species can adapt to ecosystem modification because they are naturally plastic in terms of their morphology [42,43]. On the other hand, morphological characteristics are determined when the fish develops. Morphological characteristics in fish are often associated with development, growth rate, nutrition, and environmental conditions [44,45]. Morphological information has significant advantages in fish bioecology, so it has a scientific theoretical basis for the automated analysis of fish morphological characteristics [46,47]. The analysis of morphological characteristics allows a detailed description of their shape and outline, which can be compared intra- or interspecifically [48,49]. Morphological characteristics have been used for the automated identification of fish [17,50]. A key need for the monitoring of fisheries is comprehensive and reliable fish data [51]. Hence, the automated analyses of fish morphology can facilitate electronic access to information, and the automated identification of tuna would contribute to the precise quantification of the catch, thus fostering the development of modern fishery management [52].

The two morphological characterization methods could well identify the three *Thunnus* species. In the morphological reconstruction of the EFT method, the change process of tuna morphology was from abstract to concrete. When the harmonic number was 10, the first dorsal, second dorsal, pelvic, and anal fins were initially formed. When the harmonic number was 20, each fin part was completely formed. At the same time, three different tuna characteristics could be clearly distinguished at harmonic 20. In the morphological visualization analysis of the CNN method, the change process of tuna morphology was from concrete to abstract. In this method, the change from morphological features (shallow features) to semantic features (deep features) could be visually seen from block 1 to block 5. VGG16 had five blocks. During the convolution process from block 1 to block 4, the morphological differences in the three *Thunnus* species could be found. Especially in block 4, the fin parts of the three tuna species were all perceived by CNN as important feature parts (where bright colors appeared in the image). By comparing the visual images of the EFT and CNN methods, we confirmed that CNN could also qualitatively analyze the morphological diversity of tuna.

4.2. Automated Identification of Different Tuna Species Using Machine Learning Algorithms

The data of the two different morphological features were analyzed using principal component analysis. The cumulative contribution rate of the two morphological features in the top ten principal components was more than 80%. It was further found that each principal component of morphological characteristic data showed significant interspecific differences among *Thunnus* species. It was preliminarily verified that the obtained data of different morphological features were good, and they provided a good database for subsequent automated identification.

The principal component data of different tuna morphologies were analyzed for evaluation metrics, and the identification performance of deep feature data was somewhat better than that of EFT feature data. In the average performance analysis of the three machine learning algorithms for EFT feature data, the accuracy rate was 81% for *T. obesus*, 82% for *T. albacores*, and 82% for *T. alalunga*. In the performance analysis of individual machine learning algorithms for EFT feature data, KNN performed the best, with an average of 88% for *T. obesus*, 88% for *T. albacores*, and 89% for *T. alalunga*. In the average performance analysis of the three machine learning algorithms for deep feature data, the accuracy rate was 87% for *T. obesus*, 83% for *T. albacores*, and 89% for *T. alalunga*. In the performance analysis of individual machine learning algorithms, SVM performed the best in deep feature data, with an average of 90% for *T. obesus*, 86% for *T. albacores*, and 95% for *T. alalunga*. In addition, the ROC curve and AUC value analysis results were similar to

the evaluation metrics, which further verified that the identification performance of deep feature data was better than that of EFT feature data.

By comparing EFT features with deep features, the results of the confusion matrix analysis showed that the most advanced CNN had significant advantages in extracting the morphological characteristics of tuna outlines. The results of the confusion matrix analysis were consistent with the performance analysis of the evaluation metrics. The morphological outline analysis of deep features had better robustness in the average performance of the different machine learning algorithms (the average identification performance of EFT features was 82%, while that of deep features was 86%). Among them, the identification performance of EFT features in the KNN algorithm was the best. The identification accuracies for *T. obesus*, *T. albacores*, and *T. alalunga* were 90%, 90%, and 85%, respectively. KNN had two, two, and three errors in the prediction of *T. obesus*, *T. albacores*, and *T. alalunga*, respectively. Deep features showed the highest identification performance using the SVM algorithm, with 80%, 90%, and 100% accuracy for *T. obesus*, *T. albacores*, and *T. alalunga*, respectively. The number of errors in the prediction of *T. obesus*, *T. albacores*, and *T. alalunga* using SVM was four, two, and zero, respectively. Due to the high similarity of the overall morphology of tuna, the overall identification accuracy was only above 80% on average. The difference among species was mainly reflected in the fin. Therefore, the identification performance of each species of tuna using different morphological characteristics and different machine learning algorithms revealed significant differences. However, each tuna species has essential diversity characteristics, and therefore the identification performance score was sufficiently high to effectively distinguish different tuna.

In addition, similar to previous studies, deep features combined with machine learning algorithms directly extracted all the information of the original fish images for research and consequently obtained good identification results [31,32]. This further verifies that deep features combined with a machine learning algorithm can be applied to fish species identification [31,32]. Future studies may better apply CNN to fish biology research and fish diversity analysis by identifying tuna of the same genus and different species based on their morphological outlines.

5. Conclusions

An automated method for the identification of three *Thunnus* species through morphological features based on different machine learning algorithms was proposed in this study. This method uses EFT features (morphometrics) and deep features (CNN) methods to conduct morphological visualization analysis of tuna, and perform principal component analysis and identification performance analysis, which can effectively study the interspecific differences and biodiversity of morphological characters of tuna species of the same genus. Among different machine learning algorithms to identify tuna, EFT features revealed the best identification performance in KNN, with an average identification accuracy rate of 88%. Deep features showed the best performance in SVM, with an average of 90%. In conclusion, the results obtained indicate that the analysis of morphological outlines (e.g., EFT features and deep features) supports the existence of three morphotypes (diversity of interspecific differences) within the study *Thunnus*. This study verifies the interspecific differences in the morphological characteristics of different tuna species through different performance analytic methods. Automated identification using different machine learning algorithms is helpful to further understand the morphological specificity of tuna. The comparative analysis of different morphological characteristics shows that the morphological data of deep features can be used to analyze the biodiversity of tuna. This method will promote the development of tuna fishery monitoring and provide a feasible strategy for fish biodiversity research in the future.

Author Contributions: Conceptualization, L.O. and B.L.; data curation, L.O.; investigation, L.O.; methodology, L.O.; project administration, B.L. and X.C.; resources, L.O.; software, L.O.; supervision, B.L., X.C., Q.H., W.Q. and L.Z.; validation, L.O.; writing—original draft preparation, L.O.;

writing—review and editing, L.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Program on the Survey, National Key R&D Plan (2019YFDO901404); Program on the Survey, Monitoring, and Assessment of Global Fishery Resources (comprehensive scientific survey of fishery resources at the high seas) sponsored by the Ministry of Agriculture and Rural Affairs; and Follow-Up Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning under Contract (22GZ03).

Institutional Review Board Statement: The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of Shanghai Ocean University (Approval Code 2022031501, approved on 15 March 2022).

Data Availability Statement: The data in this study are part of a larger research project. The findings of this study are not currently publicly available; however, they would be available when the overall research project is completed.

Acknowledgments: We thank those who helped us with the collection of tuna images.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, H.; Zhang, S.; Zhao, S.; Lu, J.; Wang, Y.; Li, D.; Zhao, R. Fast detection of cannibalism behavior of juvenile fish based on deep learning. *Comput. Electron. Agric.* **2022**, *198*, 107033. [\[CrossRef\]](#)
2. McCluney, J.K.; Anderson, C.M.; Anderson, J.L. The fishery performance indicators for global tuna fisheries. *Nat. Commun.* **2019**, *10*, 1641. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Herpandi, N.H.; Rosma, A.; Wan Nadiah, W.A. The tuna fishing industry: A new outlook on fish protein hydrolysates. *Compr. Rev. Food Sci. Food Saf.* **2011**, *10*, 195–207. [\[CrossRef\]](#)
4. Mata, W.; Chanmalee, T.; Punyasuk, N.; Thitamadee, S. Simple PCR-RFLP detection method for genus and species-authentication of four types of tuna used in canned tuna industry. *Food Control.* **2020**, *108*, 106842. [\[CrossRef\]](#)
5. Lin, Q.; Chen, Y.; Zhu, J. A comparative analysis of the ecological impacts of Chinese tuna longline fishery on the Eastern Pacific Ocean. *Ecol. Indic.* **2022**, *143*, 109284. [\[CrossRef\]](#)
6. Guisande, C.; Manjarrés-Hernández, A.; Pelayo-Villamil, P.; Granado-Lorencio, C.; Riveiro, I.; Acuña, A.; Prieto-Piraquive, E.; Janeiro, E.; Matías, J.M.; Patti, C.; et al. IPEz: An expert system for the taxonomic identification of fishes based on machine learning techniques. *Fish. Res.* **2010**, *102*, 240–247. [\[CrossRef\]](#)
7. Batubara, A.S.; Muchlisin, Z.A.; Efizon, D.; Elvyra, R.; Fadli, N.; Irham, M. Morphometric variations of the genus *Barbonymus* (Pisces, Cyprinidae) harvested from Aceh waters, Indonesia. *Fish. Aquat. Life* **2018**, *26*, 231–237. [\[CrossRef\]](#)
8. Rahayu, S.R.; Muchlisin, Z.A.; Fadli, N. Morphometric and genetic variations of two dominant species of snappers (Lutjanidae) harvested from the Northern Coast of Aceh waters, Indonesia. *Zool. Anz.* **2023**, *303*, 26–32. [\[CrossRef\]](#)
9. Díaz-Cruz, J.A.; Alvarado-Ortega, J.; Ramírez-Sánchez, M.M.; Bernard, E.L.; Allington-Jones, L.; Graham, M. Phylogenetic morphometrics, geometric morphometrics and the Mexican fossils to understand evolutionary trends of enchodontid fishes. *J. S. Am. Earth. Sci.* **2021**, *111*, 103492. [\[CrossRef\]](#)
10. Li, Y.; Burrige, C.P.; Lv, Y.; Peng, Z. Morphometric and population genomic evidence for species divergence in the *Chimarrichthys* fish complex of the Tibetan Plateau. *Mol. Phylogenet. Evol.* **2021**, *159*, 107117. [\[CrossRef\]](#)
11. Hanif, M.A.; Siddik, M.A.; Islam, M.A.; Chaklader, M.R.; Nahar, A. Multivariate morphometric variability in sardine, *Amblygaster clupeioides* (Bleeker, 1849), from the Bay of Bengal coast, Bangladesh. *J. Basic Appl. Zool.* **2019**, *80*, 53. [\[CrossRef\]](#)
12. Nur, F.M.; Batubara, A.S.; Fadli, N.; Rizal, S.; Siti-Azizah, M.N.; Muchlisin, Z.A. Elucidating species diversity of genus *Betta* from Aceh waters Indonesia using morphometric and genetic data. *Zool. Anz.* **2022**, *296*, 129–140. [\[CrossRef\]](#)
13. Yulianto, D.; Indra, I.; Batubara, A.S.; Nur, F.M.; Rizal, S.; Siti-Azizah, M.N.; Muchlisin, Z.A. Morphometrics and genetics variations of mullets (Pisces: Mugilidae) from Aceh waters, Indonesia. *Biodiversitas J. Biol. Divers.* **2020**, *21*, 3422–3430. [\[CrossRef\]](#)
14. Zhao, S.; Zhang, S.; Liu, J.; Wang, H.; Zhu, J.; Li, D.; Zhao, R. Application of machine learning in intelligent fish aquaculture: A review. *Aquaculture* **2021**, *540*, 736724. [\[CrossRef\]](#)
15. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [\[CrossRef\]](#)
16. Alsmadi, M.K.; Almarashdeh, I. A survey on fish classification techniques. *J. King. Saud. Univ. Com. Inf. Sci.* **2020**, *34*, 1625–1638. [\[CrossRef\]](#)
17. Hu, J.; Li, D.; Duan, Q.; Han, Y.; Chen, G.; Si, X. Fish species classification by color, texture and multi-class support vector machine using computer vision. *Comput. Electron. Agric.* **2012**, *88*, 133–140. [\[CrossRef\]](#)
18. Xu, X.; Li, W.; Duan, Q. Transfer learning and SE-ResNet152 networks-based for small-scale unbalanced fish species identification. *Comput. Electron. Agric.* **2021**, *180*, 105878. [\[CrossRef\]](#)

19. Ogunlana, S.O.; Olabode, O.; Oluwadare, S.A.A.; Iwasokun, G.B. Fish classification using support vector machine. *Afr. J. Comput. ICT* **2015**, *8*, 75–82.
20. Tharwat, A.; Hemedan, A.A.; Hassanien, A.E.; Gabel, T. A biometric-based model for fish species classification. *Fish. Res.* **2018**, *204*, 324–336. [[CrossRef](#)]
21. Andayani, U.; Wijaya, A.; Rahmat, R.F.; Siregar, B.; Syahputra, M.F. Fish species classification using probabilistic neural network. *J. Phys. Conf. Ser.* **2019**, *1235*, 012094. [[CrossRef](#)]
22. Strachan, N.J.C.; Nesvadba, P.; Allen, A.R. Fish species recognition by shape analysis of images. *Pattern. Recogn.* **1990**, *23*, 539–544. [[CrossRef](#)]
23. Salman, A.; Jalal, A.; Shafait, F.; Mian, A.; Shortis, M.; Seager, J.; Harvey, E. Fish species classification in unconstrained underwater environments based on deep learning. *Limnol. Oceanogr. Methods* **2016**, *14*, 570–585. [[CrossRef](#)]
24. Taheri-Garavand, A.; Nasiri, A.; Banan, A.; Zhang, Y.D. Smart deep learning-based approach for non-destructive freshness diagnosis of common carp fish. *J. Food. Eng.* **2020**, *278*, 109930. [[CrossRef](#)]
25. Bui, H.M.; Lech, M.; Cheng, E.; Neville, K.; Burnett, I.S. Object recognition using deep convolutional features transformed by a recursive network structure. *IEEE Access* **2016**, *4*, 10059–10066. [[CrossRef](#)]
26. Iqbal, M.A.; Wang, Z.; Ali, Z.A.; Riaz, S. Automatic fish species classification using deep convolutional neural networks. *Wireless. Pers. Commun.* **2021**, *116*, 1043–1053. [[CrossRef](#)]
27. Khan, S.; Yairi, T. A review on the application of deep learning in system health management. *Mech. Syst. Signal. Process.* **2018**, *107*, 241–265. [[CrossRef](#)]
28. Rathi, D.; Jain, S.; Indu, S. Underwater fish species classification using convolutional neural network and deep learning. In Proceedings of the 2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR), Bangalore, India, 27–30 December 2017.
29. Villon, S.; Mouillot, D.; Chaumont, M.; Darling, E.S.; Subsol, G.; Claverie, T.; Villéger, S. A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecol. Inform.* **2018**, *48*, 238–244. [[CrossRef](#)]
30. Rekha, B.S.; Srinivasan, G.N.; Reddy, S.K.; Kakwani, D.; Bhattad, N. Fish detection and classification using convolutional neural networks. In *Computational Vision and Bio-Inspired Computing: ICCVBIC 2019*; Springer: Cham, Switzerland, 2020.
31. Tamou, A.B.; Benzinou, A.; Nasreddine, K. Underwater live fish recognition by deep learning. In *Image and Signal Processing*; Springer: Cham, Switzerland, 2018.
32. Deep, B.V.; Dash, R. Underwater fish species recognition using deep learning techniques. In Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 7–8 March 2019.
33. Kuhl, F.P.; Giardina, C.R. Elliptic Fourier features of a closed contour. *Comput. Graph. Image Process.* **1982**, *18*, 236–258. [[CrossRef](#)]
34. Fauzan, M.H.N.; Rakun, E.; Hardianto, D. Feature Extraction from Smartphone Images by Using Elliptical Fourier Descriptor, Centroid and Area for Recognizing Indonesian Sign Language SIBI (Sistem Isyarat Bahasa Indonesia). In Proceedings of the 2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS), Singapore, 28 February–2 March 2019.
35. Wei, S.J.; Al Riza, D.F.; Nugroho, H. Comparative study on the performance of deep learning implementation in the edge computing: Case study on the plant leaf disease identification. *J. Agric. Food Res.* **2022**, *10*, 100389. [[CrossRef](#)]
36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv Prepr.* **2014**, arXiv:1409.1556.
37. Ren, L.; Tian, Y.; Yang, X.; Wang, Q.; Wang, L.; Geng, X.; Wang, K.; Du, Z.; Lia, Y.; Lin, H. Rapid identification of fish species by laser-induced breakdown spectroscopy and Raman spectroscopy coupled with machine learning methods. *Food. Chem.* **2023**, *400*, 134043. [[CrossRef](#)]
38. Zhu, Y.; Weiyi, X.U.; Luo, G.; Wang, H.; Yang, J.; Lu, W. Random Forest enhancement using improved Artificial Fish Swarm for the medial knee contact force prediction. *Artif. Intell. Med.* **2020**, *103*, 101811. [[CrossRef](#)] [[PubMed](#)]
39. Politikos, D.V.; Petasis, G.; Chatzispayrou, A.; Mytilineou, C.; Anastasopoulou, A. Automating fish age estimation combining otolith images and deep learning: The role of multitask learning. *Fish Res.* **2021**, *242*, 106033. [[CrossRef](#)]
40. Dogan, M.; Taspinar, Y.S.; Cinar, I.; Kursun, R.; Ozkan, I.A.; Koklu, M. Dry bean cultivars classification using deep cnn features and salp swarm algorithm based extreme learning machine. *Comput. Electron. Agric.* **2023**, *204*, 107575. [[CrossRef](#)]
41. Hsieh, C.L.; Chang, H.Y.; Chen, F.H.; Liou, J.H.; Chang, S.K.; Lin, T.T. A simple and effective digital imaging approach for tuna fish length measurement compatible with fishing operations. *Comput. Electron. Agric.* **2011**, *75*, 44–51. [[CrossRef](#)]
42. Elliott, M.; Whitfield, A.K.; Potter, I.C.; Blaber, S.J.; Cyrus, D.P.; Nordlie, F.G.; Harrison, T.D. The guild approach to categorizing estuarine fish assemblages: A global review. *Fish Fish.* **2007**, *8*, 241–268. [[CrossRef](#)]
43. Whitfield, A.K.; Elliott, M. Fishes as indicators of environmental and ecological changes within estuaries: A review of progress and some suggestions for the future. *J. Fish. Biol.* **2002**, *61*, 229–250. [[CrossRef](#)]
44. Canty, S.W.; Truelove, N.K.; Preziosi, R.F.; Chenery, S.; Horstwood, M.A.; Box, S.J. Evaluating tools for the spatial management of fisheries. *J. Appl. Ecol.* **2018**, *55*, 2997–3004. [[CrossRef](#)]
45. Floeter, S.R.; Bender, M.G.; Siqueira, A.C.; Cowman, P.F. Phylogenetic perspectives on reef fish functional traits. *Biol. Rev.* **2018**, *93*, 131–151. [[CrossRef](#)]
46. Strachan, N.J.C. Length measurement of fish by computer vision. *Comput. Electron. Agric.* **1993**, *8*, 93–104. [[CrossRef](#)]

47. Khotimah, W.N.; Arifin, A.Z.; Yuniarti, A.; Wijaya, A.Y.; Navastara, D.A.; Kalbuadi, M.A. Tuna fish classification using decision tree algorithm and image processing method. In Proceedings of the 2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Bandung, Indonesia, 5–7 October 2015.
48. Almeida, P.R.; Monteiro-Neto, C.; Tubino, R.A.; Costa, M.R. Variações na forma do otólito sagitta de *Coryphaena hippurus* (Actinopterygii: Coryphaenidae) em uma área de ressurgência na costa sudoeste do Oceano Atlântico. *Iheringia. Série Zool.* **2020**, *110*. [[CrossRef](#)]
49. Bakhshalizadeh, S.; Abbasi, K.; Rostamzadeh Liafuie, A.; Bani, A.; Pavithran, A.; Tiralongo, F. Morphometric Analyses of Phenotypic Plasticity in Habitat Use in Two Caspian Sea Mulletts. *J. Mar. Sci. Eng.* **2022**, *10*, 1398. [[CrossRef](#)]
50. Saputra, W.A.; Herumurti, D. Integration GLCM and geometric feature extraction of region of interest for classifying tuna. In Proceedings of the 2016 International Conference on Information & Communication Technology and Systems (ICTS), Surabaya, Indonesia, 12 October 2016.
51. Qiao, M.; Wang, D.; Tuck, G.N.; Little, L.R.; Punt, A.E.; Gerner, M. Deep learning methods applied to electronic monitoring data: Automated catch event detection for longline fishing. *ICES J. Mar. Sci.* **2021**, *78*, 25–35. [[CrossRef](#)]
52. Khokher, M.R.; Little, L.R.; Tuck, G.N.; Smith, D.V.; Qiao, M.; Devine, C.; O'Neill, H.; Pogonoski, J.; Arangio, R.; Wang, D. Early lessons in deploying cameras and artificial intelligence technology for fisheries catch monitoring: Where machine learning meets commercial fishing. *Can. J. Fish. Aquat. Sci.* **2022**, *79*, 257–266. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.