EDIBLE FISH IDENTIFICATION BASED ON MACHINE LEARNING

Israa Mohammed Hassoon[1]

[1]*Department of Mathematics, College of Science, University of Mustansiriyah (UOM), Baghdad-Iraq*

isrmo9@uomustansiriyah.edu.iq

Shaymaa Akram Hantoosh[2]

[2]*Middle Technical University, Continuous Education Center, Baghdad, Iraq*
shymmaakram35@mtu.edu.iq

Abstract - **Automated fish identification system has a beneficial role in various fields. Fish species can usually be identified based on visual observation and human experiences. False appreciation can cause food poisoning. The proposed system aims to efficiently and effectively identify edible fish from poisonous ones based on three machine learning (ML) techniques. A total of 300 fish images are used, collected from 20 species with differences in shapes, sizes, and colors. Hybrid features were extracted and then fed to three types of ML techniques: k-nearest neighbor (K-NN), support vector machine (SVM), and neural networks (NN). The ٣٠٠ fish images are divided into two: 70% for training and 30% for testing. The accuracy rates for the presented system were 91.1%, 92.2%, and 94.4% for KNN, SVM, and NNs, respectively. The proposed system is evaluated using four terms: precision, sensitivity, F1-score, and accuracy. Results show that the proposed approach achieved higher accuracy compared with other recent pertinent studies.**

Index Terms - Edible Fish, High Order Statistical Features, Machine Learning, Poisonous Fish, Second Order Statistical Features.

I. INTRODUCTION

Various types of fish found in seas and rivers occasionally have the same shapes and sizes, resulting in difficulty distinguishing them from one another. Given that consumers are looking for fresh or high-quality fish, the huge challenge facing fishers is the laborious method of manually confirming the fish species with high valuable, fresh, edible, dangerous, invasive, or poisonous. Another challenge is underwater environment, where the task of identifying fish species is challenged by the complexity of the background, water turbidity, and light dispersal in deep water [1]. Distinguishing one fish species from another based on external semblance is still difficult because of the considerable similarity. Current research remains incapable of identifying which fish are dangerous or not. Accordingly, a reliable and authentic identification of fish is crucial to avert fish mislabeling in markets [2]. The best possible techniques for fish authentication are to identify fish species using morphometric and morphological characteristics [3]. Detecting edible, poisonous, or dangerous fish is a complicated and critical task, and the use of wrong information caused by lack of experience may lead to death in many cases; hence, particular knowledge of taxonomy framework and fish biological proprieties are required [4]. The correct identification of fish species (i.e., edible, poisonous

fish) helps in economic development and the preservation of human life. Accordingly, technology that can automatically detect poisonous fish must evolve to reduce waste of time and provide considerably accurate results. Machine learning (ML) plays an essential role in classification problems. In ML, computer programs can be updated automatically throughout experience. ML algorithms require large amounts of data and highly consume resources [5]. Pretrained data help to minimize data inadequacy consistent with a particular task. This technology is utilized by scientists to scrutinize large data and to automate complex tasks. ML techniques show better performance than traditional forecasting. Furthermore, ML techniques were used for defect identification in various fields [6]. ML techniques are classified into four classes: supervised, unsupervised, semi-supervised, and reinforcement. Supervised algorithms utilize labeled data. Learning data are used when identified specific goals should be achieved from a specific input [7]. Unsupervised algorithms utilize unlabeled data. Semi-supervised algorithms can be supervised and unsupervised because they can utilize unlabeled and labeled data sets [8]. Reinforcement algorithms enable machine and software agent to evaluate the optimal solution in a particular problem and are considered among the best tools for training artificial intelligent models [9]. The current study used three ML techniques, namely, k-nearest neighbors (K-NN), support vector machine (SVM), and neural networks (NNs), for reliable fish identification system by utilizing images from a significant database [10]. Fish images are preprocessed, and hybrid features (i.e., second and high order statistical features) are extracted from fish images. Thereafter, the fish classification algorithm is created using the three ML methods (i.e., K-NN, SVM, and NNs), and fish categories (i.e., edible, poisonous) can be identified thereafter.

The remainder of this paper is organized as follows. Section 2 reviews the related studies. Section 3 presents the methods and materials used in this research. Section 4 provides the experimental results. Lastly, Section 5 presents the conclusions and future research.

II. RELATED STUDIES

This section presents a concise review of some related studies.

In [1], an approach was proposed for classifying fish based on convolutional neural network (CNN) and incremental

learning strategy. A total of 15 fish species, 20 videos, and 22,000 sample images were used. The images have different sizes ranging from 20×20 to 200×200 pixels. Each video was labeled. Their data sets were based on blurry and noisy environments, poor lighting conditions, and plain/complex backgrounds. Their system trained ResNet50 on two fish species: hard and easy fish species. First, the system learn about hard fish species. Second, easy species were added thereafter by using incremental learning. Knowledge distillation loss was added to modify loss function, and used to learn a new fish species. The system achieved an accuracy of 85.26%.

In [11], an approach was proposed to classify fish into dangerous and non-dangerous fish: non-dangerous fish were classified into ornamental and edible fish, while dangerous fish were classified into poisonous or predatory fish. A combination of 39 anchor point features, 2 statistical features, and 24 features of GLCM were extracted from fish images. To classify the fish images, they used simulated the annealing and genetic algorithms with back propagation, and the accuracy rate was 82.25% and 90%, respectively.

In [12], the YOLO deep learning algorithm, which has the advantage of faster image detection, was used for fish identification. Their approach have several steps: building the YOLO algorithm, collecting and labeling fish images, training, and testing. Their approach can identify 24 fish species; the highest detection accuracy was in Holacanthus tricolor at 90.70%, while the lowest was in Forcipiger flavissimus at 68.28%. The total accuracy of their model was 82.82%. Another approach [13] used the YOLO detection method to classify six fish species from underwater video. The input fish image with size of 488×488 was divided into seven grids, with each grid representing a bounding box that provides a prediction score for the class. A total of 220 fish images were used as data set (120 for training and 100 for testing), and the accuracy of this approach was 92%.

In [14], fish species recognition for underwater images was proposed. In particular, 28,328 RGB fish images from 130 species of sizes 300×300 and 512×512 were used. As fish attribute extractors, the model used MobileNetv3-large and VGG16 backbone networks. Given that the data set was imbalanced, a class-aware loss function was used to solve this problem. The model's data set contained three categories: training data, which contain 70 samples, and validation and testing data, which contain 30 samples. The accuracy of their system was 79.7%.

In [15], a novel method for fish species detection and classification was presented. Fish images were obtained using boat cameras, and 3777 fish samples divided into 8 species were used. Data augmentation techniques were used to solve the problem of image shortage in some species. Hence, the total number of fish images was 16000. The VGG-16 network pretrained on the ImageNet data set was used as classifier. A total of 2000 patches were extracted for each species from the training data set. CNN was used for image detection and classification, and the model achieved detection and classification accuracies of 90% and 92%, respectively.

In [16], a system to monitor marine life was proposed. This system comprised three tasks. Fist, underwater moving fish were detected. Second, the detected fish were tracked to avoid errors in counting. Lastly, species of the detected fish were classified. The first task was performed using the GMM technique based on the background subtraction process. For the second task, the Kalman filter was used, which predicts the trajectory's location in each frame and determines the possibility of each detection being assigned to each trajectory. The Kalman filter could avoid errors in counting the tracked fish. The pyramid-histogram-of-visual-words features were extracted, and a classifier was built thereafter. SVM was used for classification (third task). The number of images used was 22,444. The accuracy rate of their system was 91.7%, but the accuracy of the low-quality video was 40.1%.

In [17], a hybrid local binary pattern was introduced to classify fish in Bangladesh. A new data set (i.e., BDIndigenousFish2019) was created, containing images of eight different fish species. Thereafter, hybrid features descriptors, such as LBP, NABP, LGP, CENTRIST, LAID and DTCTH, were used to extract significant features from the fish image. Meanwhile, different kernels of SVM were used for classification, and the experimental results showed accuracy of 90%.

III. RESEARCH MATERIALS AND METHODS

This research, which aims to identify edible/poisonous fish, comprises three phases; data collection, features extraction, and identification phases. Three ML techniques were used to identify fish categories (i.e., edible/ poisonous). The general architecture presenting the methods used in this research is shown in Fig. (1).
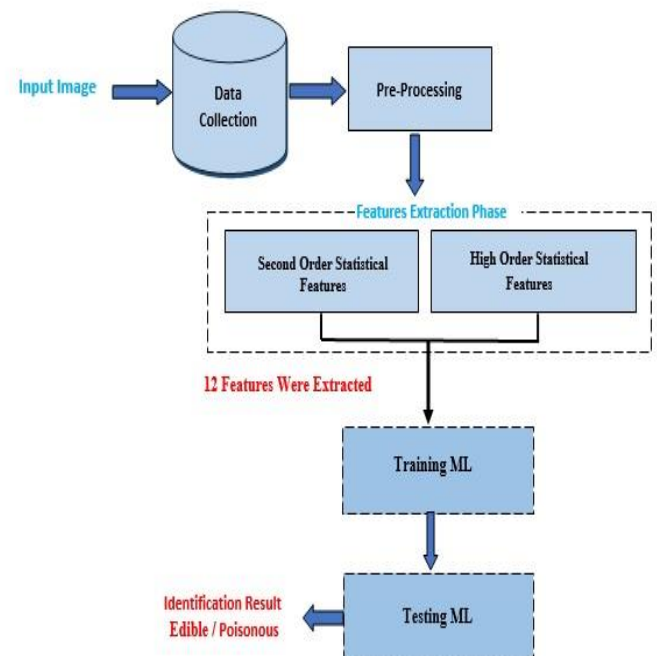

Fig. 1 Diagram of this Research

Himantura uarnak, (9) Pateobatis bleekeri, and (10) Arothron hispidus.

A . Data Collection

A total of 300 real-world fish images from 20 species were obtained in this research. The images were from major databases [10]. Moreover, the images were selected on the bases of differences in size, color, controlled/uncontrolled illumination, out of the water/in the water, and plain/complex background. The data set utilized in this research has two categories.

1) Edible Fish

A total of 10 species of edible fish were utilized in this research: 15 images for each species, for a total of 150 fish images selected. The two types of edible fish collected were freshwater and saltwater fish.

2) Poisonous Fish

A total of 10 species of poison fish were utilized in this research: 15 images for each species, for a total of 150 fish images selected. The two types of poisonous fish were poisonous and venomous fish. Poisonous fish contains toxins that cause poisoning if a person eats it. By contrast, venomous fish produces a vigorous mishmash of toxins that cause poisoning through sting, bite, or stab. Fig. (2) shows the fish image data set.
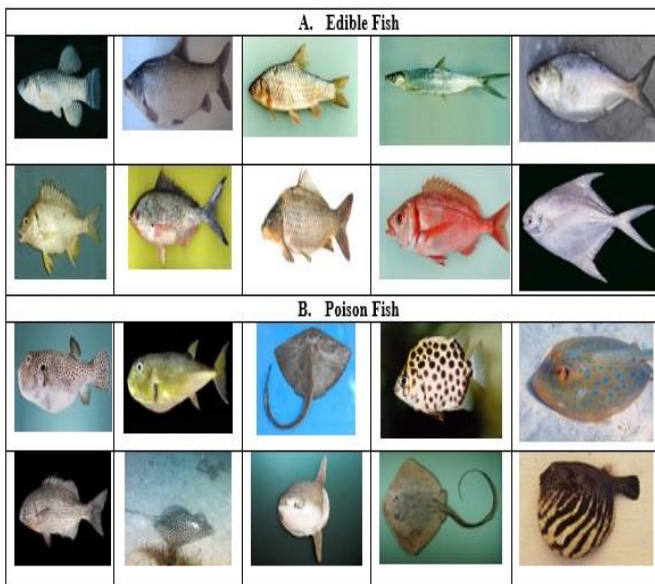

Fig. 2 Fish Image Data set; two categories: A (Edible fish) and B (Poison fish), each comprising 10 fish species. Category A: (1) Aphanius dispar, (2) barbus, (3) carps, (4) Chanos chanos, (5) Tenualosa ilisha, (6) Acanthopagrus arabicus, (7) Caesio lunaris, (8) Cyprinus carpio, (9) Dentex macrophthalmus, and (10) Pampus argenteus.
Category B: (1) Arothron stellatus, (2) Lagocephalus lunaris, (3) Pastinachus sephen, (4) Scatophagus argus, (5) Taeniura lymma, (6) Kyphosus cinerascens, (7) Mola mola, (8)

Several pre-processing steps were performed on the fish images. First, the RGB image was converted into a grayscale image. Second, background and noise were removed from an input image using the filter in [18]. Third, the fish images were cropped and resized to 256×256 to remove errors. Thereafter, all images are ready for the next stage.

B. Features Extraction

In this phase, a combination of second and high order statistical features were extracted (12). Significant and relevant features were extracted from the fish image (5). These features are second and (7) and high order statistical features.

1) Second Order Statistical Features

One of the specialized methods that used to obtain significant object details. The properties of any region can be determined by the distribution of the gray levels. The features of a region can also be described by using space relations of the gray level distribution of the input image. Different second order statistical features can be obtained based on the peculiarity of pixels to depict the image. In this research (5), features of the second order statistics features are obtained: energy, homogeneity, correlation, contrast and entropy. These features are clarified in Eqs. (1) to (5) [19]:

$$\text{Energy} = \Sigma i,j I(i,j)2, \qquad (1)$$

$$\text{Homogeneity} = \frac{I(j,i)}{1+|j-i|} \qquad (2)$$

$$\text{Correlation} = \Sigma i, j \frac{(i-\mu i)(j-\mu j)I(i,j)}{\sigma i \sigma j}, \qquad (3)$$

$$\text{Contrast} = \Sigma i,j \ (i,j)2 \ I(i,j), \qquad (4)$$

$$\text{Entropy} = - \Sigma i,j \ I(i,j)\log I(i,j), \qquad (5)$$

where:
i,j: number of rows and columns, respectively, in the gray matrix;
I(i,j): pixel intensity value probability;
μi, μj: means; and
σi, σj: standard deviations for rows and columns, respectively.

2) High Order Statistical Features

Gray level (GL) run length matrix is a high order statistical features technique that applies to a run length matrix. It is a method that can extract higher order statistical texture proprieties. Moreover, this technique is a number of straight pixels in a particular direction having an identical intensity value. It can extract image information from its GL matrix. GLRLM can keep the parameters stable, and simply changing the rotation and scale will obtain a perfectly new texture [20]. The main idea of GLRLM is computing the number of straight pixels having an identical GL in a particular direction. Thereafter, a matrix is created. The number of run length matrices can be calculated for one image. However, just one matrix can be calculated in each direction. Lastly, features are extracted from this matrix, as shown in Eqs. (6) to (12) [21]:

$$SR = \frac{1}{r}\sum_{i,j}\frac{P(i,j)}{j^2} \qquad (6)$$

$$LR = \frac{1}{r}\sum_{i,j}j^2\,p(i,j) \qquad (7)$$

$$GL = \frac{1}{r}\sum_i\left(\sum_j p(i,j)\right)^2, \qquad (8)$$

$$RPER = \sum_{i,j}\frac{1}{p(i,j)j} \qquad (9)$$

$$L= \frac{1}{r}\sum_j(\sum_i p(i,j))^2, \qquad (10)$$

$$LGR= \frac{1}{r}\sum_{i,j}\frac{p(i,j)}{i^2} \qquad (11)$$

$$HGR = \frac{1}{r}\sum_{i,j}i^2\,p(i,j), \qquad (12)$$

where:

SR: Short–Run–Emphasis;
LR: Long–Run–Emphasis;
GL: Gray–Level–Non-uniformity;
RPER: Run–Percentage;
RL: Run–Length–Non-uniformity;
LGR: Low–Gray–Level–Run–Emphasis;
HGR: High–Gray–Level–Run–Emphasis;
P(i,j): run length matrix;
i: number of runs having pixels with gray level intensity;
j: length of run on a particular direction; and
r: aggregate number of runs.

C. ML Techniques
1) K-NN

K-NN is one of the supervised ML algorithms widely used in solving classification problems. Supervised learning means that the input and output are given to the system. Some of the input data are labeled, and when a new sample is entered to the system, the output will present. K-NN uses a data set containing data points, and data points of a sample is given to K-NN as a classification problem. The K-NN algorithm finds the closest class to the input sample based on the distance between the new input sample and all labeled data sets. K-NN accuracy depends on predicting the correct K-NN [22], and finding the correct K-NN relies on how skillfully the large labeled data set has been trimmed. Meanwhile, finding the closest class is achieved using Euclidean distance, as shown in Eq. (13). K-NN is flexible, easy to implement, and appropriate for multi-modal. Although K-NN is considered the best classification method, noisy data and irrelevant characteristics will cause retrogression of accuracy [23].

$$Dis = \sqrt{(m2 - m1)^2 + (n2 - n1)^2} \qquad (13)$$

2) SVM

SVM is one of the linear classifiers widely utilized for recognition, classification, detection, and regression. In particular, SVM can separate two groups of data with as broad a margin as tolerable. This situation leads to perfect results on unseen data, thereby enabling SVM to learn from massive data [24]. SVM can classify objects based on training data sets and can deal with structured, semi-structured, and high dimensional data. However, SVM cannot work well if data are noisy. Occasionally, finding a suitable kernel function is difficult [25]. Let the data set be $(x_1, y_1, \ldots, x_k, y_k)$, inputs are $(x_1, \ldots, x_k)$, outputs are (y_1, y_k), m, n = 1, 2, 3,…., k, and the intercept is C. Accordingly, SVM is provided in Eq. (14):

$$SVM = \sum_{m=1}^{k}\frac{1}{2}\,\beta m - \sum_{m,n=1}^{k}ymynC(xm,xn)\beta m\beta n \qquad (14)$$

This research uses SVM to separate data into two categories of fish: edible and poisonous fish.

3) NNs

The third ML technique is NN, the mathematical model of which uses learning algorithms inspired by the human brain to solve complex problems. Difficult tasks and problems for humans are significantly easy to solve using NNs. In general, NNs are composite of many nodes with multi-connections. A perceptron is an NN unit performing specific computations to reveal features in the input data. It takes a set of features and their targets as inputs, and it attempts to find how classes can be separated [26]. NNs consist of input, hidden, and output layers. Layers in NN comprise any number of neurons. However, the number of neurons in the output often matches the number of categories being surmised. Neuron connections in NN are weighted. Each neuron in NN have an activation function that transforms a neuron's input into a coveted output [27]. In NNs, learning is constraining of weights generating the closest outputs to the desired results. Learning is attained by optimizing the entire errors with regard to weights. The simplified structure of NNs provides leverage in ML. A supervised learning algorithm can learn based on well-represented input data to guess results for unforeseen data. When choosing a learning algorithm, several aspects must be considered: small memory requirements, speed and stability, and time saving [28]. Activation function A_n (I, W) of NN is the summation of input I_m multiplied with weight W_{mn}, as shown in Eq. (15). The sigmoid activation function of NN with the output function O_m (I, W) is given in Eq. (16).

$$A_n = \sum_{m=0}^{k}(I_m, W_{mn}) \qquad (15)$$

$$O_m = \frac{1}{1+e^{-A_n(I,W)}} \qquad (16)$$

D. **Performance Evaluation**

Performance evaluation is the process of measuring model quality. Various evaluation methods should be used to evaluate our model. Evaluation metrics can be used to assess the quality of the ML model. Several types of evaluation metrics are available to evaluate the ML model. These metrics provide accurate information of model

performance and help in determining the best model when several models are used. The most important evaluation metric is the confusion matrix. Confusion matrix is based on the number of true and false predictions in each class. Fig. (3) presents an example of a confusion matrix.

Fig. 3 Confusion Matrix

The performance of the three ML techniques can be evaluated in terms of precision, sensitivity, F1-score, and accuracy using Eqs. (17) to (20), respectively [29]:
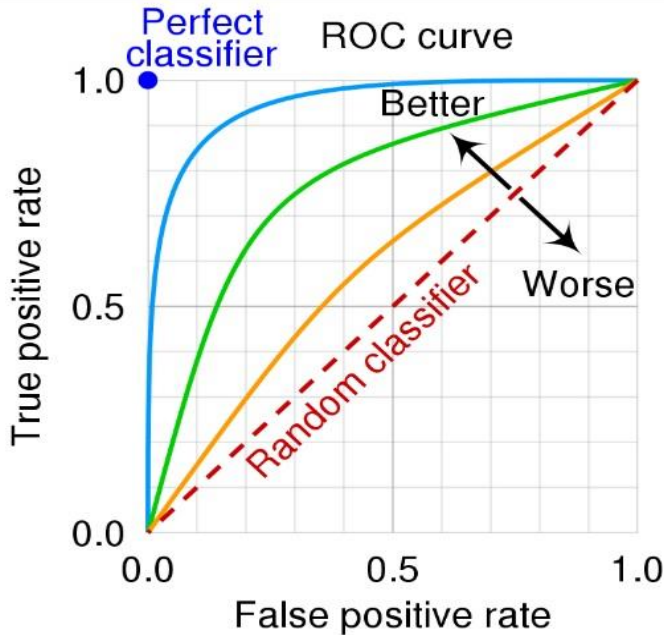
$$PR = \frac{TP}{TP+FP}, \qquad (17)$$

$$SN = \frac{TP}{TP+FN}, \qquad (18)$$

$$\text{F1-Score} = 2* \frac{SN*PR}{SN+PR}, \qquad (19)$$

$$Ac = \frac{TP+TN}{TP+TN+FP+FN}. \qquad (20)$$

The other method to evaluate a model or algorithm is the receiver operating characteristic (ROC) curve. The ROC curve explains the ratio between the true

positive rate against the false positive at different classification thresholds. ROC can be plotted using two axes (i.e., x-, y-axes), which refer to the false and true positive rates, respectively. Fig. (4) shows an ROC example [30].

Fig.4 Receiver Operating Characteristic Curve

IV.　RESULTS

This study was implemented using a CPU with Intel (R) Core (TM) i5-2430M processor CPU at 2.40 GHz, with 4 GB RAM, x64-based processor, MATLAB R2020a, and Windows 10. A total of 300colored fish images (i.e., edible and poisonous) were used. Colored fish images were collected from a significant database [10].

A. Training

This study used a 300-input fish image data set and 3 ML techniques (i.e., K-NN, SVM, and NNs). For all techniques, 210 and 90 samples were used for training and testing, respectively. For NNs, 12, 8, and 2 neurons were used for the input, hidden, and output layers, respectively.

Fish-Inputs and Fish-Targets matrices represent the input and target layers, respectively, for NN, and the Fish-Inputs matrix represents the relevant

features extracted from the fish images. A total of 300-input fish images data set is obtained: 150 images each for edible and poisonous fish. The size of the input features was 12, which exemplifies the combination of features extracted from each sample, and the size of output was 2, which represents edible or poisonous fish. Binary exemplification was likewise uses [1 0]; [0 1]. Table 1 presents the scaled conjugate gradient (SCG) backpropagation learning algorithm parameters. The organization of data set is as follows: 210 and 90 samples for training and testing, respectively.

TABLE 1 DEFAULT PARAMETERS OF THE SCALED CONJUGATE GRADIENT BACKPROPAGATION LEARNING ALGORITHM

| Types of fish	Encoding for target matrix
Edible fish	10
Poisonous fish	01

B. Testing

A total of 90 fish samples (i.e., edible/poisonous) were used for testing in the three ML techniques, selected from a significant database [10]. Prediction results of the number of correct and incorrect states are explained using a confusion matrix, while the performance of the presented work can be explained using the ROC curve. Fig. (5)
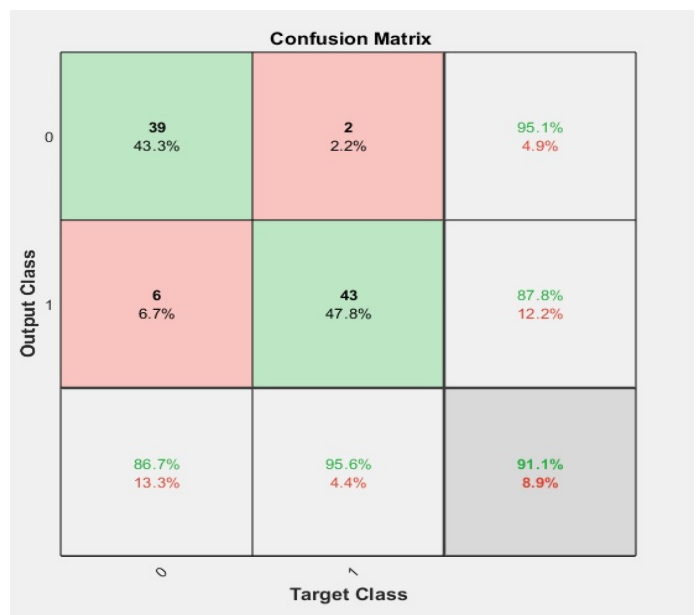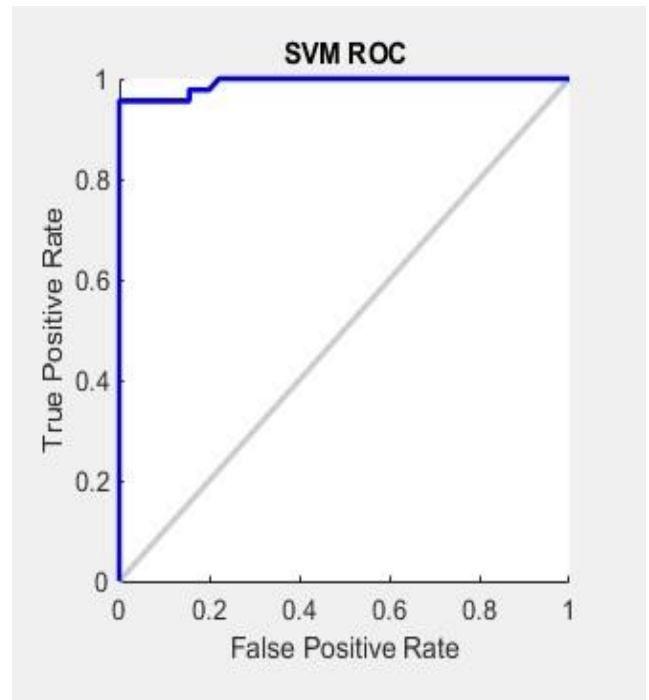
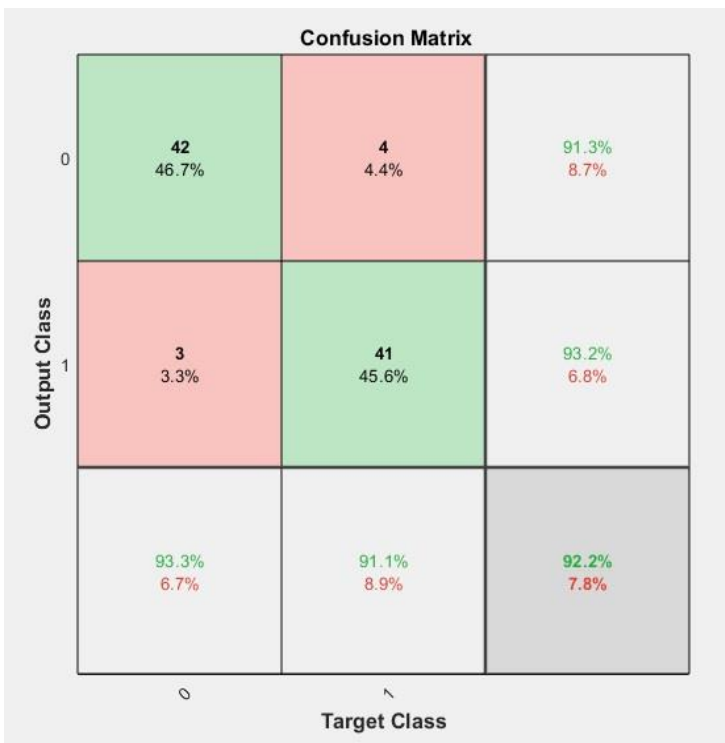shows the confusion matrix and ROC for the three ML techniques.
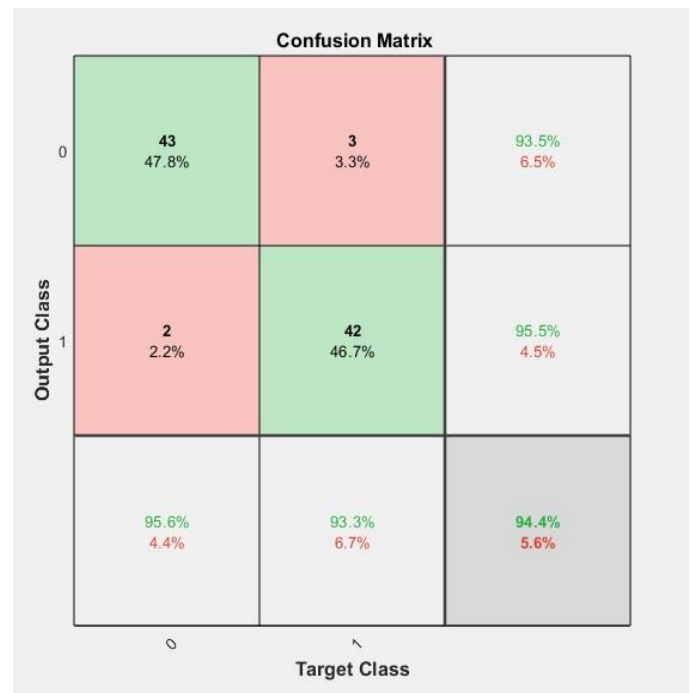
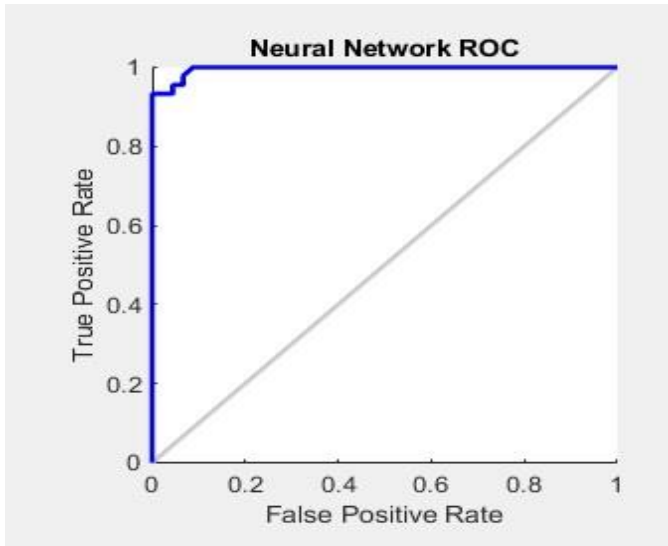(a)   Confusion matrix of the K-NN techniq
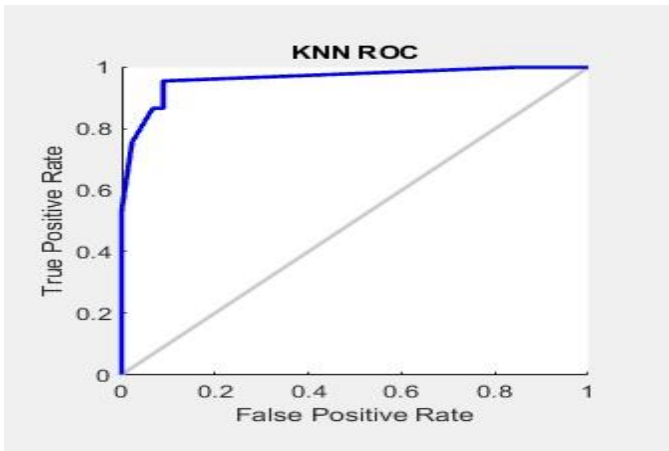
(b)   ROC of the K-NN technique

(d) ROC of the SVM technique

(c) Confusion matrix of the SVM technique

(d) Confusion matrix of the NN technique

(d) ROC of the NN technique
(e) **Fig.5** Confusion Matrices and ROC Curves for the Three ML Techniques

TABLE II EVALUATION SCORES OF THE ML TECHNIQUES

Nos.	Classifiers	Precision	Sensitivity	F1-Score	Accuracy
1	K-NN	0.9412	0.8700	0.9100	0.9111
2	SVM	0.91300	0.9333	0.9231	0.9222
3	NN	0.9500	0.9650	0.9500	0.9444

Table 2 shows that the NN method achieved the highest results in terms of precision, sensitivity, F1 score, and accuracy. Fig. (6) presents the evaluation of the presented work in terms of precision, sensitivity, F1-score, and accuracy for the three ML techniques.
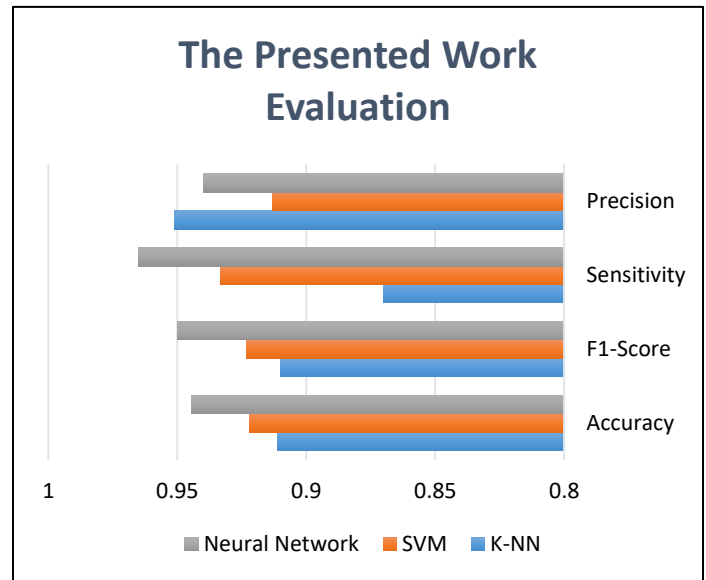

Fig. 6 Evaluation of the Presented Work

TPR explains the number of true positive results among all positive samples obtainable during testing. FPR explains the number of wrong positive results among all negative samples obtainable during testing.

As shown in Fig (5), the confusion matrix presents a comparison between the actual and predicted results of the edible/poisonous fish classes for the three ML methods. For the K-NN method, (a) shows the misclassification for 8 out of the 90 images. For the SVM method, (b) shows the misclassification for 7 out of the 90 images. For the NN method, (c) shows the misclassification for 5 out of the 90 images. For the K-NN method, the accuracy is 100% for six fish species, but misclassifications appeared in four species: barbus and *Tenualosa ilisha* (edible species) and *Lagocephalus lunaris* and *Himantura uarnak* (poisonous species). For the SVM method, the accuracy is 100% for seven fish species, but misclassifications appeared in three species: *Dentex macrophthalmus* and *Tenualosa ilisha* (edible species) and *Himantura uarnak* (poisonous species). For the NN method, the accuracy is 100% for eight fish species, but misclassifications appeared in two species: *Tenualosa ilisha* (edible species) and *Lagocephalus lunaris* (poisonous species). The evaluation scores of the presented work are provided in Tables 2.

The accuracy of the three ML techniques is shown in Fig. (7).
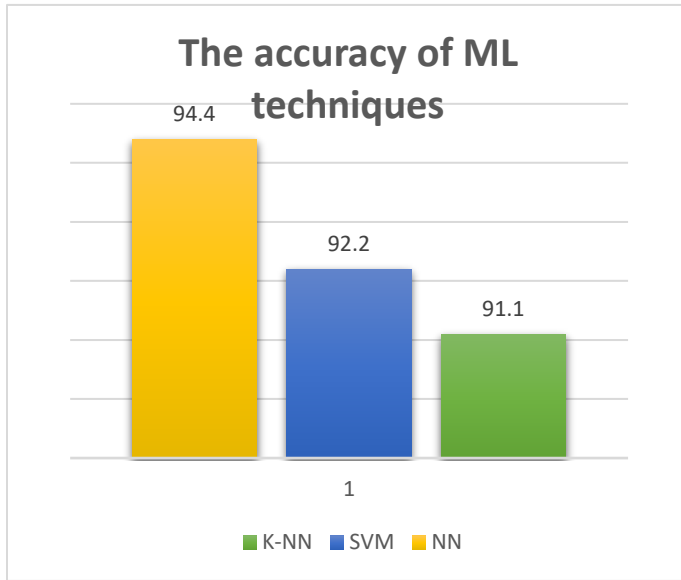
Fig.7 Accuracy Rate of the Three ML Techniques

Proposed ML approach	K-NN	91.1%
	SVM	92.2%
	NN	94.4%

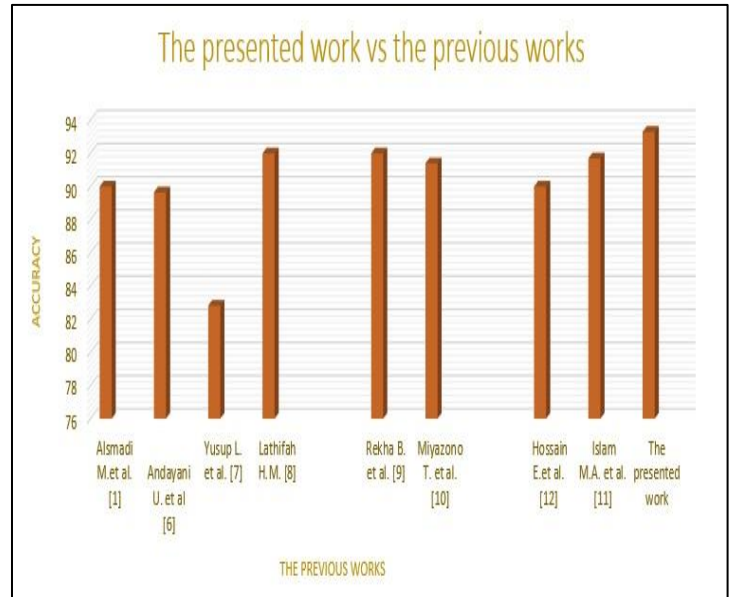The results of this research are compared with those of previous research. Table 3 presents the results. Fig. (8) shows a comparison between the presented and previous research.

TABLE III COMPARISON BETWEEN THE PRESENTED AND PREVIOUS STUDIES.

References	Methods	Accuracy
Alsmadi et.al. [1]	CNN	85.26%
Alsmadi M. et al. [11]	BP algorithm and MA-B classifier	82.25% and 90% respectively.
Tamou A.B. et al. [12]	YOLO deep learning Algorithm	82.82%.
Lathifah H. M. [13]	YOLO algorithm	92%,
Alaba S.Y. et al. [14]	MobileNetv3-large and VGG16 backbone networks	79.7%.
Rekha B. et al. [15]	CNN	90% - 92%
Islam M. A. et al. [16]	pyramid-histogram-of-visual-words with SVM	91.7%
Hossain E. et al. [17]	SVM	90%.

Fig. 8 Comparison between the Presented and Previous Research.

Despite the excellent results, there are some expected limitations in this study. The current research works on images from a data set not from underwater videos in natural environments. Moreover, the images were not captured from local rivers, seas, or local markets.

V.   CONCLUSION

This study underscores the importance of having a system for automatically identifying edible and poisonous fish. This approach provides a framework that can accurately classify fish into edible and poisonous ones based on ML. A total of 300 fish images with different sizes, shapes, and colors were selected from 20 fish species. Hybrid features (second order statistical features, high order statistical features) were extracted and three ML techniques (K-NN, SVM, and NNs) were used to classify fish into edible/poisonous. The proposed approach is evaluated in terms of precision, sensitivity, F1-score, and accuracy. The results are promising for the three ML techniques. NNs showed

the highest performance. The accuracies of the ML techniques are 91.1%, 92.2%, and 94.4% for K-NN, SVM, and NNs, respectively. This approach is invaluable for taxonomists, data scientists, and those interested in livestock. Future research will focus on increasing the number of fish samples and improving the proposed system to identify additional fish types (e.g., invasive, omnivorous, and ferocious fish) by using additional fish attributes (e.g., fish appearance and behavioral attributes).

Acknowledgments

REFERENCES

[1] A. Ben Tamou, A. Benzinou, and K. Nasreddine, "Live Fish Species Classification in Underwater Images by Using Convolutional Neural Networks Based on Incremental Learning with Knowledge Distillation Loss," *Mach. Learn. Knowl. Extr.*, vol. 4, no. 3, pp. 753–767, 2022. https://doi.org/10.3390/make4030036.

[2] M. Paraschiv *et al.*, "Classification of Underwater Fish Images and Videos Via Very Small Convolutional Neural Networks," *J. Mar. Sci. Eng.*, vol. 10, no. 6, pp. 1-21, 2022. https://doi.org/10.3390/jmse10060736.

[3] S. Franceschini, E. Gandola, M. Martinoli, L. Tancioni, and M. Scardi, "Cascaded Neural Networks Improving Fish Species Prediction Accuracy: The Role of The Biotic Information," *Sci. Rep.*, vol. 8, no. 1, pp.1-12, 2018. 10.1038/s41598-018-22761-4.

[4] Y.-C. Lu, C. Tung, and Y.-F. Kuo, "Identifying the Species of Harvested Tuna and Billfish Using Deep Convolutional Neural Networks," *ICES J. Mar. Sci.*, vol. 77, no. 4, pp. 1318–1329, 2020. https://doi:10.1093/icesjms/fsz089.

[5] A. D. J. van Dijk, G. Kootstra, W. Kruijer, and D. de Ridder, "Machine Learning in Plant Science and Plant Breeding," *iScience*, vol.24, no.101890, pp.1-12, 2021. https://doi.org/10.1016/j.isci.2020.101890.

[6] B. Wu *et al.*, "An Ameliorated Deep Dense Convolutional Neural Network for Accurate Recognition of Casting Defects in X-Ray Images," *Knowl. Based Syst.*, vol. 226, no. 107096, pp.1-24, 2021. https://doi.org/10.1016/j.knosys.2021.107096.

[7] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 160, pp. 1-21, 2021. https://doi.org/10.1007/s42979-021-00592-x.

[8] Y. Mohamadou, A. Halidou, and P. T. Kapen, "A review of Mathematical Modeling, Artificial Intelligence and Datasets Used in The Study, Prediction and Management of COVID-19," *Appl. Intell.*, vol. 50, no. 11, pp. 3913–3925, 2020. https://doi.org/10.1007/s10489-020-01770-9.

[9] I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity Data Science: An Overview from Machine Learning Perspective," *J. Big Data*, vol. 7, no. 41, pp.1-29, 2020. https://doi.org/10.1186/s40537-020-00318-5.

[10] https://www.fishbase.in/search.php.

[11] M. K. Alsmadi, M. Tayfour, R. A. Alkhasawneh, U. Badawi, I. Almarashdeh, and F. Haddad, "Robust Features Extraction for General Fish Classification," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 9, no. 6, pp.5192-5204, 2019. http://doi.org/10.11591/ijece.v9i6.pp5192-5204.

[12] I. M. Yusup, M. Iqbal, and I. Jaya, "Real-Time Reef Fishes Identification Using Deep Learning," *IOP Conference Series: Earth and Environmental Science*, 2020. 10.1088/1755-1315/429/1/012046.

[13] H. M. Lathifah, L. Novamizanti, and S. Rizal, "Fast and Accurate Fish Classification from Underwater Video Using You Only Look Once," *IOP Conference Series: Materials Science and Engineering*, 2020. https://doi:10.1088/1757-899X/982/1/012003.

[14] S. Y. Alaba *et al.*, "Class-Aware Fish Species Recognition Using Deep Learning for An Imbalanced Dataset," *Sensors (Basel)*, vol. 22, no. 8268, pp. 1-18, 2022. https://doi.org/10.3390/s22218268.

[15] B. S. Rekha, G. N. Srinivasan, S. K. Reddy, D. Kakwani, and N. Bhattad, "Fish Detection and Classification Using Convolutional Neural Networks," in *Computational Vision and Bio-Inspired Computing*, Cham: Springer International Publishing, 2020, https://doi.org/10.1007/978-3-030-37218-7_128.

[16] M. A. Islam, M. R. Howlader, U. Habiba, R. H. Faisal, and M. M. Rahman, "Indigenous Fish Classification of Bangladesh Using Hybrid Features with SVM Classifier," in 2019 *International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, 2019. Doi:10.1109/IC4ME247184.2019.9036679.

[17] E. Hossain, S. M. S. Alam, A. A. Ali, and M. A. Amin, "Fish Activity Tracking and Species Identification in Underwater Video," in 2016 *5th International Conference on Informatics, Electronics and Vision (ICIEV)*, 2016. 10.1109/ICIEV.2016.7760189.

[18] S. khudayer Jadwa, "Wiener Filter Based Medical Image De-noising," *Int. J. Sci. Eng. Appl.*, vol. 7, no. 9, pp. 318–323, 2018. 10.7753/IJSEA0709.1014.

[19] R. Josphineleela et. al., "Feature Extraction Techniques in Medical Imaging: A Systematic Review", *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11 no. 5, pp. 23-29, 2023. DOI: https://doi.org/10.17762/ijritcc.v11i5.6521.

[20] I. M. Hassoon, S. A. Qassir, M. Riyadh," PDCNN: FRAMEWORK for Potato Diseases Classification Based on Feed Foreword Neural Network", *Baghdad Science Journal*, Vol. 18 No.2, pp. 1012- 1019, 2021. DOI: http://dx.doi.org/10.21123/bsj.2021.18.2(Suppl.).1012

[21] Ş. Öztürk and B. Akdemir, "Application of Feature Extraction and Classification Methods for Histopathological Image using GLCM, LBP, LBGLCM, GLRLM and SFTA," *Procedia Comput. Sci.*, vol. 132, pp. 40–46, 2018. https://doi.org/10.1016/j.procs.2018.05.057.

[22] H. Saadatfar, S. Khosravi, J. H. Joloudari, A. Mosavi, and S. Shamshirband, "A New K-Nearest Neighbors' Classifier for Big Data Based on Efficient Data Pruning," *Mathematics*, vol. 8, no. 286, pp.1-12, 2020. https://doi.org/10.3390/math8020286.

[23] R. S. A. Daulay, S. Efendi and Suherman, " Review of Literature on Improving the KNN Algorithm

", Transactions on Engineering and Computing Sciences, vol. 11 No. 3, pp. 63-72, 2023. DOI: https://doi.org/10.14738/tecs.113.14768.

[24] S.O. Ogunlana, O. Olabode, S.A. A. Oluwadare , G. B. Iwasokun, "Fish Classification Using Support Vector Machine", *Afr J. of Comp & ICTs.*, vol.8, no.2, pp.75-82, 2015.

[25] M.M Taya," Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions", *Computers*, *vol. 12*, no. 5, 2023. https://doi.org/10.3390/computers12050091

[26] A. Ratku and D. Neumann, "Derivatives of Feed-Forward Neural Networks and Their Application in Real-Time Market Risk Management," *OR Spectr.*, vol. 44, no. 3, pp. 947–965, 2022. https://doi.org/10.1007/s00291-022-00672-1

[27] I. H. Sarker, "Deep cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective," *SN Comput. Sci.*, vol. 2, no. 3, pp.1-16, 2021. https://doi.org/10.1007/s42979-021-00535-6.

[28] K. M. Z. Othman and A. M. Salih, "Scaled Conjugate Gradient ANN for Industrial Sensors Calibration," *Bull. Electr. Eng. Inform.*, vol. 10, no. 2, pp. 680–688, 2021. 10.11591/eei.v10i2.2738.

[29] Shaveta, "A review on machine learning", *International Journal of Science and Research Archive*, vol. 09, no.01, pp. 281–285, 2023. DOI: https://doi.org/10.30574/ijsra.2023.9.1.0410.

[30] H. Han, " The Utility of Receiver Operating Characteristic Curve in Educational Assessment: Performance Prediction", *Mathematics*, vol.10, no. 1493, pp. 1-11, 2022. https://doi.org/10.3390/math10091493.