Fish Species Classification using Optimized Deep Learning Model

J.M. Jini Mol[1]
Research Scholar, Department of Computer Science, Malankara Catholic College Mariagiri
Affiliated to Manonmaniam Sundaranar University
Tirunelveli, Tamilnadu, India

Dr. S. Albin Jose[2]
Research Supervisor, Department of Computer Science, Malankara Catholic College Mariagiri
Affiliated to Manonmaniam Sundaranar University
Tirunelveli, Tamilnadu, India

***Abstract*—Classification of fish species in aquatic pictures is a growing field of research for researchers and image processing experts. Classification of fish species in aquatic images is critical for fish analytical purposes, such as ecological auditing balance, observing fish populations, and saving threatened animals. However, ocean water scattering and absorption of light result in dim and low contrast pictures, making fish classification laborious and challenging. This paper presents an efficient scheme of fish classification, which helps the biologist understand varieties of fish and their surroundings. This proposed system used an improved deep learning-based auto encoder decoder method for fish classification. Optimal feature selection is a major issue with deep learning models generally. To solve this problem efficiently, an enhanced grey wolf optimization technique (EGWO) has been introduced in this study. The accuracy of the classification system for aquatic fish species depends on the essential texture features. Accordingly, in this study, the proposed EGWO has selected the most optimal texture features from the features extracted by the auto encoder. Finally, to prove the efficacy of the proposed method, it is compared to existing deep learning models such as AlexNet, Res Net, VGG Net, and CNN. The proposed method is analysed by varying iterations, batches, and fully connected layers. The analysis of performance criteria such as accuracy, sensitivity, specificity, precision, and F1 score reveals that AED-EGWO gives superior performance.**

Keywords*—*Fish species classification; deep learning; GW optimization; auto encoder decoder; feature selection

I. Introduction

Object classification is an important area of research for underwater environments. To perform this, a high-resolution camera is used to scatter light and its absorption nature underwater [1]. Numerous researchers are interested in analyzing the health status of aquatic organisms, specifically the population and distribution of fish species [2]. Warming of the oceans will weaken aquatic life by increasing the pressure on fish species [3]. Accordingly, a cost-effective approach must be designed for underwater fish species analysis. In the past, fish species were classified by a laborious process involving the capture of fish or visual surveys conducted by deep-sea divers. Low contrast in the aquatic environment leads in very blurry pictures [4]. Due to the low quality of the images captured underwater, several minute features are lost. This will certainly impact the performance of the underwater image analysis system.

With the advent of powerful graphical processing units (GPU) and massive amounts of data, deep learning algorithms have become popular in classification and pattern reorganization [5][6]. This study's primary objective is to develop the deep learning model in order to create a completely automated system for classifying fish species. However, the presence of noise in underwater images limits the deep learning models training capacity. Additionally, it makes deep learning models more computationally demanding. This study employs the popular deep learning architecture auto encoder-decoder to obtain texture features from underwater images. Feature selection and hyper parameter (learning rate selection, weight updating process, and others) tuning is the most challenging aspect of building deep learning models.

GWO is an evolutionary optimization technique based on grey wolves' hunting mechanism and leadership hierarchy. Comparing Genetic Algorithms (GA) and particle swarm optimization (PSP), numerous studies have demonstrated that the performance of the GWO optimization method is superior. Unfortunately, traditional GWO requires more iterations to determine the optimal value when the data size and image noise rise. So in this research, the existing GWO algorithm has been enhanced to discover the optimal features using the newly introduced EGWO algorithm with fewer iterations.

Finally, three types of experiments have been conducted to prove the efficiency of the proposed fish species classification system. First, the fish species classification efficiency has been evaluated with the most essential parameters, such as accuracy, recall, specificity, precision, and F1-Score. Secondly, the training and validation efficiency of the proposed AED-based deep learning model has been evaluated. Finally, the computational efficiency of the proposed method has been evaluated. Through these three types of experimental analysis, a comparative study is conducted between the proposed method and existing deep learning algorithms such as AlexNet, ResNet, VGGNet, and CNN. On the basis of these three types of experimental observations, it has been proven that the proposed methodology has excellent training efficiency, high accuracy, and low computing overhead. This study's significant contributions are summarized below.

1) First, the R, G, and B channels of the underwater images are normalized to enhance the object visibility.

2) Second, the fish morphology localization method has been implemented to eliminate objects that do not have an impact on classification.

3) Next, an auto encoder-decoder deep learning model is used to extract underwater images' texture and color features.

4) Finally, the EGWO approach is introduced to improve feature selection efficiency.

Highlights of the proposed methodology are as follows: In the second section of this study, the previously developed computer-based classification system for fish species is examined in greater detail. The proposed auto encoder-decoder and enhanced grey wolf optimization are discussed in detail in the third section. In the fourth section, the experimental analysis of the proposed fish species classification system has been comparatively analyzed with existing deep learning algorithms. Finally, the proposed method is concluded.

II. Literature Review

Classification of underwater images is a challenging task. Manual classification methods demand considerable time and effort. In underwater classification, image size, color, texture, inter-class similarity, and intra-class dissimilarity pose the greatest obstacle. Recently, researchers have developed several machine learning and deep learning methods for classifying underwater fish species [7]. This section reviews the existing fish species classification methods.

The classification is mainly carried out on dead sections depending on shape and texture [8, 9]. The fish's length, width, and thickness are identified using laser light [10]. Classification of fish species is very difficult due to variation in luminosity, background, turbidity of water. Moreover, the similarity of shape and color of various fish species is very difficult to classify.

The fishes are categorized depending on the shape and texture patterns in unconstrained nature [11]. Classification of fish species depends on the biomass content [12]. Classification and identification of fish species depend on morphology, texture and geometry [13]. Fish species identification is done using live fish in the open sea [14].

Recognition of fish species can be done from low resolution images [15]. Combining sparse representation and PCA for classification will provide an accuracy of 82.8%.Gaussion mixture combined with support vector machine will provide a recognition rate of 78%.Few conventional methods are done at a constrained manner. Classification of fish species using shape and color of dead fish sections in organized background.

Grouping of fish species is done based on texture and physical features in unrestrained environments [16] [17]. These techniques will provide good results in fish species with the big difference in texture. The biometric approach [18] takes fish species from various distances and different illuminations. Weber's local descriptor provides classification by the Adaboost classifier.

Classification techniques depending on a mixture of various feature extracting approaches and clustering algorithms with input classifiers are less time-consuming and cheaper [19]. Regions with Convolutional Neural Networks (RCNN) provides good accuracy of 82% in fish species classification [20]. PCA method combined with binary hashing function along with SVM classifier is used for fish species recognition and provides an accuracy of 98% [21].

AlexNet model is pre-trained to perform fish species classification [22]. The CNN-based model and transfer learning use Res Net-152, and SVM classifier provides an accuracy of 95% [23]. Classification of fish species by combination of YOLO deep neural network with Gaussian mixture provides good accuracies in two data sets. VGGNet along with convolutional layers, provides fish species classification [24].

A convolutional neural network is built to rapidly classify the behaviorist of fishes. Few pressure settings are completed in the research center, the fishes' behavior positions are acknowledged, and the model database is recognized [25]. To order the fishes, convolutional neural systems are used. The Faster Regional Convolutional Neural Network strategy is used to eliminate the high spot of images [26]. A Deep CNN Fondest is used for fish recognition, and localization and grouping are done using visual data got from cameras [27].

A Mask R-CNN, together with GOTURN is used for real-time applications of fish recognition and classification [28]. A fish grouping using transfer learning and Matlab is used as the primary step of undertaking the issue. FishNet is an adjustment form of AlexNet to categorize different varieties of fishes [29]. An automatic fish classification based on sonar videos classify fishes based on shape and Movement [30].

The limitations of these surveyed publications are summarized below.

- Due to the poor contrast of underwater photographs, a number of existing methods are semi-automated.
- In many existing systems, the classification algorithm is trained on dead underwater images. Clearly, these techniques cannot be utilized to classify fish species in real-time.
- The accuracy of the fish species classification system depends on the color and texture features of the underwater images. However, no specific optimization approaches have been developed in existing methods for selecting the appropriate color and texture features from underwater images.

III. Proposed Methodology

This section discusses the process flow and methodologies of the proposed fish species classification system. The entire process flow of the proposed methodology is illustrated in Fig. 1.

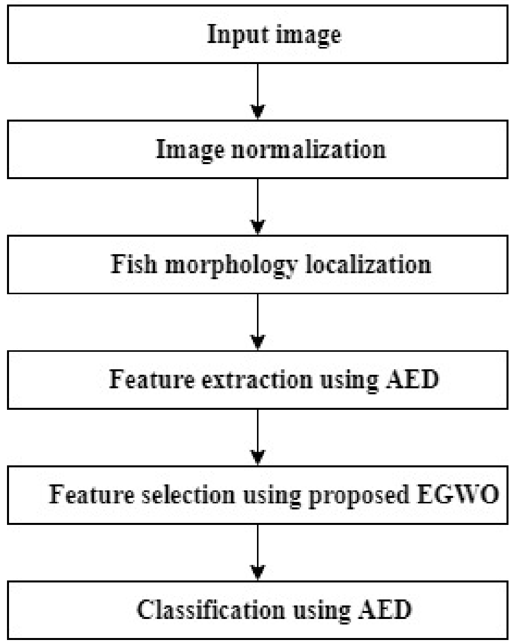

Fig. 1. Overall Process Flow of the Proposed System.

A. Dataset

In this study, the dataset is downloaded from GitHub (https://github.com/primepake/Fishes_classification) and internet sources. There are 1000 images of aquatic fish in this data set. Five type's fish species are included in the data set: aulonocara fire, discus, flame fish, king fish, and molly. Each fish species in the dataset is partitioned into a 4:1 ratio for training and validation purposes. Table I demonstrates the partitioning information for the dataset. Accordingly, 800 images are obtained for the training phase, and 200 are obtained for the testing phase. To boost training efficiency, fish images are artificially augmented using the following image processing techniques: image rotation (90, 180, and 360 degrees), horizontal and vertical flipping, and zooming.

B. Image Normalization

Table I demonstrates very clearly that underwater images contain a great deal of noise, particularly undesired dark regions and water backgrounds around the images. These undesirable dark regions impose an additional computational overhead on the classification algorithms. Hence it necessitates the implementation of pre-processing techniques. Generally, all images captured by underwater cameras are RGB images. These images have different color combinations. The majority of images captured from the bottom of the water are quite dark, and the visibility of the objects within them is extremely poor. At the same time, images captured from the surface of the water are extremely bright. When training the model with such images, the accuracy will undoubtedly suffer. Due to the un-normalized nature of underwater images, the red, blue, and green channels should be normalized. Therefore, in this research, the three colour channels of the underwater images are normalized to reduce classification loss. The following equation is used to carry out the underwater image normalization procedure.

$$r(\alpha,\beta) = \frac{R(\alpha,\beta)}{R(\alpha,\beta)+G(\alpha,\beta)+B(\alpha,\beta)} \tag{1}$$

$$g(\alpha,\beta) = \frac{G(\alpha,\beta)}{R(\alpha,\beta)+G(\alpha,\beta)+B(\alpha,\beta)} \tag{2}$$

$$b(\alpha,\beta) = \frac{B(\alpha,\beta)}{R(\alpha,\beta)+G(\alpha,\beta)+B(\alpha,\beta)} \tag{3}$$

Where $\alpha\ and\ \beta$ indicates the dark and bright regions in the underwater images. $(r, g\ and\ b)$ specifies the red, green and blue colour channels of the underwater images. The range $r(\alpha,\beta)$, $g(\alpha,\beta)$, $b(\alpha,\beta)$ is from 0 to 255, the value of $r(\alpha,\beta)$, $g(\alpha,\beta)$, $b(\alpha,\beta)$ can vary from 0 to 1. Fig. 12(a) shows the underwater images before the image normalization. Fig. 12(b) shows the underwater images after image normalization.

C. Fish Morphology Localization

Fish morphology localization is a crucial element in fish classification. This will eliminate pixels that have no impact on fish classification. This study uses the Simple Linear Iterative Clustering (SLIC) approach to localize fish morphology from aquatic images. SLIC is the most used super pixels segmentation method, and its key benefits include it separates fish regions from aquatic images with little computational cost. SLIC method integrates image plan space and color dimensions to build consistent and realistic super pixels. In order to perform local clustering, the SLIC method performs clustering the pixel dimensions, which is the CIELAB color space (l*a*b*). Euclidean distances in l*a*b* are measured by the following formulas, which calculate the pixel variation in the images.

$$D_{LAB} = \sqrt{((L_i - L_j)^2 + (A_i - A_j)^2 + (B_i - B_j)^2}, \tag{4}$$

$$D_{xy} = \sqrt{((x_i - x_j)^2 + ((x_i - y_j)^2}, \tag{5}$$

$$D = D_{LAB} + \frac{m}{s} D_{xy} \tag{6}$$

L=Lightness.

A=Red/Green Values.

B=Blue/Yellow Values.

The SLIC approach begins with cluster sample centers that are uniformly separated. The centers are then moved to initialization places based on the gradient position with the lowest value. Gradient values are calculated by the following formula (7).

$$\boldsymbol{g(x,y)} = ||L(x+1,y) - L(x-1,y)||^2 + ||L(x,y+1) - L(x,y-1)||^2 \tag{7}$$

$L(x,y)$ is the pixel coordinates. According to the SLIC methodology, pixels in the neighbourhood of a large section will have the same labeling. The method then establishes relationships by relabeling disconnected portions with the labels of the closest neighboring cluster. Fig. 11(c) shows the proposed fish morphology localization procedure.

TABLE I. FISH DATASET DETAILS

Fish image	Total number of images	Training	Testing	Class Label
	200	140	60	King fish
	200	140	60	Discus
	200	140	60	Flame fish
	200	140	60	Molly
	200	140	60	Aulonocara fire

D. Auto Encoder Decoder

An auto encoder decoder is a deep learning method that uses unsupervised learning to conduct encoding and decoding [31][32]. Similar to an artificial neural network, it consists of input, hidden, and output layers [33]. Each layer has neurons, with the input and output layers having the same amount, but the hidden layer has fewer neurons than the input layer.

Fig. 2 depicts the architecture of auto encoder decoders. The pre-processed images from the fish dataset are supplied to the auto encoder for feature extraction. There are encoder and decoder sections here. Each encoder comprises a convolution with a filter bank, max pooling, and subsampling to generate the feature map. The encoder is composed of two convolution layers and an intermediate layer. Here, the convolution process of feature maps is not performed. Following batch normalisation, the convergence of local minima is enhanced. Additionally, the decoder comprises two layers of convolution. The convolutional auto encoder consists of output data Y that generates from input data x that is comparable to the original input data. When the time reaches infinity, the optimal value of the cost function will be reached. F () and F*() are the encode and decode functions. In this study, the nonlinear activation function Rectified Linear Unit (ReLU) is employed, which is represented by the following equation (8). If the function receives negative input, it returns 0, but for positive input, it returns the input value.

$$f(x) = max(0, x) \tag{8}$$

By optimising the weight and bias term, it is possible to minimise the error. Using the back propagation method, the weight is modified.

$$F(X) = \sigma(wX + B) \tag{9}$$

$$X = \sigma(w(F(X)) + B) \tag{10}$$

The first layer of encoder has an input X. The next layer has input data X^L , weight $W^L$and bias term B^L. So the above equations are changed as

$$F(X^{L+1}) = \sigma(W^L X^L + B^L) \tag{11}$$

$$X^{L+n+1} = \sigma\left(W^{n-L}\left(F(X^{n+L})\right) + B^{n-L}\right) \tag{12}$$

This auto encoder is trained for 200 epochs. The weights are optimized in order to find the local minima. Hidden layer takes output of the preceding layer. Throughout the training phase σ value was 4. The mean squared error decreased till 14 for 100 epochs and the performance during the training phase is 0.00058. The error rate is reduced smoothly without over fitting. The trained parameters will be provided to the next layer. Each and every auto encoder is subjected to max polling layer and the result is subsampled by a factor of 2.

$$y_{i,j} = maximum(x_{i,j}) \tag{13}$$

Where $x_{i,j}$ *and* $y_{i,j}$ denotes the i region of j feature-map which is the feature map of input and i neuron of j feature-map which is the feature map of output respectively. The number feature-map remain the same both at input and output. In each and every decoder up sampling is done for its feature-map by applying the indices of maximum pooling from the corresponding feature-map of encoder. As a result sparse feature-map is produced. Convolution of decoder filter bank and feature map is done to regenerate the input.

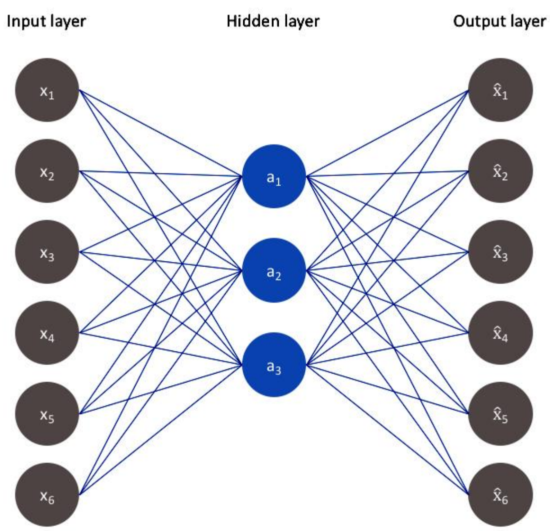

Fig. 2. Auto Encoder Decoder.

E. Traditional GWO Method

This algorithm is based on the searching and hunting behaviour of grey wolves. According to GWO, the fitness value has three parameters one is alpha (α), the other is beta (β), the third is the delta (δ). The hierarchical structure of grey wolves is illustrated in Fig. 3. The remaining solutions are called as omegas (ω). Three grey wolves will guide omegas in searching step. At time t=1, first iteration begins, at the time when a prey is found out. The omegas will encircle the prey with the help of alpha, beta and delta wolves. Three coefficients $(a^{\rightarrow}, c^{\rightarrow}, d^{\rightarrow})$will help in encircling process. They are.

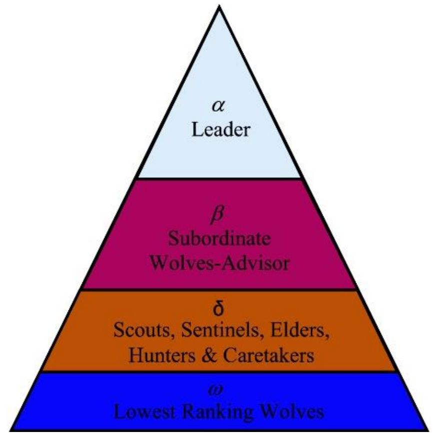

Fig. 3. Hierarchy of Grey Wolves.

$$d_{\alpha}^{\rightarrow} = \left| c_1^{\rightarrow}.x_{\alpha}^{\rightarrow} - x_{(t)}^{\rightarrow} \right| \tag{14}$$

$$d_{\beta}^{\rightarrow} = \left| c_2^{\rightarrow}.x_{\beta}^{\rightarrow} - x_{(t)}^{\rightarrow} \right| \tag{15}$$

$$d_{\delta}^{\rightarrow} = \left| c_3^{\rightarrow}.x_{\delta}^{\rightarrow} - x_{(t)}^{\rightarrow} \right| \tag{16}$$

Where t is the current iteration. $x_{1,}^{\rightarrow} x_{2,}^{\rightarrow} x_{3,}^{\rightarrow}$denote the position vector of alpha, beta and delta respectively.

$$x_{1,}^{\rightarrow} = x_{\alpha}^{\rightarrow} - a_1^{\rightarrow}.\, d_{\alpha}^{\rightarrow} \tag{17}$$

$$x_{2,}^{\rightarrow} = x_{\beta}^{\rightarrow} - a_2^{\rightarrow}.\, d_{\beta}^{\rightarrow} \tag{18}$$

$$x_{3,}^{\rightarrow} = x_{\delta}^{\rightarrow} - a_3^{\rightarrow}.\, d_{\delta}^{\rightarrow} \tag{19}$$

$$x_{(t)}^{\rightarrow} = \frac{x_{1,}^{\rightarrow} + x_{2,}^{\rightarrow} + x_{3,}^{\rightarrow}}{3} \tag{20}$$

The parameters $a^{\rightarrow}, c^{\rightarrow}$ are given by

$$a^{\rightarrow} = 2\alpha\, r_1^{\rightarrow} - \alpha \tag{21}$$

$$c^{\rightarrow} = 2\, r_2^{\rightarrow} \tag{22}$$

Where $r_1^{\rightarrow}$ *and* $r_2^{\rightarrow}$ are random numbers. The controlling parameter is α whch alters the value of $a^{\rightarrow}$. If the value of $a^{\rightarrow}$ is greater than 1 the grey wolves will move farther away from the prey, which denotes that the omega will run away, representing global optimization. If the value of $a^{\rightarrow}$ is lesser than 1,the omega will move towards the prey showing local optimization. The controlling parameter will decline from 2 to zero in a linear manner. This is given by

$$\alpha = 2\left(1 - \frac{it}{n}\right) \tag{23}$$

Where n is the maximum value of iteration number which is a cumulative iteration number.

Every grey wolf runs away or moves towards the prey with a suitable mean weight for an alpha, beta, and delta. On the start of the searching method, the weight of the alpha must be larger than the others ie, beta and delta. The hierarchy in weight must be such that the weight of alpha is higher than that of beta and delta. Similarly, the weight of beta is greater than that of delta. The alpha wolf is given greater importance than others. The alpha wolf is considered to be nearer to the prey. The alpha wolf governs the searching. Beta and delta have a less important role. If beta or delta finds the best position, it is transferred to the alpha wolf. In the searching process, the hypothesized prey is encircled, but in hunting, the real prey is encircled. The alpha is the nearest one to the prey than the beta. But the delta is farther away to beta. The omega wolf will alter and give their best positions to these dominants. Initially, the value of alpha is 1 and the value of beta and delta is zero. Finally, the dominants will encircle the prey, since they have the same weight. The weight of alpha has to be reduced, and the weight of beta and delta has to arise due to the cumulative iteration number. The position is updated as

$$\vec{x}_{(t+1)} = w_1\vec{x_1} + w_2\vec{x_2} + w_3\vec{x_3}$$

Where $w_1 + w_2 + w_3 = 1 \; such \; that \; w_1 \geq w_2 \geq w_3$ (24)

Algorithm 1 explain the process flow of traditional GWO. In the GWO algorithm, the initial population is generated by randomly. In this study, the population refers to the number of texture features. When initialising at random, there is a significant likelihood of irrelevant or redundant features. This raises the number of iterations of the GWO method, which increases the algorithm's running time and computational overhead. Therefore, the traditional GWO method must be enhanced in order to choose the desirable features from the fish data set with fewer iterations and in less time.

Algorithm 1

Step1: Random initialization of the population of grey wolves Xi
Step2: Initialise the values of α, β, δ.
Step3: Fitness of search agent is calculated
$\vec{x_a}$=search agent in first position
$\vec{x_\beta}$=search agent in second position
$\vec{x_\delta}$=search agent in third position

Step4: when (t< the number of iterations)
For each search agent
update the position of current agent using equation (24)
end for
Step5: Update the values of α, β, δ.
Step6: The fitness of all agents are computed.
Step7: Update $\vec{x_a}, \vec{x_\beta}, \vec{x_\delta}$
Step8: Increment the value of t
End while
Step9:Return the value of $\vec{x_a}$

F. Feature Selection using Enhanced GWO Method

Feature selection is a very important part of artificial intelligence based prediction models [33][34]. Fig. 4 shows the overall process flow of the proposed AED-EGWO. The proposed research AED extracts different types of colour and texture features from underwater images. When extracting features from underwater images using AED, redundant and irrelevant features are likely to be extracted. The formulas 25-33 determine the important texture features of this proposed method. The accuracy of classification algorithms is relies on effective feature selection techniques. The precision of deep learning models can be significantly enhanced by picking the most beneficial feature.

In addition, the computational burden of deep learning models is drastically lowered. The EGWO approach is used to eliminate irrelevant and redundant texture features from this classification system for fish species.

$$\text{Autocorrelation} = \sum_i \sum_j (i,j) p(i,j) \quad (25)$$

$$Contrast = \sum_{n=0}^{N_e-1} n^2 \sum_{n=i}^{N_e} \sum_{n=j}^{N_e} \{p(i,j)\} \quad (26)$$

$$Correlation = \frac{\sum_i \sum_j (i,j) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (27)$$

$$Energy = \sum_i \sum_j p(i,j) \quad (28)$$

$$Dissimilarity = \sum_i \sum_j |i-j| * p(i,j) \quad (29)$$

$$Entropy = \sum_i \sum_j p(i,j) \log(p(i,j)) \quad (30)$$

$$Homogeneity = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j) \quad (31)$$

$$Variance = \sum_i \sum_j (i-\mu)^2 p(i,j) \quad (32)$$

$$Cluster\ shade = \sum_i \sum_j (i+j-\mu_x-\mu_y)^2 p(i,j) \quad (33)$$

Here, an improved version of the GWO algorithm is proposed. Particularly in the population initialization phase, provide an intelligent initialization approach to achieve the optimal solution in early iterations. This intelligent initialization strategy accelerates the convergence of the algorithm. The key difference is that the population is now initialized using a correlation-based technique rather than at random method. The initial population is formed based on the correlation value, which decides whether a feature value is selected or not. The following equation represents the computation of the correlation for a feature f:

$$Cor_F = \frac{\sum (F_i - \bar{F})(C_i - \bar{C})}{\sqrt{\sum (F_i - \bar{F})^2 \sum (C_i - \bar{C})^2}} \quad (34)$$

Above equation, F indicates the texture features of underwater images. C represents the class values. $\bar{F}$ and $\bar{C}$ represents the mean values of features and classes, respectively.

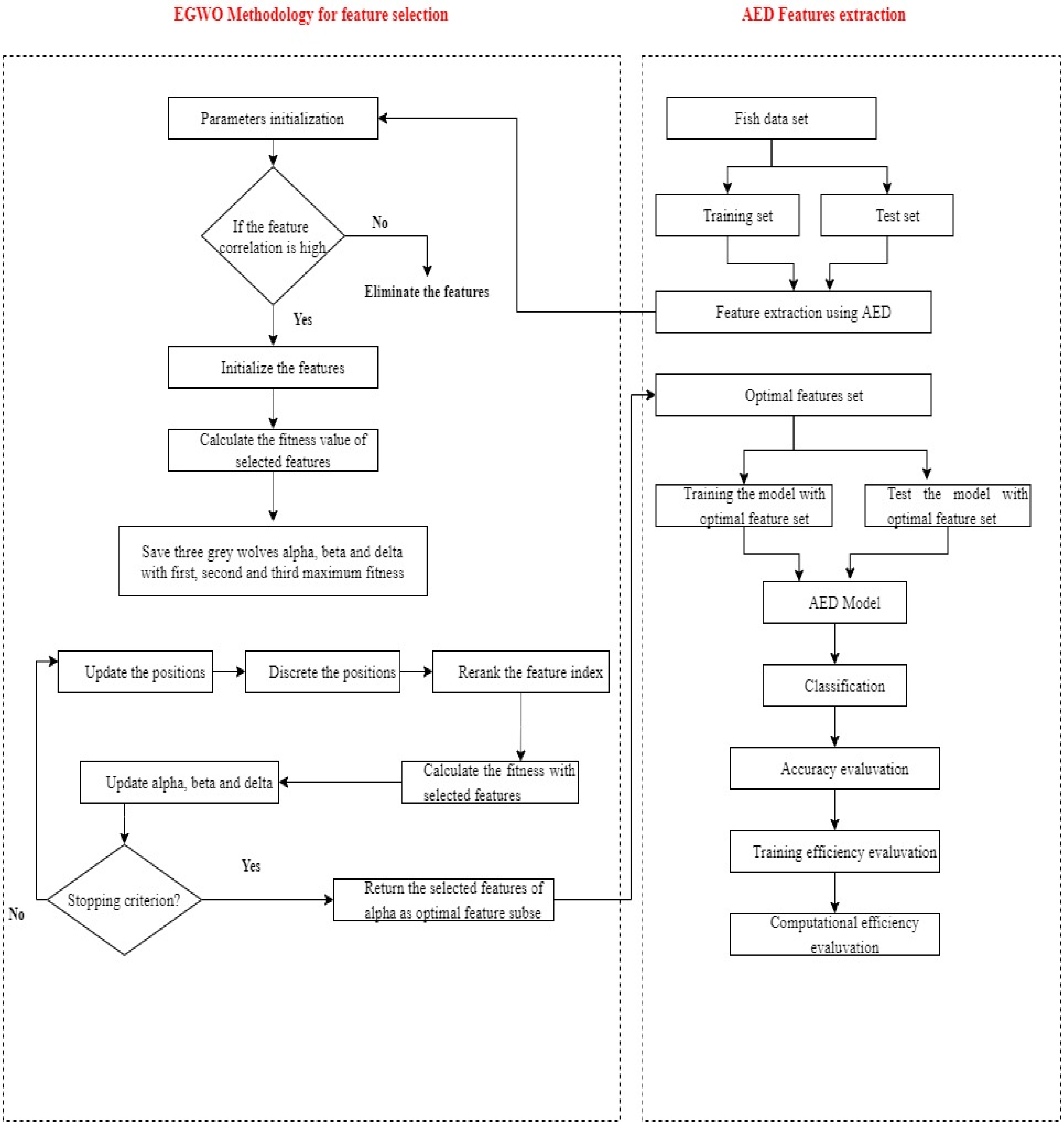

Fig. 4. Overall Process Flow of the Proposed AED-EGWO.

A feature with higher correlation values is crucial for the classification. Therefore, employing the following formula, the suggested method assures that features with high correlation values are included in the initial population. Then, based on the correlation values, the new population is initialized as follows.

$$P(i) = \begin{cases} 1, if\ rand < IG(i) \\ 0, if\ rand \geq IG(i) \end{cases} \quad (35)$$

$rand$ is a random number between 0 and 1, and $P(i)$ is the binary representation of the ith feature in the initial population. According to formula (35), the features values with high correlation values are initialized for the initial population. This allows the traditional GWO method to get optimal features in a minimal number of iterations. Algorithm 2 explain the proposed EGWO based texture feature selection.

Algorithm 2 proposed feature selection

Input: AED extracts texture features from fish training data set
$(T_1, T_2, T_3 \ldots\ldots T_n, T_c) T_n$ is the number of texture features and T_c is the target class.
Output: O_{tf} (Optimal texture features)
Step 1: **Begin**
Step 2 : **for** i=1 to n **do**
C=calculate the coefficient(T_n, T_c) using formula
End
Set the threshold level Ʈ=0.6
Step 3: if **Ʈ**<0.6 // means there is no significant correlation between T_n and T_c
Step 4: **for** I =1 to m **do**
r=Calculate the significant between (C,Ʈ)
if significant is high
Then
Add the features to ⇒ O_{tf}
End
End
Return O_{tf}
End

IV. Results and Discussion

A. System setup and Configuration

The software tools used to develop this fish species categorization system are Matlab 2018 and deep learning libraries. Additionally, Windows 10 is used as an operating system. To run deep learning libraries and design the implementation model for this fish species classification system, a graphics processing unit with 4GB NVIDIA 1650, an Intel 10th Gen Core i5 processor, 256GB SSD, and 16GB RAM is utilized.

B. Classification Model Performances Evaluation

Existing deep learning approaches such as AlexNet, ResNet, VGGNet, and CNN are compared to the proposed methodology. The proposed method is evaluated with different configurations of the deep learning model, including fully connected layers, iterations, and batches, with and without optimization. The test image and the ground truth image must be compared. Five common performance measures that are typically used for classification are used to the analysis the performance of the proposed classification system: accuracy, sensitivity, specificity, precision, and F1 score. The mentioned accuracy measures are depending on the following variables.

True positive fish species classification: If the proposed method correctly recognises and classifies fish species from underwater images, this classification is known as a true positive fish species classification. The variable TP specifies the classification of true positive fish species.

True negative fish species classification: If the proposed method correctly identifies and classifies non-fish species from underwater images, this classification is known as true negative fish species classification. The variable TN specifies classification of true negative fish species.

False positive fish species classification: If the proposed method mistakenly identifies and classifies non-fish species as fish species from underwater images, this is referred to as a false positive fish species classification. The variable FP specifies the classification of false fish species.

False negative fish species classification: False negative fish species classification happens when the proposed approach fails to recognise and classify non-fish species from underwater images. The variable FN specifies classification of false negative fish species.

The confusion matrices of the proposed and existing deep learning models are depicted in Fig. 5. According to Table I, 300 images are utilized to test the AED-EGWO approach. Accuracy, recall, specificity, precision, and F1-Score values are calculated based on the Confusion matrix, which are shown in Fig. 6 to 10.

In this study, the definition of accuracy is the correct classification of fish species from underwater images. To ensure the reliability of proposed model, it is essential to determine the proportion of true positives and true negatives among all of the instances that have been analyzed. Mathematically, accuracy is expressed as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{36}$$

According to the classification of fish species, sensitivity involves accurately identifying fish species. Sensitivity can be determined by examining the proportion of true positives. Sensitivity has the following mathematical expression:

$$\text{Sensitivity (Recall)} = \frac{TP}{TP+FN} \tag{37}$$

According to the classification of fish species, specificity determines the reliability of non-fish classification results. To measure it, calculate the proportion of genuine negatives. Specificity has the following mathematical expression:

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{38}$$

Precision is the ratio of the number of accurate fish classifications to the total number of positive fish predictions. Precision is calculated by the following formula.

$$Precision = \frac{TP}{TP+FP} \tag{39}$$

In an AI-based classification system, the F1-score is calculated using the overall precision and recall values. Formula for calculating the F1-score is as follows.

$$F1 - score = \frac{2(Recall \times Precision)}{Recall+Precision} \tag{40}$$

The confusion matrices of the proposed and existing deep learning models are depicted in Fig. 5. Accuracy, recall, specificity, precision, and F1-Score values are calculated based on the Confusion matrix, which are shown in Fig. 6 to 10.

Fig. 5. TP, TN, FP and FN Ratios of Proposed and Existing Methods.

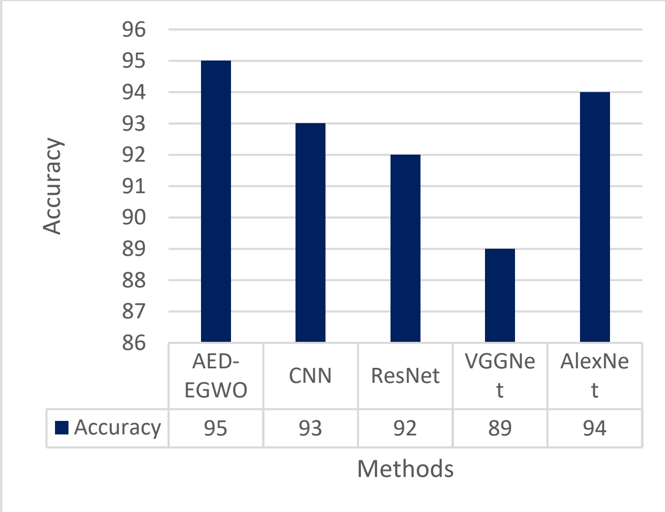

Fig. 6. Accuracy Comparison Results.

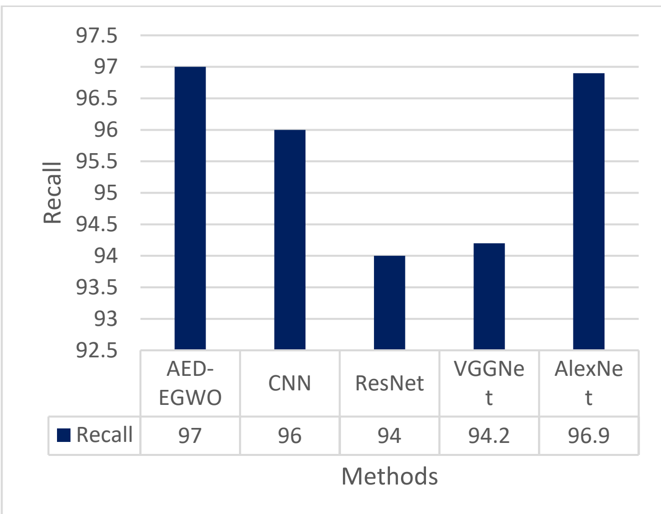

Fig. 7. Recall Comparison Results.

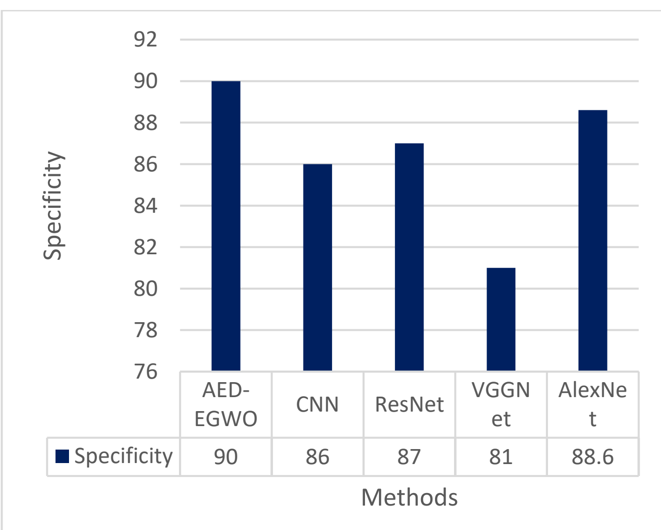

Fig. 8. Specificity Comparison Results.

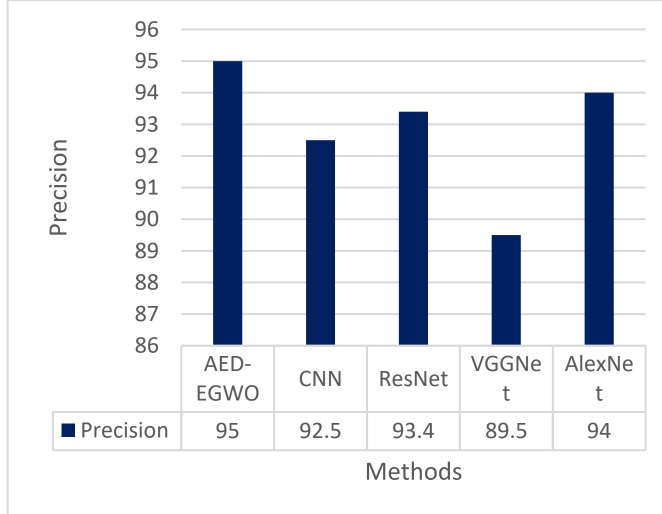

Fig. 9. Precision Comparison Results.

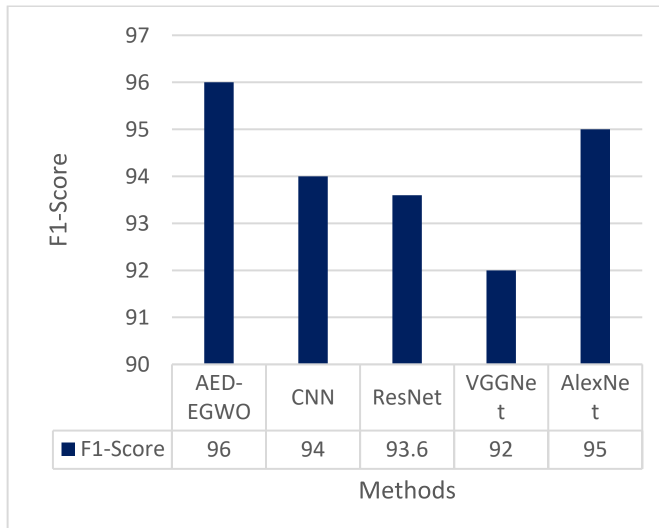

Fig. 10. F1-score Comparison Results.

The proposed methodology's performance is compared to traditional methods AlexNet, ResNet, VGGNet, and CNN. For a better analysis, the methods are experimentally implemented and their performance metrics are compared with the existing methods. The experimental results (accuracy, recall, specificity, precision and F1-Score) are presented in the Fig. 6 to 10. According to the experimental results, it is discovered that the proposed method is more accurate than the other state-of-the-art methods. Proposed fish classification system will have a greater number of true-positive fish pixel classifications; this will obviously improve the accuracy metrics. Following the proposed system, CNN and AlexNet have the highest accuracy (accuracy, recall, specificity, precision, and F1-Score). The ROC curve depicted in Fig. 10 has an AUC between 0.5 and 1.0, which indicates that it ranks a random positive sample higher than a random negative sample more than fifty percent of the time. Based on the results, it can be seen that the proposed methodology has fewer false positive fish pixels than other techniques. In addition, it is recognized that the number of true negative fish pixels is greater for the proposed method than for the traditional deep learning techniques. In addition, the experimental results proves suggested technique reduces the number of false negative fish pixels.

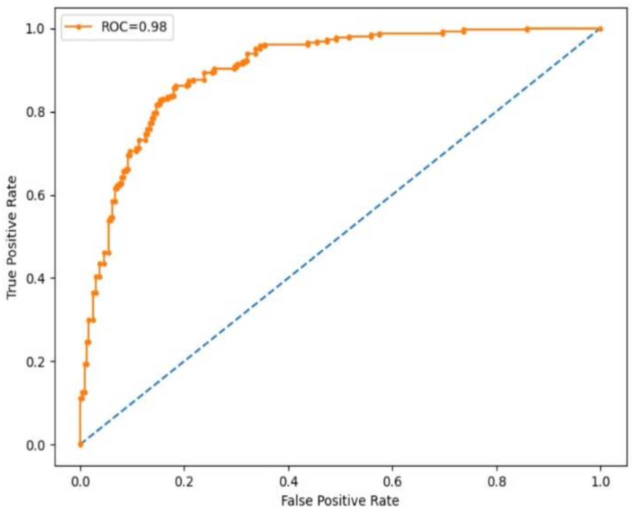

Fig. 11. ROC Curve.

C. Training efficiency analysis

This section examines the training efficiency of the proposed and existing approaches. For that, mean absolute error and mean squared error are two of the most important training efficiency evaluation metrics applied in this research. If the error values are minimal, therefore the training loss of the suggested technique is also limited.

1) Mean absolute error: Mean absolute error (MAE) is the average of the difference between actual and predicted values. It describes the variation between the predicted and the actual values. The following formula calculates MAE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \widehat{yi}| \tag{41}$$

y represents the actual class value, while $\hat{y}$ represents the outcome predicted by the proposed model.

2) Mean squared error: Mean squared error is the average of the squared differences between the actual and predicted values. The equation below is used to compute the MSE.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}|yi - \widehat{yi}|^2 \tag{42}$$

MAE and MSE provide positive integer values during training. If the value is near to zero, the deep learning model's training loss is very low; otherwise, the training loss is high. The MAE and MSE comparisons are summarized in Table II, it is observed that the proposed system has lower MAE and MSE. The proposed system is to less error-prone than existing traditional methods. As the error value is lower for the proposed approach, it is expected that the proposed system will be more effective than other conventional deep learning methods.

D. Computational Efficiency Analysis

Table III summarized the computational efficiency of the AED-EGWO and existing methods. According to the experimental data, AED-EGWO takes 16 minutes to train the model, which is the shortest training time in the table, while RsetNet takes the longest of 23 minutes.

E. Discussion

Classification of fish species is an essential component of marine research and oceanography. It contributes significantly to the migration, breeding, and monitoring of endangered fish species. In the meantime, the manual classification of fish species is a labour-intensive and time-consuming process. To automate the classification procedure, numerous computer-aided fish species classification systems have been developed. In existing methods, the deep learning or machine learning models are trained using images of dead fish taken out from under water. When training deep learning models using these photos instead of real-time underwater photographs, the FN and FP rates are increased. High FN and FP rates have a noticeable impact on the classification model's precision. To prevent this, the AED model in this study was trained using underwater photos of live fish However, developing fish classification system using deep learning models with photos of underwater fish presents numerous challenges. Especially underwater, light penetration is very low, therefore pictures captured from this environment are extremely dim. Accordingly, the visibility of objects in underwater photographs are very poor. Therefore, underwater images have been normalized in this research to correct the problem. The normalized images through this research are shown in Fig. 12(b). Also, in the images taken from underwater, the color of water and other objects besides fish are occupied excessively, which increases the computational burden of the deep learning model. To correct it, in this research, the morphology of fish has been localized using the Simple Linear Iterative Clustering (SLIC) method. The results of fish morphology localization are shown in Fig. 12(b). Further, this study proposes the AED-EGWO framework for the classification of aquatic fish species. EGWO is being proposed for two significant responsibilities here.

TABLE II. TRAINING EFFICIENCY COMPARISON RESULTS

Methods	MAE	MSE
AED-EGWO	0.23	0.0529
CNN	0.42	0.1764
ResNet	0.77	0.5929
VGGNet	0.47	0.2209
AlexNet	0.31	0.0961

TABLE III. COMPUTATIONAL EFFICIENCY COMPARISON RESULTS

Methods	Training time	Classification time
AED-EGWO	16 minutes	1.37 Seconds
CNN	21 minutes	2.18 Seconds
ResNet	23 minutes	2.71 Seconds
VGGNet	19 minutes	1.92 Seconds
AlexNet	17 minutes	1.68 Seconds

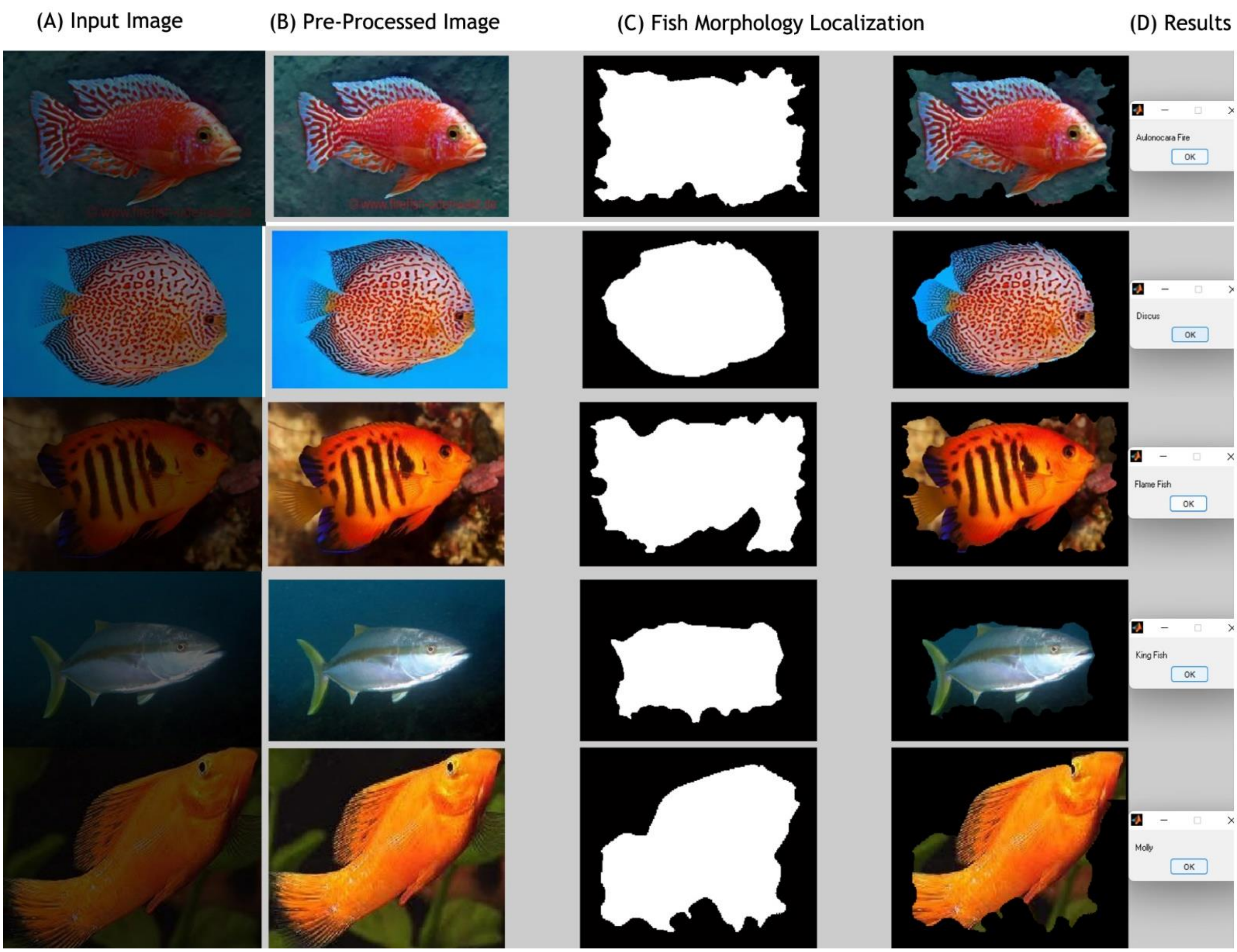

Fig. 12. Experimental Results.

Initially, EGWO eliminates irrelevant and redundant texture information from the data set during AED model training. Next, EGWO method select the optimal texture features from the data set with the lowest possible iterations.

According to the proposed EGWO, feature initialization is based on correlation rather than randomness. Therefore, it only initializes features with a strong relationship to classification results. Thus, the optimization method can find optimal features with less iterations. This eliminates the unnecessary processing resources required for AED training. Further, it enhances the fish classification accuracy to some extent.

V. Conclusion

In this study, the AED-EGWO methodology is developed for classifying fish species. This recommended classification method has two major components: optimal feature selection using EGWO and fish classification using AED. First, an improved grey wolf optimization approach is designed to identify the most important texture feature in the fish data set with the fewest possible iterations. Second, an auto encoder decoder network is developed to classify fish species based on the identified features. Finally, experimental study has been conducted to evaluate the proposed method's reliability and classification effectiveness. The experimental results of the proposed AED-EGWO approach were compared to existing deep learning models. The comparison results demonstrate that the proposed strategy has greater classification precision and reduced training loss.

This fish classification system was evaluated using images of clear sea water. In the future, this research can be expanded to classify fish species in blurry fresh water images.

References

[1] D. Rathi, S. Jain and S. Indu, "Underwater Fish Species Classification using Convolutional Neural Network and Deep Learning," 2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR), 2017, pp. 1-6, doi: 10.1109/ICAPR.2017.8593044.

[2] F. J. P. Montalbo and A. A. Hernandez, "Classification of Fish Species with Augmented Data using Deep Convolutional Neural Network," 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET), 2019, pp. 396-401, doi: 10.1109/ICSEngT.2019.8906433.

[3] S. Hasija, M. J. Buragohain and S. Indu, "Fish Species Classification Using Graph Embedding Discriminant Analysis," 2017 International

Conference on Machine Vision and Information Technology (CMVIT), 2017, pp. 81-86, doi: 10.1109/CMVIT.2017.23.

[4] M. T. A. Rodrigues, F. L. C. Pádua, R. M. Gomes and G. E. Soares, "Automatic fish species classification based on robust feature extraction techniques and artificial immune systems," 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), 2010, pp. 1518-1525, doi: 10.1109/BICTA.2010.5645273.

[5] Sarker IH. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Comput Sci. 2021;2(6):420. doi: 10.1007/s42979-021-00815-1. Epub 2021 Aug 18. PMID: 34426802; PMCID: PMC8372231. FLEX Chip Signal Processor(MC68175/D), Motorola, 15(3) (1996) 250-275.

[6] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8.

[7] Fouad, M. M., Zawbaa, H. M., Gaber, T., Snasel, V., & Hassanien, A. E. (2016). A fish detection approach based on BAT algorithm. In The 1st international conference on advanced intelligent system and informatics, AISI 2015 (pp. 273–283). Cham: Springer.

[8] Spampinato, C., Giordano, D., Di Salvo, R., Chen-Burger, Y. H. J., Fisher, R. B., & Nadarajan, G., (2010). Automatic fish classification for underwater species behavior understanding. In Proceedings of the first ACM international workshop on analysis and retrieval of tracked events and motion in imagery streams (pp. 45–50). ACM.

[9] Nagashima, Y., & Ishimatsu, T. (1998). A morphological approach to fish discrimination. In IAPR workshop on machine vision applications, Nov. 17–19 (pp. 306–309).

[10] Storbeck, F., & Daan, B. (2001). Fish species recognition using computer vision and a neural network. Fisheries Research, 51(1), 11–15.

[11] Rova, A., Mori, G., & Dill, L. M. (2007). One fish, two fish, butterfish, trumpeter: Recognizing fish in underwater video. In IAPR conference on machine vision applications, Tokyo, Japan (pp. 404–407).

[12] Shafait, F., Mian, A., Shortis, M., Ghanem, B., Culverhouse, P. F., Edgington, D., et al. (2016). Fish identification from videos captured in uncontrolled underwater environments. ICES Journal of Marine Science, 73(10), 2737–2746.

[13] Hernández-Serna, A., & Jiménez-Segura, L. F. (2014). Automatic identification of species with neural networks. PeerJ, 2, e563.

[14] Huang, P. X., Boom, B. J., & Fisher, R. B. (2012). Hierarchical classification for live fish recognition. In BMVC student workshop paper.

[15] Sun, X., Shi, J., Dong, J., & Wang, X. (2016). Fish recognition from low-resolution underwater images. In 9th International congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI), IEEE (pp. 471–476).

[16] Ludwig Bothmann,Michael Windmann,GoeranKauermann,” Realtime classification of fish in underwater sonar videos”2016.

[17] [Spampinato C, Giordano D, Salvo RD, Chen-Burger Y-HJ, Fisher RB, Nadarajan G (2010) Automatic fish classification for underwater species behavior understanding. In: Proceedings of the first ACM international workshop on analysis and retrieval of tracked events and motion in imagery streams,pp 45–50.

[18] Zhao J, Cao Y, Fan D, Cheng M, Li X, Zhang L (2019) Contrast prior and fluid pyramid integration for rgbd salient object detection. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 3922–3931.

[19] Rauf HT, Ikram Ullah Lali M, Zahoor S, Shah SZH, Rehman AU, Bukhari SAC (2019) Visual features based automated identification of fish species using deep convolutional neural networks. Comput Electron Agric 167:105075.

[20] Tharwat A, Hemedan AA, Hassanien AE, Gabel T (2018) A biometric-based model for fish speciesclassification. Fish Res 204:324–336.

[21] Jin L, Liang H (2017) Deep learning for underwater image recognition in small sample size situations. In: OCEANS 2017 - Aberdeen, pp 1–4.

[22] Sun, X., Shi, J., Dong, J., & Wang, X. (2016). Fish recognition from low-resolution underwater images. In 9th International congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI), IEEE (pp. 471–476).

[23] Qin JLH, Li X, Zhang C (2016) Deepfish: accurate underwater live fish recognition with a deep architecture. Neurocomputing 187:49–58 101088:57

[24] Rova A, Mori G, Dill LM, “One fish, two fish, butterfish, trumpeter: recognizing fish in underwater video”, In: IAPR conference on machine vision applications (2007).

[25] Fangfang Han, Junchaozhu, Bin Liu, “Fish Shoals Behavior Detection based on Convolutional Neural N/Wand Spatiotemporal Information”2020.

[26] Aditya Agarwal , Tushar Malani , Gaurav Rawal , Navjeet Anand, Manonmani S, 2020, Underwater Fish Detection, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 04 (April 2020),

[27] Christensen, J. H., Mogensen, L. V., Galeazzi, R., & Andersen, J. C. (2019). Detection, Localization and Classification of Fish and Fish Species in Poor Conditions using Convolutional Neural Networks. In Proceedings of 2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV) (pp. 1-6).

[28] J. H. Christensen, L. V. Mogensen, R. Galeazzi and J. C. Andersen, "Detection, Localization and Classification of Fish and Fish Species in Poor Conditions using Convolutional Neural Networks," 2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV), 2018, pp. 1-6, doi: 10.1109/AUV.2018.8729798.

[29] C.S.Arvind, R.Prajwal,Prithvi Narayana Bhat,A.Sreedevi, K.N.Prabhudeva ” , Fish Detection and Trackingin Pisciculture Environmentusing Deep InstanceSegmentation, 2019.

[30] SuryadiputraLiawatimena, Yaya Heryadi, Lukas,” A Fish Classification on Images using Transfer Learning and Matlab”, 2018.

[31] Bourlard, H., Kabil, S.H. Autoencoders reloaded. Biol Cybern 116, 389–406 (2022). https://doi.org/10.1007/s00422-022-00937-6.

[32] Li, Y., Wang, Z., Yang, X. et al. Efficient convolutional hierarchical autoencoder for human motion prediction. Vis Comput 35, 1143–1156 (2019). https://doi.org/10.1007/s00371-019-01692-9.

[33] Qiang Li, Huiling Chen, Hui Huang, Xuehua Zhao, ZhenNao Cai, Changfei Tong, Wenbin Liu, Xin Tian, "An Enhanced Grey Wolf Optimization Based Feature Selection Wrapped Kernel Extreme Learning Machine for Medical Diagnosis", Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 9512741, 15 pages, 2017. https://doi.org/10.1155/2017/9512741.

[34] Muni, MK, Parhi, DR, Kumar, PB. Implementation of grey wolf optimization controller for multiple humanoid navigation. Comput Anim Virtual Worlds. 2020; 31:e1919. https://doi.org/10.1002/cav.1919.