ORIGINAL CONTRIBUTION





Automated Freshwater Fish Species Classification using Deep CNN

Jayashree Deka¹ · Shakuntala Laskar¹ · Bikramaditya Baklial²

Received: 9 February 2022 / Accepted: 8 April 2023 / Published online: 21 April 2023 © The Institution of Engineers (India) 2023

Abstract Freshwater fish is considered a poor man's protein supplement as they are easily available in lakes, rivers, natural ponds, paddy fields, beels, and fisheries. There are various freshwater fish species that resemble each other, making it difficult to classify them by their external appearance. Manual fish species identification always needs expertise and so, is erroneous. Recently, computer vision along with deep learning plays a significant role in underwater species classification research where the number of species under investigation is always limited to a maximum of eight (8). In this article, we choose deep-learning architectures, AlexNet and Resnet-50, to classify 20 indigenous freshwater fish species from the North-Eastern parts of India. The two models are fine-tuned for training and validation of the collected fish data. The performance of these networks is evaluated based on overall accuracy, precision, and recall rate. This paper reports the best overall classification accuracy, precision, and recall rate of 100% at a learning rate of 0.001 by the Resnet-50 model on our own dataset and benchmark Fish-Pak dataset. Comprehensive empirical analysis has proved that with an increasing Weight and Bias learning rate, the validation loss incurred by the classifier also increases.

Keywords Fish classification · Fish species recognition · Resnet-50 · Deep-learning

Introduction

The aquaculture of South Asian countries like India, Pakistan, Bangladesh, Sri Lanka, and Nepal contributes in volume 7.9% to Asian and 7.1% to world fish production [1]. The predominant fish groups in these countries are the carps and the catfishes [2]. Among freshwater fishes, the Indian major carps are most cultured [3], followed by Exotic Carps, Minor Carps, Catfish, and Trout. Apart from these, there are plenty of small indigenous fish species (i.e.,-mola, climbing perch, barbs, batá, etc.) that grow in rivers, lakes, ponds, beels, streams, wetlands, lowland areas, and paddy fields. These indigenous fish are rich in nutrients, often containing high levels of zinc, iron, and vitamin A [4]. As per FishBase, India reports a total of 1035 freshwater fish species, while it is 241 from Pakistan and 256 from Bangladesh. Out of 1035 freshwater species, 450 are Small Indigenous Fish Species (SIFS, body length maximum 26 cm). Almost, 216 SIFS are reported from Assam, the North-Eastern part of India [5]. These indigenous fish species in the polyculture system have the potential of enhancing the nutritional security of the poor which in turn can provide greater employment opportunities in these underdeveloped Asian countries. As the production and marketing of these indigenous fish species are at the local level, they are mostly invisible in government statistics. Also, manual identification of fish species needs information about the number of fins, fin location, scales, and lateral line, head shape, color, and texture of the body. However, a lack of appropriate taxonomic knowledge or user expertise can potentially cause unsought repercussions for fishery management. Manual fish identification based on morphology is often considered an erroneous, inaccurate, and inefficient task. This shortcoming can be eliminated by using computer



[☐] Jayashree Deka jayashreedeka@gmail.com

Electrical and Electronics Engineering, Assam Don Bosco University, Guwahati, Assam 781017, India

Department of Zoology, Bahona College, Jorhat, Assam 785101, India

vision techniques, as they have shown outstanding performance in the various image-based classification task.

Diversification of fish species in aquaculture has become a popular research topic in India and Bangladesh. In poly-culture fisheries, it has been observed that the presence of certain fish species (including SIFS) may result in an increase or decrease in the growth of some other species (mainly carp) [6–10]. So, it is important to identify/classify rival species for early segregation to promote better growth. The study of fish diversification is necessary to evaluate selectivity, dietary overlap, and food competition among the cultured species. Therefore, it is crucial to correctly classify all the fish species reared in one environment to attain the targeted sustainable growth, and productivity. The correct classification will also help fish farmers with biomass estimation, disease identification, feeding planning, and cost estimation.

Based on computer vision algorithms, much research has been conducted for fish species recognition using fish images and underwater videos. Most of the recognition algorithms are based on hand-crafted feature extraction techniques (color, texture, and shapes) from images until the evolution of deep learning [11–22]. As deep-learning gains popularity in animal and plant species recognition, it becomes a common choice for fish classification research too. Li et al. 2022 have studied how recent machine vision techniques have changed the outlook of feature-based fish species classification experiments [23]. As the underwater fish images do not possess much color and texture information due to poor contrast and varied illumination, so only shape information plays a significant role in underwater fish image classification models. It has been observed that the research on fish identification mostly deals with marine-water fish resources and only a handful of research has been executed on fresh-water fish species. Meanwhile, a few researchers utilized pre-trained CNN models for fresh-water fish image classification where a constant background was maintained for all images captured in a laboratory environment [24–27]. Although these works demonstrate satisfactory performance, further research into the effect of varied backgrounds on the performance of these deep-learning models is still required. Moreover, there is still scope for improvement in the accuracy of the small-indigenous fish classification problems [24, 28]. Indian water bodies carry varieties of fresh-water fish species, still, we are unable to track down any research article focusing related to these species. Here, in this work, we perform a deep-learning-based based experiment using MATLAB to classify images of 20 varieties of fresh-water fish including small indigenous species that are native to Assam, India. We deploy a deep residual 50-layer network, ResNet-50, and an Alex-Net architecture to build a robust model to work on different background images. A total of

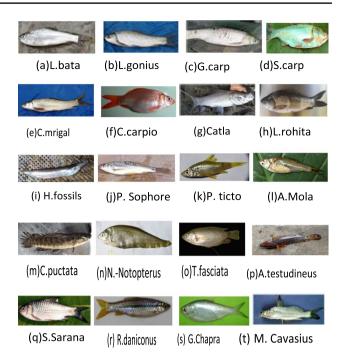


Fig. 1 Fish sample images used

587 images of these species are captured under different background conditions under natural lighting. Figure 1 shows the fish sample images that we have collected for the experiment.

The proposed method will help in the accurate classification of carps and small indigenous fish species reared in a ploy-culture environment. The proposed work will also help different stakeholders in fisheries, i.e., government officials, policymakers, scientists, and farmers in developing effective conservation measures for these fresh-water fish species for stock management.

The main contributions of this paper are:

- I. The paper reports for the first time, the application of deep learning networks for automated classification of 20 freshwater fish species including carps and SIFS native to North-eastern parts of India. Images captured from collected samples are labeled accurately with the help of our third author who has an experience with fish anatomy study.
- II. The proposed method utilizes a cost-effective image acquisition device by using a smartphone and the same is being verified by the concerned expert.
- III. The paper investigates the appropriate learning rate, weight, and bias learning rate factors to achieve the best classifier performance.
- IV. As local fish farmers are exploring mixed culture farming, early detection, and accurate classification of fish species will help those farmers to segregate the rival



species in order to gain maximum growth and better vield.

Literature Search

Research on fish classification and recognition tasks started back in the year 1993 [11]. Fish recognition was first investigated on dead demersal and pelagic fishes where color and shape features were used for recognition [11]. They reported 100% sorting for demersal and 98% for pelagic fish, respectively.

Cadieux et al. 2000 developed a system for monitoring fish using a Silhouette Sensor]. Fish silhouettes were used for extracting the shape features using Fourier Descriptor and other shape parameters i.e., area, perimeter, length, height, compactness, convex hull area, concavity rate, image area, and the height of the image [12]. A multi-classifier approach was used for classification by incorporating Bayes maximum likelihood classifier, Learning Vector Quantification classifier, and One-Class-One-Network (OCON) neural network classifier. The combined classifier resulted in an accuracy of 77.88%.

The fish classification from videos in the constrained environment was explored based on the texture pattern and shape of the Striped Trumpeter and the Western Butterfish from 320 images [13]. The SVM classifier with a linear kernel outperforms the polynomial kernel with an accuracy of 90%

Spampinato et al. performed a classification of 10 fish species using a dataset of 360 images. They extracted texture features using Gabor filtering, and boundary features by using Fourier descriptors and curvature analysis [14]. The system showed an accuracy of 92%. Alsmadi et al. 2011 developed a backpropagation classifier-based fish image recognition. They extracted fish color texture features from 20 different fish families using GLCM and achieved an overall testing accuracy of 84% on 110 testing fish images [15].

Images from Birch Aquarium were studied using a Haar detector and classifier for classifying Scythe Butterflyfish images [16]. Hu et al. utilized color, texture, and shape features of six freshwater fish images that are collected from ponds in China [17]. Color features were extracted from skin images by transforming RGB images into HSV color space, whereas statistical texture features and wavelet texture features were obtained from a total of 540 fish images [17]. The accuracy achieved was 97.96% with some miss-classification due to similar texture features among the species.

In a separate experiment, features were extracted by SIFT and SURF detectors for Tilapia fish identification [18]. It was observed that the performance with SIFT (69.57%) features was lower than that with SURF (94.4%) features. Another fish recognition model was introduced for 30 fish

species from Thailand based on shape and texture features. The system achieved a precision rate of 99% with the ANN classifier [19]. Rodrigues et al. used SIFT feature extraction & Principal Component Analysis (PCA) for the classification of 6 fish species from a dataset of 162 images and achieved an accuracy of 92% [20].

Fish recognition from live fish images of the South Taiwan sea was performed using 66 different attributes based on color, texture, and shape features [21]. As the recognition task depended on the detection of the fish in the video, the system showed a poor recognition rate. Different feature descriptors for different parts of fish images were employed in [22]. SIFT and color histograms methods extracted partial features from images and a partial hierarchical classifier predicts the images with a score of 92.1% on the FISH4KN-LOWLEDGE dataset [22]. Mattew et al. 2017 used landmark selection and geometrical features for the identification of four Indian marine fish using ANN and they achieved an accuracy of 83.33% [29].

The popularity of Complex Neural Networks (CNNs) has been proved in fish detection and species classification tasks too. A deep CNN based on the Alex-Net model for fish species classification on the LifeClef-15 dataset was used and it gained an F-score of 73.5% on underwater videos [30]. Video clips of the 16 species collected from the Western Australia sea were studied using a pre-trained deep CNN where a cross-layer pooling was used as feature extractor and classification was performed using an SVM classifier thereby achieving an accuracy of 94.3% [31].

Different fish images were collected from the internet by the authors in [32]. They used a deep residual 50-layer network for residual feature extraction and an SVM classifier on those images. They examined the effect of the frozen layer on the overall system performance. The work reported an accuracy of 97.19%. Salman et al.2019 used CNN-SVM for best accuracy on the same dataset as in [31] where the maximum accuracy they claimed was 90% [33]. Khalifa et al. [34] proposed a 4-layer CNN on QUT and LifeClef-15 dataset image classification and achieved a testing accuracy of 85.59%.

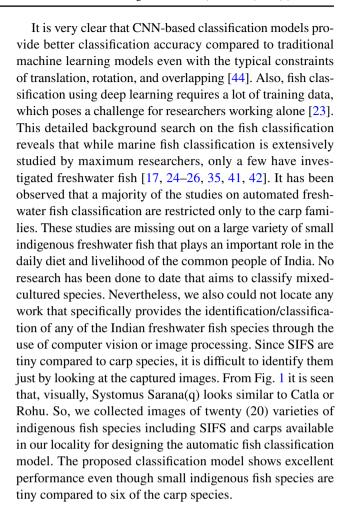
Islam et al. performed eight (8) different types of small indigenous fish classification experiments in Bangladesh [35]. They developed an HLBP (hybrid census transform) feature descriptor for feature extraction and used an SVM classifier to classify the fish images [35]. They achieved an accuracy of 94.97% with a linear SVM kernel. A 32-layer CNN network was employed on 6 freshwater fish species from the Fish-Pak [24] dataset [25]. They investigated the effect of different learning rates on the system performance and reports the best classification accuracy of 87.78% at a learning rate of 0.001 and epoch = 350 [25]. Another six SIFS image classifications from Bangladesh were studied using color, texture, and geometrical features [26]. A total



of 14 features were extracted, which was further reduced by using PCA and then fed to an SVM classifier resulting in a classification accuracy of 94.2% [26]. For the first time, an experiment for finding fish species, order, and family of 68 Pantanal fish species was executed using Inception layers [28]. These layers extracted the features from the fish images and then three branches were implemented with the help of fully connected layers to predict the class, family, and species of the fish images. The technique showed maximum accuracy for order classification.

In a separate experiment conducted in the Lifeclef19 and QUT dataset, an Alex-Net with a reduced no. of convolutional layers and an extra soft attention layer added in between the last fully connected layer and the output layer was used. This gave an accuracy of 93.35% on the testing dataset [27]. A deep YOLO network on the LifeClef-15 dataset and UWA datasets were used for fish detection and classification [36]. The classification accuracy observed was 91.64% and 79.8% on both datasets respectively. Deep DenseNet121 and SPP-Net were used for the classification of six different kinds of freshwater fish images downloaded from the internet in [37]. They used an 'sgd' optimizer for training the models and a true classification rate(recall) of 99.2% was achieved by SPP-Net [37]. In the later year, a pre-trained VGG-16 model was utilized on four major carp images and was able to achieve an accuracy of 100% with fivefold validation [38]. A modified Alex-Net was studied using QUT and LifeClef dataset [39]. They adopted a dropout layer before the SoftMax layer that helped in a higher accuracy over the traditional Alex-Net model and reported 90.48% accuracy on the validation dataset. To eliminate the problem of class imbalances, the focal loss function together with the SeResNet-152 model was used for training and achieved an accuracy of 98.8% [40]. It has been observed that, with the class balanced focal loss function, the performance increased for the species with fewer datasets too.

Dey et al. 2021 used a FishNet app based on 5-layer CNN for the recognition of eight varieties of indigenous freshwater fish species from Bangladesh. [41]. An 'adam' optimizer was chosen for simulation and three different drop-out rates were applied to the last three fully-connected layers of the CNN. This method attained a validation accuracy of 99% with a 3.25% validation loss [41]. Abinaya et al. 2021 used the Fish_Pak dataset and segmented fish heads, scales, and body are segmented from those images [42]. They used a Naïve Bayesian fusion layer in each training segment to increase the classification accuracy to 98.64% with the Fish-Pak dataset while with the BYU dataset accuracy achieved was 98.94% [42]. Jose et al. [43] used deep octonion networks as feature extractors and a soft-max classifier for tuna fish classification. This method reported an accuracy of 98.01% on the self-collected tuna fish dataset and 98.17% accuracy on the Fish-Pak dataset.



Materials and Methods

Overview of the Proposed System

This work proposes the classification of freshwater fish using a deep Convolutional Neural Network. Figure 1 presents the layout of the proposed classification model.

Data Acquisition and Data preparation

Twenty different varieties of freshwater fish species are considered for the fish classification purpose. For this purpose, fifty-seven (57) fish samples of each of these 20 varieties are captured from different local ponds in Assam, India using fishing nets. The images from these species are acquired in natural lighting conditions using a Samsung GalaxyM30s mobile phone that has a triple camera. We use the wide view mode of the main rear camera with a specification of 48-megapixel, aperture (f/2.0) lens, and focal length of 26 mm. The sensor size is 1/2.0" and the pixel size is 0.8 µm. The third author of the paper, who is actively involved with the identification and study of fish



Table 1 Data collection per species

Species	Fish_Pak	Ours	Total
Labeo gonius	_	20	20
Labeo bata	_	28	28
Channa punctata	_	13	13
Amblypharyngodon_mola	_	15	15
Mystus Cavasius	_	16	16
Guduchia Chapra	_	18	18
N. Notopterus	_	14	14
R. daniconus	_	18	18
Trichogaster fasciata	_	20	20
Systomus Sarana	_	16	16
Puntius Ticto	_	9	9
Heteropneustes_fossils	_	20	20
Anabas_testudineus	_	23	23
Puntius Sophore	_	17	17
Cirrinhus Mrigal	70	22	92
Labeo rohita(Rou)	73	16	89
Cyprinus Carpio	50	10	60
Catla	20	12	32
CIdella(grass-carp)	11	4	15
Silver carp	47	5	52
Total	271	316	587

fauna across Assam, has supported the labeling of the collected images. Table 1 shows the fish image distribution per species used in the experiment. We use Fish-Pak [24] dataset for simulating the system in its initial phases. Images of the major carps are downloaded from the Fish-Pak dataset as these are popular carps native to South Asian countries. A total of 587 images of these species are accumulated for the

Fig. 2 Proposed methodology for Fish Classification

Pre-processing IOIN resize, augmenting) Feature Validation images extraction Training images Image acquisition Feature extraction *Define layers *define training Knowledge options(momentu Train the CNN m, learning rate, weight and bias factor Predict the classify image Fish Image with model with class-label name

desired fish classification task and labeled with the help of the third author.

As these images are not enough for training the classifier, so, without acquiring new images, the size of the training images is increased with the help of various image augmentation techniques. The process of image augmentation generates duplicate images from the original images without losing the key features of the original images (Fig. 2).

Image Pre-Processing

Since all the images, we have collected are of different sizes and in different formats (jpeg and png), we first convert all images into 'png' format in 'Paint' app. Again, as the number of images in some species are few, so, image augmentation techniques are applied to increase the size of the dataset. As the dataset is very small, so, we have considered offline image augmentation technique which involves storing of images on the computer disk after each augmentation operation.

Let I(x, y) be the image. Following augmentation techniques are employed to each of these original images.

Cropping

Cropping is the basic image reframing operation where a desired portion of the image is retrieved. Let (x_1, y_1) be the origin of the image and (x_e, y_e) be the end point of the image. The new image co-ordinate (x_p, y_p) after cropping operation will be

$$I(x_{p}, y_{p}) = I(x_{e} - x_{1}, y_{e} - y_{1})$$
(1)



Flipping

All the original images in the dataset were flipped upside down. Then the images were again flipped horizontally as it is common than flipping in the vertical axis. It is one of the easiest ways to implement and has proven useful on datasets such as CIFAR-10 and ImageNet. Upside down images have a new co-ordinate of I(x, width-y-1) whereas, a horizontally flipped image co-ordinate is I(width-x-1, y) in the new image.

Rotation

The image rotation routine involves an input image, a rotation angle θ , and a point about which rotation is done. If a point (x_1, y_1) in original image is rotated around (x_0, y_0) by an angle of θ^0 then the new image co-ordinate (x_2, y_2) , can be presented by following equation.

$$x_2 = (x_1 - x_0) * \cos\theta + (y_1 - y_0) \times \sin\theta + x_0$$
 (2)

$$y_2 = -\sin\theta * (x_1 - x_0) + \cos\theta * (y_1 - y_0) + y_0$$
 (3)

If $x_0 = y_0 = 0$ and $\theta = 90^{\circ}$ clockwise, then (x_2, y_2) becomes

$$x_2 = y_1 \tag{4}$$

All the original images and upside-down images are rotated 90⁰ clockwise, anticlockwise and 180⁰ clock-wise. The image count post augmentation techniques goes up to 2781.

Convolutional Neural Network

A Convolutional neural-network (CNN) is a deep structured algorithm commonly applied to visualize and extract the hidden texture features of image datasets. Since the last decade, CNN features and deep networks have become the primary choice to handle any computer vision tasks. Different CNN architecture models utilize different layers i.e., convolution layers, batch-normalization, max-pooling layer, fully connected layer, and a soft-max layer. The major advantages of the deep CNN architecture over the manual supervised method are its self-learning and self-organized characteristics [38].

The first layer in the CNN architecture is the image input primary layer that takes up 2-D and 3-D images as input. The size of the image must be initialized into this layer [45]. Let $I = m \times n$ is an input image where m and n represent the rows and columns respectively. Equation (5) represents the image input layer by a function:



where p is a deep learning architecture model which can be series or parallel network. The layer In_p performs some pre-processing if required, before passing to the next layer called convolution layer. The convolution layer uses filters that perform convolution operations and scans the input I with respect to its dimensions. The input image becomes less noticeable at the output of a convolution layer. However, particulars such as the edges, orientation, and patterns become more prominent, and these are the key features from which a machine actually learns. The output obtained from the convolution operation of the input, I, is called as feature map, I

$$f_m = Conv(I, k, d, r_l) \tag{6}$$

where k is the filter window, d is the stride, and r_l is RELU activation function. Next, batch normalization layer evaluates the mean and variance for each of this feature map. Batch normalization process makes it possible to apply much higher learning rates without caring about initialization.

The pooling layer performs the down-sampling operation on the feature maps obtained from the preceding layer and produces a new set of feature maps with compressed resolution. The main objective of a pooling layer is to gather crucial information only by discarding unnecessary details. Two main pooling operations are average pooling and max pooling. In average pooling, down-sampling is performed by partitioning the input into small rectangular pooling regions and then computing the average values of each region.

In the Max-pooling operation, the spatial size of the previous feature map is reduced. It helps to bring down the noise by eliminating noisy activations and hence is superior than average pooling. The pooling operator carries forward the maximum value within a group of T activations. Then *n*-th max-pooled band consists of K-related filters

$$p_{k,m} = \max(h_{j,(m-1)U+t}) \tag{7}$$

where $U \in \{1, 2, 3, ..., T\}$ is pooling shift operation that permits overlap between pooling regions if U < T. This results in shrinking the dimensionality of the K convolutional band to N pooled band where, N = (k - T)/U + 1 and final layer becomes $p = [p_1, p_2 ... p_N] \in T^{NJ}$.

At the end of every CNN, there lies one or more fully-connected (FC) layers. The output of pooling layer is flattened and then, put in to the fully-connected layer. The neurons in the FC layer performs a linear transformation to its input and later, a nonlinear transformation takes place as follows

$$Y = f\left(\sum_{i=1}^{nH} W_{jk} X + B\right) \tag{8}$$



where f is the activation function, X is the input, W is the weight matrix, j is the number of neurons in previous layer, k is the neuron in current layer, and B is the bias weight. It is important to have the same number of output nodes in the final FC layer as the number of classes.

Finally, Soft-max function computes the probability of each of the class that is outputted by the final FC layer. A soft-max function is represented by

$$\operatorname{softmax}(y_i) = \exp^{(y_i)} / \sum_{i} \exp(y_i)$$
(9)

From the background search on fish classification and recognition, it is observed that CNN has shown outstanding result in fish classification compared to traditional method of manual feature-based image classification. Some of the species in our dataset had fewer images that resulted in class imbalance. So, to overcome this problem we have employed transfer learning [46] method where a pretrained model of CNN(ResNet-50 and Alex-Net) is being chosen to solve the task.

AlexNet

Alex-Net is considered one of the most popular CNNs used for pattern recognition and classification/identification applications. It incorporates 60 million parameters and 650,000 neurons [47]. The architecture consists of 5 convolution layers, max-pooling layers, and 3 consecutive fully connected layers. The input layer takes the images of size $227 \times 227 \times 3$ and then it filters the images by using a total 96 number of kernels of size 11×11 with a stride of four. The output from this layer is passed through the second convolutional layer of 256 feature maps and the size of each kernel is 5×5 with a stride of one. The output from this layer passes through a pooling and normalization layer resulting in a $13 \times 13 \times 256$ output. The third and fourth convolution layer use a kernel of 3×3 with a stride of one. Then the output of the last convolutional layer is flattened through a fully connected layer of 9216 feature maps which is again to a fully connected layer with 4096 units. The last layer is the output that uses the SoftMax layer with six units (classes) for the Fish Pak dataset and 20 units (classes) for our dataset.

ResNet-50

The ResNet-50 architecture comprises of 5 stages each with a convolution and identity block, one average pool layer, and one fully connected layer with 1000 neurons [48]. Each convolution block has 3 convolution layers and each identity block also has 3 convolution layers. The input to Resnet-50 is of size $224 \times 224 \times 3$. Each convolutional layer is followed by a batch normalization layer and a nonlinear

RELU function. Batch normalization layers standardize the activations of a given input volume before passing it into the succeeding layer. It calculates the mean and standard deviation for each convolutional filter response across each mini-batch at each iteration to normalize the present layer activation. Resnet-50 has over 23 million trainable parameters. The original residual unit is modified by bottleneck design as shown in Fig. 3. For each residual function F, a stack of 3-layers was used instead of original 2-layers.

Experimental Settings and Evaluation Criteria

The experimental simulation is performed in MAT-LAB2018a environment in a system with an Intel Core i5 7300HQ Processor of 2.50 GHz turbo up to 3.50 GHz with NVIDIA GeForce GTX 1050 Ti with a computation capability of 6.1.

We use fine-tuned Resnet-50 and Alex-net from MAT-LAB's Deep-learning Toolbox. Due to variations in the image sizes, we use 'augmentedImageDatastore' function from MATLAB to resize all the manually augmented images.

Our simulation-based experiments are performed first on Fish_Pak dataset. A total of 1,191 augmented images from this dataset are segregated randomly into 70% training (834) and 30% validation (357) images. As our dataset is small, we use fine-tuned approach where the top 3 layers of Resnet-50 and Alex-Net models are removed and then again, we re-arranged the model by adding the three layers with six output neurons for the output layer. The weight and

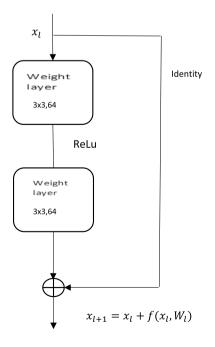


Fig. 3 basic residual building unit of Resnet-50



bias learning rate of the fully connected layer are varied for best performance of the classifier while the bottom layer weights are frozen. Freezing the bottom layer weight helps in increased network speed during training. For training the classifier an "sgdm" optimizer with a "momentum" of 0.9 is used. Training data is divided into a "mini-batch" size of 20 and we choose a maximum epoch = 30. The learning capability of both the CNNs, i.e., Alex-Net and Resnet-50 is evaluated in terms of classification accuracy, loss function (training and validation loss) and confusion matrix (CM). CM is an ideal performance measure to assess accuracy of multi-class balanced classification problem like fish classification. It is a two-dimensional matrix with target and predicted class at the rows and columns, respectively. It provides insight into class-specific and also overall classification accuracies. It gives the overall classification accuracy evaluated based on the number of images that have been classified correctly.

As both the CNN models show satisfactory performance with Fish-Pak dataset, so next, we apply our own dataset consisting of a total 2781 augmented images. The training of 1946 images(70% of 2781) is carried out and the 834 images are chosen for the validation of the proposed CNN classifier. A 'mini-batch size of 32' is chosen for both training and testing purposes. We use a gradient descent solver momentum

optimizer('sgdm') with a momentum of 0.9 to train our dataset for the two architectures under consideration.

The learning rate (Lr) in every CNN architecture produces a prominent impact on the overall accuracy of the classifier. While a low learning rate (Lr) can exhibit a slow execution rate of the entire model and thus increasing the time complexity. On the other hand, using too large a learning rate, Lr, there is a possibility that the model may get stuck at the suboptimal output results. So, we have experimented with different values of the learning rate, i.e., 0.01, 0.001, and 0.0001. Tables 2 and 3 represent performance analysis of the two network models at different learning rate with variations in the 'WeightLearnRate' factor and 'Bias-LearnRate' factor of the *fully connected layers*.

The performance of the classifier is determined by four evaluation parameters which are important for multi classification problems: overall classification accuracy, mean precision, recall, and mean F-score. The recall is an important parameter in a multi-class classification task as it refers to the no. of positive classes that are labeled correctly. Precision gives information about the actual positive cases out of the total positive prediction. All these parameters are calculated as per Eqs. (10)–(12).

Table 2 Performance Analysis of Resnet-50 and Alex-Net classifier on Fish-Pak dataset with Max epoch = 30, validation frequency = 3, momentum = 0.9

Parameters	Resnet-50			AlexNet			
	Lr = .01	Lr = .001	Lr=.0001	Lr = .001	Lr=.0001	Lr=0.00001	
WeightLearnRateF	actor' = 10,'I	BiasLearnRate	Factor' = 10				
Accuracy	100	100	98.04	98.60	99.16	94.69	
Precision	100	100	97.78	98.43	99.24	94.21	
Recall	100	100	98.21	98.43	99.22	95.61	
F-score	100	100	98.00	98.43	99.23	94.91	
Training loss	.0003	.0001	.0147	.0014	.0086	.4798	
Validation loss	.0238	.0382	.1265	.0315	.0232	.1766	
WeightLearnRateF	actor' = 20,'I	BiasLearnRate	Factor' = 20				
Accuracy	99.44	99.16	98.88	97.76	98.04	91.34	
Precision	99.24	98.77	98.74	98.07	97.34	89.93	
Recall	99.45	99.43	99.72	97.85	98.09	93.92	
F-score	99.35	99.09	98.99	97.96	97.71	91.88	
Training loss	.0145	.0013	.0162	.0043	.0020	.2763	
Validation loss	.1009	.0526	.0790	.0771	.0460	.2322	
WeightLearnRateF	actor' = 30,'I	BiasLearnRate	Factor' $= 30$				
Accuracy	97.21	99.72	98.88	98.04	98.32	96.09	
Precision	96.93	99.80	98.96	98.44	98.79	95.99	
Recall	96.87	99.76	99.05	98.73	98.17	96.11	
F-score	96.90	99.72	99.00	98.14	98.48	96.09	
Training loss	.0620	.0002	.0128	.0546	.0057	.0725	
Validation loss	.1288	.0165	.1177	.0985	.0823	.1435	



Table 3 Performance Analysis of Resnet-50 and Alex-Net classifier on own dataset with Max epoch = 30, validation frequency = 3, momentum = 0.9

Parameters	Resnet-50			AlexNet			
	Lr = .01	Lr = .001	Lr=.0001	Lr=.001	Lr=.0001	Lr=0.00001	
WeightLearnRateF	actor' = 10,'I	BiasLearnRate	Factor' = 10,				
Accuracy	99.52	99.76	98.68	97.60	97.00	92.08	
Precision	99.52	99.71	98.52	97.17	96.87	90.13	
Recall	99.43	99.91	98.87	97.48	96.75	92.11	
F-score	99.47	99.81	98.70	97.32	96.81	91.11	
Training loss	.0003	.0020	.0328	0.0597	.0145	.3374	
Validation loss	.0352	.0184	.1010	0.0949	.0796	.2774	
WeightLearnRateF	actor' = 20,'I	BiasLearnRate	Factor' = 20				
Accuracy	99.52	100	98.80	96.76	97.12	93.28	
Precision	99.40	100	98.32	96.13	97.04	92.00	
Recall	99.31	100	98.83	96.12	97.12	93.64	
F-score	99.36	100	98.57	96.12	97.08	92.81	
Training loss	.0005	.0004	.0235	.0173	.0457	.2150	
Validation loss	.0429	.0206	.0728	.0903	.0840	.2105	
WeightLearnRateF	actor' = 30,'1	BiasLearnRate	eFactor' = 30				
Accuracy	99.16	99.76	98.44	95.92	97.84	93.88	
Precision	98.97	99.62	98.18	95.68	97.62	92.50	
Recall	99.20	99.78	98.35	96.47	98.38	94.38	
F-score	99.16	99.70	98.26	96.07	98.00	93.43	
Training loss	.0028	.0002	.0103	.0201	.0014	.2114	
Validation loss	.0720	.0269	.0707	.1495	.0556	.1938	

$$Precision = \frac{True \ Positive}{True \ positive + False \ positive}$$
 (10)

$$Recall = \frac{True \ Positive}{True \ Positive + False \ Negative}$$
 (11)

$$F - score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$
 (12)

Results and discussions

The CMs shown in Fig. 4, 5 and 6 give the idea of how all the fish species are classified to its corresponding classes by Resnet-50 on the Fish-Pak dataset. The rows in the CM correspond to the Output Class (predicted class) and the columns correspond to the true class (target class). The diagonal values represent the number of correctly classified species and the percentage value represented in the diagonal cells is the recall value, i.e., out of all predicted classes how many are actually correctly classified(positive). All confusion matrix plot is obtained by using the MATLAB function "Plotconfusion". Figure 4 represents the CM of Resnet-50 architecture that gives the best classification accuracy of 100% at a learning rate of 0.01 and weight = bias = 10 on Fish Pak dataset. Here, all the fish test images are predicted

successfully into their corresponding classes. Figure 5 is the best CM achieved with bias and weight factor 20 at a learning rate 0.01 for ResNet-50 and the model in this case fails to classify two(2) of the Catla fishes and mistakes it as Grass-carp(G.carp). With Lr=0.001 and Weight=Bias=30, the maximum accuracy achieved with Resnet-50 is 99.72% shown in Fig. 7 and in this case, only one mrigal fish is misclassified as "Rou" while all others are 100% correctly classified. As maximum validation accuracy is achieved at

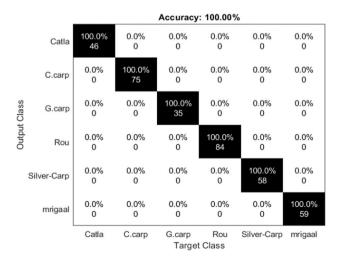


Fig. 4 CM of Resnet-50 on Fish-Pak dataset at Lr=.01, Weight=10=Bias



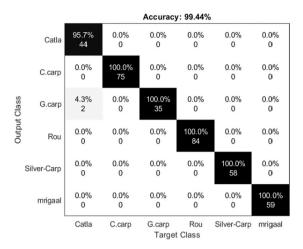


Fig. 5 CM of Resnet-50 on Fish Pak dataset at Lr=.01, Weight=20=Bias

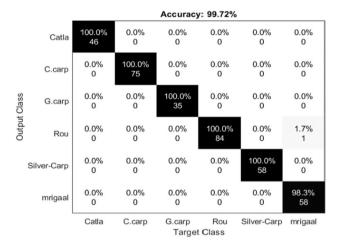


Fig. 6 CM of Resnet-50 on Fish Pak dataset at Lr=.001, Weight=30=Bias

learning rate, Lr = 0.01 and weight = bias rate = 10, so the variation of accuracy and loss for this is shown in Fig. 10.

With Alex-Net, the maximum classification accuracy of 99.16% is obtained at Lr = 0.0001 and Weight = Bias factor = 10. The CM corresponding to this is shown in Fig. 7. From the CM, we come to know that all the predicted fish of 'C. carp', 'G. carp', 'Silver-carp' and 'mrigaal' are 100% correctly classified. On the other hand, one (1) 'Rou' is misclassified as 'C. carp' and another one (1) is mistaken as 'Catla', thus bringing down the recall rate of 'Rou' to 97.6%. Here, one of the Catla is also mis-classified as 'Rou' so, the recall value of 'Catla' is reduced to 97.8%. The CM in Fig. 8 tells us how the AlexNet 100% correctly predicted all the 'C. carp', 'G. carp', 'Silver-Carp', but it fails in predicting 7 other fish images. Two

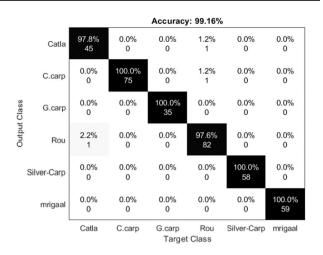


Fig. 7 CM of Alex-Net on Fish Pak dataset at Lr=.0001 and Weight=10=Bias

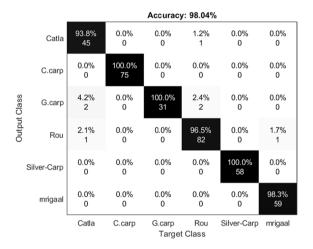


Fig. 8 CM of Alex-Net on Fish Pak dataset at Lr = .0001 and Weight = 20 = Bias

of the 'Catla' fish are mis-classified as 'G.carp' while one is predicted wrong as Rou. Additionally, two of the Rou fish are predicted wrong as 'G. carp', one as 'Catla' and only one 'mrigaal' is wrongly predicted as 'Rou'. With Lr=0.0001 and Weight=Bias=30, total 6 fish images are classified inaccurately and overall accuracy achieved is 98.32%. From Fig. 9 we see that one (1) 'Catla' and one (1) 'C.carp' fish is wrongly classified as 'Rou' (Fig. 10). Whereas, one (1) Rou fish is misclassified as 'mrigaal' and one as 'Catla'. Figure 11 shows the best performance of AlexNet from training at learning rate, Lr=0.0001 and weight=bias learning rate=10.

Figure 12, 13 and 14 are the best confusion matrices derive from ResNet-50 at different learning rate, Lr and "Weight and Bias Learning Rate" on our own dataset. In Fig. 12, all the fish species are correctly classified to its



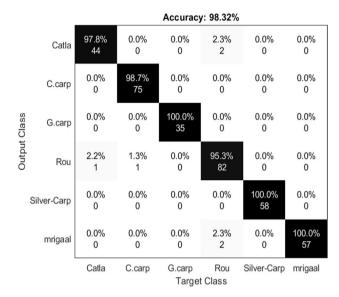


Fig. 9 CM of Alex-Net on Fish Pak dataset at Lr = .0001 and Weight = 30 = Bias

corresponding classes. Though in Figs. 13 and 14, accuracies are same, their mis-classification cases are different. For same accuracy, in Fig. 13, two of the Rou fish are predicted

wrongly as G. Carp, whereas in Fig. 14, one of the 'Catla' fish is misclassified as 'S. Sarana' and one 'Rou' as 'Catla'. With AlexNet on our dataset, mis-classification rate is quite high which is clearly visible from CMs shown in Fig. 15, 16 and 17. These are the best CMs of Alex-Net when we vary the Weight and Bias Learn Rate factor from 10 to 30 on three different learning rates. In Fig. 15, total mis-classification is 20, in Fig. 16 it is 24 and in Fig. 17 it is 18. The training progress of best performance by Resnet-50 and Alex-Net on our dataset is shown in Figs. 18 and 19, respectively.

It is clearly visible from the confusion matrices (CMs) that ResNet-50 is the best at correctly classifying all the fish species into their corresponding classes with few miss-classification.

The performance parameters of both the CNN models i.e., Resnet-50 and Alex-Net are shown in Tables 2 and 3 on the Fish-Pak dataset and our own dataset, respectively. For Fish-Pak dataset, Resnet-50 offers 100% accuracy, precision and recall for both Lr = 0.01 and 0.001 at a weight and bias factor = 10.However, the validation loss is more with Lr = 0.001 than Lr = 0.01. So, therefore, we can say best performance is achieved at Lr = 0.01. On the other hand, the maximum achievable accuracy with AlexNet on the Fish_Pak dataset is 99.16% at a learning rate of 0.0001 with

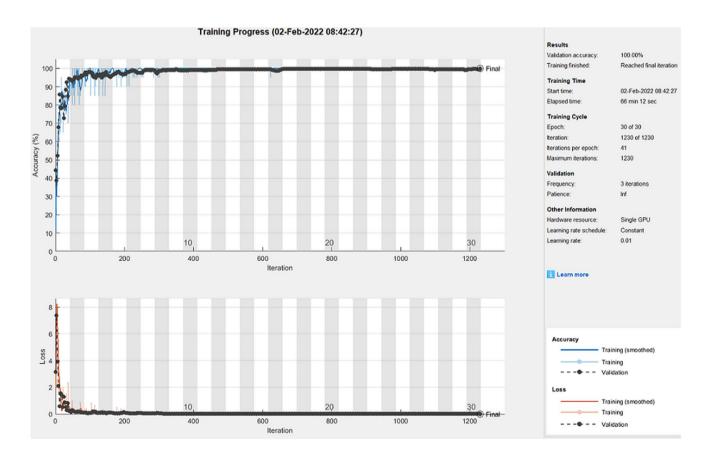


Fig. 10 Accuracy and Loss curve of ResNet-50 at Lr = 0.01 and weight = 10 for Fish-Pak dataset



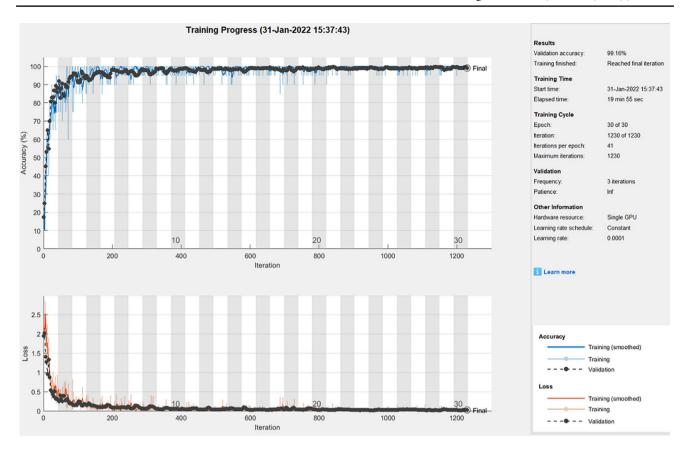


Fig. 11 Accuracy and Loss curve of AlexNet-50 at Lr=0.0001 and Weight=10 for Fish-Pak dataset

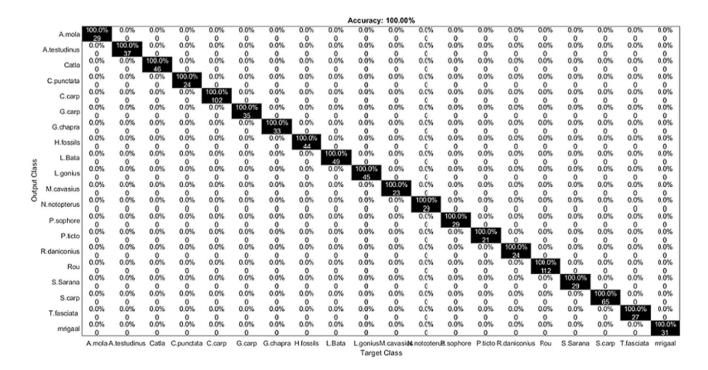


Fig. 12 CM of Resnet-50 at Lr=0.001 and Weight=20=Bias factor on own dataset



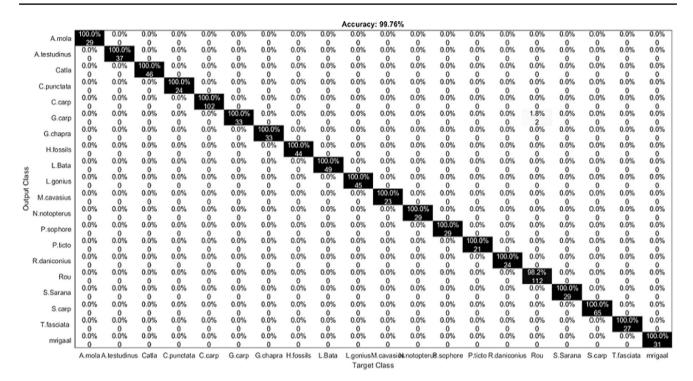


Fig. 13 CM of Resnet-50 at Lr=0.001 and Weight=10=Bias factor on own dataset

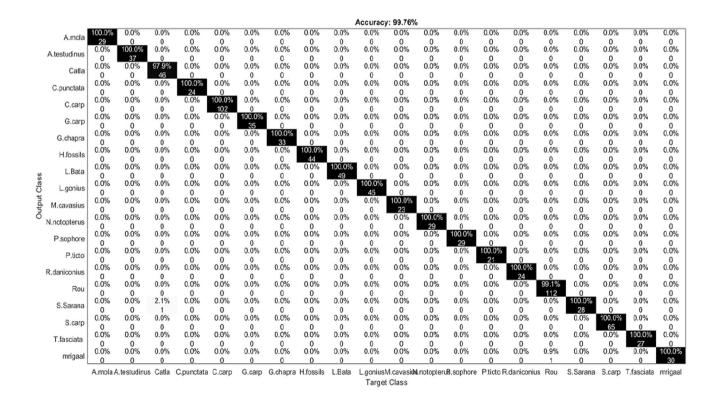
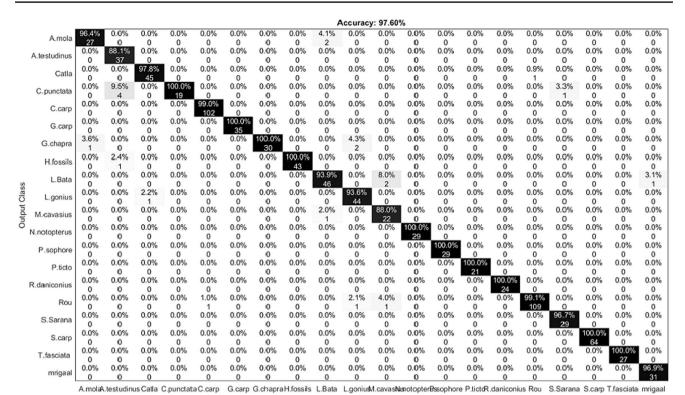


Fig. 14 CM of Resnet-50 at Lr = 0.001 and Weight = 30 = Bias factor on own dataset





Target Class

Fig. 15 CM of Alex-Net at Lr = .001 and Weight = 10 = Bias factor on own dataset

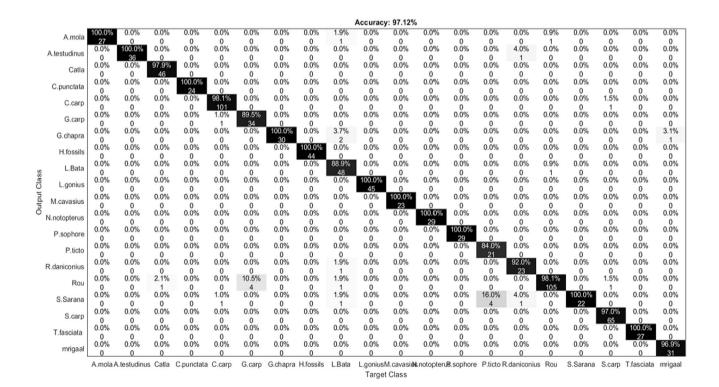


Fig. 16 CM of Alex-Net at Lr = .0001 and Weight = 20 = Bias factor on own dataset



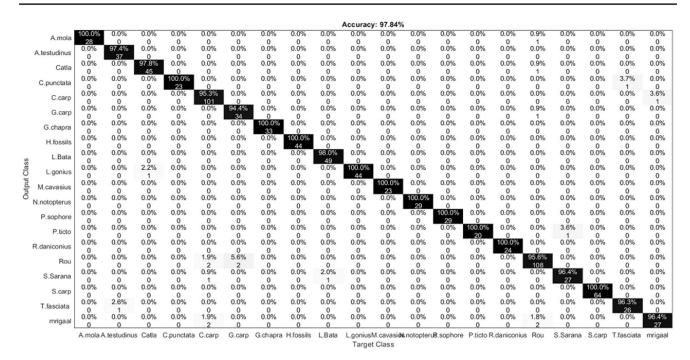


Fig. 17 CM of Alex-Net at Lr = .0001 and Weight = 30 = Bias factor on own dataset

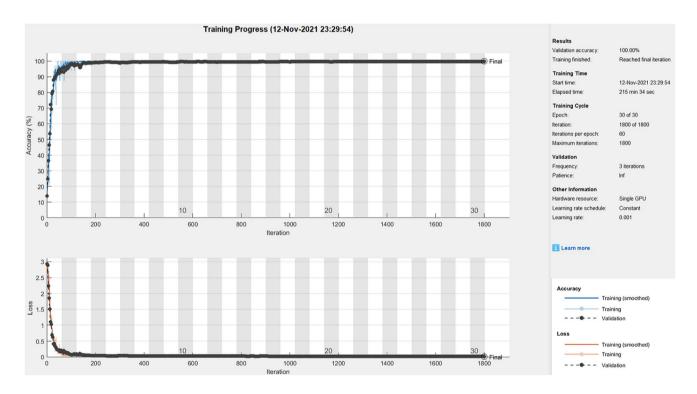


Fig. 18 Accuracy and loss curve (performance) of Resnet-50 at Lr = .001 and weight = 20 = bias factor on our dataset

Weight and Bias learning factor = 10. During the experiment, we observed that, performance measures at Lr = 0.01 for Alex-Net is below 40% for all weight and bias factors at

same training options that we feed to Resnet-50. Therefore, we totally discard results associated with this learning rate.

From Table 3, with our own dataset, the best classification (100%) performance measures by ResNet-50 are obtained at



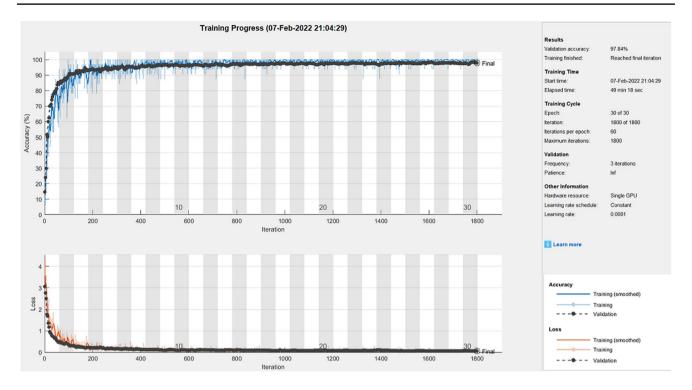


Fig. 19 Accuracy and loss curve (performance) of Alex-Net at Lr = .0001 and weight = 30 = bias factor on our dataset

Table 4 Mis-classification analysis with Resnet-50 and Alex-Net

Parameters	Weight and bias learning rate	Mis-classification on Fish_Pak dataset (validation image = 357)	Mis-classification on own data- set (validation image = 834)
Resnet-50	10	0	2
	20	2	0
	30	1	2
Alexnet	10	3	20
	20	7	24
	30	6	18

Lr=0.001 when we choose the Weight and Bias factor as 20. It also experiences the least validation and training loss with this training options. Alex-Net best performance is only 97.84%. validation accuracy at Lr=0.0001 with weight and bias learning factor=30. Table 4 gives the total misclassification numbers by both the models on the two datasets. So, from the above empirical analysis, we observe that the best classification parameters among the two classifiers are offered by ResNet-50 for both datasets at a learning rate of 0.001.

For both the datasets, it has been observed that training loss and validation loss in the case of ResNet-50 decrease first when we train the model from learning rate, Lr = 0.01 to Lr = 0.001, and it again increases for the learning rate, Lr = 0.0001. While in the case of Alex-Net, a decrease in learning rate increases the training and validation loss. Also, an increase in Weight and Bias learning factors increase the training and validation loss of the classifiers. From Tables 2 and 3, we can draw an inference that with a gradual increase in Weight and Bias learning factor, the loss function increases for both the datasets.

The performance of the proposed classification method is compared with the most recent works that dealt with freshwater fish species for classification and a detailed analysis is shown in Table 5. Those works use similar kinds of fish species that are common in South Asian countries.

We have considered multiple parameters for showing a detailed comparison of the proposed work with previous works. The highest classification accuracy on the Fish-Pak dataset is 98.8% [39] followed by 98.64% to date [42], whereas with our fine-tuned Resnet-50 network, we are able to reach 100% classification accuracy, precision, and recall on the Fish-Pak dataset. We also achieve a considerable improvement in terms of validation (2.38%) and training



Table 5 Performance comparison of the proposed model with state-of-the-art methods

Papers	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	Validation Loss (%)	Training Loss (%)	Dataset	No. of species
Islam et al. [35]	HLBP feature with SVM	94.78	Not men- tioned	Not men- tioned	Not men- tioned	Not men- tioned	BDIndig- enous- Fish2019	8
Rauf et al. [25]	32-layer VGG- Net, (Resnet-50)	98.5%, (93)	94.83, (89.88)	95.67, (90.17)	Not men- tioned	Not men- tioned	Fish_Pak	6
Sharmin et al. [26]	Handcrafted fea- tures + SVM- classifier	94.2%	93%	94.59	NA	NA	self-collected SIFS	6
Wang et al. [37]	SPP-densenet	97%	97.62%	99.2%	0.1(from graph)	Not men- tioned	Google Images	6
Banan et al. [38]	VGG-16	100%	Not men- tioned	Not men- tioned	0.0014	.0154	Own images of major carps	4
Iqbal et al. [39]	AlexNet	90.48%	Not men- tioned	Not men- tioned	Not men- tioned	Not men- tioned	LifeClef'15	6
Xu et al. [40]	Combined SE- Resnet-152	98.8%	Only per species is mentioned	Only per species is mentioned	NA	NA	Fish_Pak	6
Dey et al. [41]	5-layer CNN	99	99.01	99.01	.0325	.0167	BDIndig- enous- Fish2019	8
Abinaya et al. [42]	AlexNet with fuse Naive Bayesian layer	98.64	99.80	98.99	Not mentioned	Not mentioned	Fish_Pak	6
Jose et al. [43]	Octonion Net- work + soft- max classifier	98.17% (98.01%)	98.11% (98.05%)	98.13% (98.07%)	NA	NA	Fish_Pak (Self-col- lected Tuna fish data)	6 (3)
Proposed Work (Ours)	Fine-tuned Resnet-50	100	100	100	.0238	0.0003	Fish_Pak	6
Proposed Work (Ours)	Fine-tuned Resnet-50	100	100	100	.0206	.0004	Own dataset	20

loss (0.03%). Our analysis shows that this is the best result on the Fish_Pak dataset to this date.

Although a 100% classification accuracy is achieved by using VGG-16 [38], their target species is only 4 major carps. Another important thing we want to highlight in this paper is that the learning curve where the validation loss is lesser than the training loss which is considered an underfit problem. This is a case of an unrepresentative validation dataset [49–51] which means validation dataset is easier for the model to predict than the training data. Another possibility could be validation dataset considered in the simulation experiment is scarce but widely represented by the training dataset therefore the model behaves pretty well with those few validation data. Generally, a good training means validation loss should be slightly greater than the training loss. Also, it is advised to keep training the classifier if validation

loss < training loss. With our own dataset of 20 varieties of indigenous fish images, we achieve a 100% validation accuracy, precision and recall. We accomplish even less training and validation loss than Fish-Pak dataset on our simulation result. So, we can claim that our work is the first to report successful automatic classification of 20 varieties of fresh-water fish from India with best classification scores and minimum loss than any of the state-of-the-art methods.

Conclusions

The research article investigates different learning rate, weight and bias factor for achieving maximum classifier accuracy by ResNet-50 and Alex-Net on fish classification problem. From the performance analysis of these two deep-learning models,



we have observed that a 100% classification accuracy is achievable by Resnet-50 architecture at a learning rate of 0.001 on both the datasets under investigation. The proposed ResNet-50 model not only improves the target classification accuracy but reduces network under-fitting too. It also reveals that the misclassification rate by ResNet-50 is minimum compared to Alex-Net. The paper reports the best classification accuracy on the Fish-Pak dataset till date. The paper also reports maximum classification accuracy of 100% first time for 20-class classification model. The proposed method will help in accurate classification of carps and small indigenous fish species reared in one environment. With the help of this technique, fish farmers would be able to manage early segregation of the fish species according to their food preference and growth rate for better yield. Also, accurate classification of fish species can create awareness among fish loving locals of the North-East India in limiting over exploitation of some of the threatened fish species. In future, we would like to incorporate a live monitoring-based classification system to develop a robust and realistic classification model to help local and poor fish farmers for proper management of their fisheries as their livelihood depends on this.

Acknowledgements The authors are grateful to Mr. Balaram Mahalder (Technical Specialist, WorldFishCenter), Mr. Hamid Badar Osmany (Assistant Biologist, Marine Fisheries Department, Karachi, Pakistan), Mostafa A. R. Hossain (Professor, Professor, Aquatic Biodiversity & Climate Change, Department of Fish. Biology & Genetics, Bangladesh Agricultural University), Dr. Shamim Rahman(Asstt. Professor in Zoology, Devicharan Barua Girls' College, Jorhat, India), Dr. Dibakar Bhakta(Scientist (SS), Riverine and Estuaries Fisheries Division ICAR-Central Inland Fisheries Research Institute, Barrackpore, India), Dr. Mosaddequr Rahman(Kadai · Graduate School of Agriculture Forestry and Fisheries, Kagoshima University) and Bhenila Bailung(Research Scholar, Dibrugarh University) for allowing us to use their fish images from FishBase for initial simulation of the deep-learning networks. Due to the COVID-19 led pandemic and lockdown, our own fish image collections was hampered initially, but these individuals saved us by permitting us to use their collection of images which helped in initial planning of the classification model.

Author contribution statement JD: Conceptualization, Methodology, Software, Investigation, Writing—original draft, Writing—review & editing. Dr. SL: Supervision, Conceptualization. Dr. BB: Data collection and validation.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

FAO, Aquaculture development trends in Asia (2000). http://www.fao.org/3/ab980e/ab980e03.htm#TopOfPage

- 2. T.M. Berra, An Atlas of Distribution of the Fresh Water Fish Families of the World (University of Nebarska Press, Lincoln, 1981)
- D. Kumar, Fish culture in undrainable ponds. A manual for extension, FAO Fisheries Technical Paper No. 325. (Rome, FAO, 1992). 239 p
- M. Karim, H. Ullah, S. Castine et al., Carp—mola productivity and fish consumption in small-scale homestead aquaculture in Bangladesh. Aquac. Int. 25, 867–879 (2017). https://doi.org/10. 1007/s10499-016-0078-x
- B.K. Bhattacharjya, M. Choudhury, V.V. Sugunan, Ichthyofaunistic resources of Assam with a note on their sustainable utilization, in *Participatory approach for Fish Biodiversity Conservation in North East India*, ed. by P.C. Mahanta, L.K. Tyagi (Workshop Proc. NBFGR, Lucknow, 2003), pp. 1–14
- S. Dewan, M.A. Wahab, M.C.M. Beveridge, M.H. Rahman, B.K. Sarker, Food selection, electivity and dietary overlap among planktivorous Chinese and Indian major carp fry and fingerlings grown in extensively managed, rain-fed ponds in Bangladesh. Aquac. Res. 22, 277–294 (1991)
- M.M. Rahman, M.C.J. Verdegem, M.A. Wahab, M.Y. Hossain, Q. Jo, Effects of day and night on swimming, grazing and social behaviours of rohu *Labeo rohita* (Hamilton) and com-mon carp *Cyprinus carpio* (L.) in simulated ponds. Aquac. Res. 39, 1383– 1392 (2008)
- M.A. Wahab, M.M. Rahman, A. Milstein, The effect of common carp *Cyprinus carpio* (L.) and mrigal *Cirrhinus mrigala* (Hamilton) as bottom feeders in major Indian carp polycultures. Aquac. Res. 33, 547–557 (2002)
- D.M. Alam, M. Hasan, Md. Wahab, M. Khaleque, M. Alam, Md. Samad, Carp polyculture in ponds with three small indigenous fish species—Amblypharyngodon mola, Chela cachius and Puntius sophore. Progress. Agric. 13, 117–126 (2018)
- A.S.M. Kibria, M.M. Haque, Potentials of integrated multi-trophic aquaculture (IMTA) in freshwater ponds in Bangladesh. Aquac. Rep. 11, 8–16 (2018). https://doi.org/10.1016/j.aqrep.2018.05.004
- N.J.C. Strachan, Recognition of fish species by colour and shape. Image Vis. Comput. 11, 2–10 (1993)
- S. Cadieux, F. Michaud, F. Lalonde, Intelligent system for automated fish sorting and counting, Proceedings, in 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000) (Cat. No.00CH37113), vol. 2 (Takamatsu, Japan, 2000), pp. 1279–1284
- 13. A. Rova, G. Mori, L. M. Dill, One fish, two fish, butterfish, trumpeter: recognizing fish in underwater video, in *APR Conference on Machine Vision Applications* (2007), pp. 404–407
- C. Spampinato, D. Giordano, R. Di Salvo, J. Chen-Burger, R. Fisher, G. Nadarajan, Automatic fish classification for underwater species behavior understanding. Anal. Retr. Tracked Events Motion Imagery Streams (2010). https://doi.org/10.1145/18778 68.1877881
- M.K. Alsmadi, K.B. Omar, S.A. Noah et al., Fish classification based on robust features extraction from color signature using back-propagation classifier. J. Comput. Sci. 7, 52 (2011)
- B. Benson, J. Cho, D. Goshorn, R. Kastner, Field programmable gate array (FPGA) based fish detection using Haar classifiers. Am. Acad. Underwater Sci. (2009)
- J. Hu, D. Li, Q. Duan, Y. Han, G. Chen, X. Si, Fish species classification by color, texture and multi-class support vector machine using computer vision. Comput. Electron. Agric. 88, 133–140 (2012)
- M. M. M. Fouad, H. M. Zawbaa, N. El-Bendary, A. E. Hassanien, Automatic Nile Tilapia fish classification approach using machine learning techniques, in 13th International Conference on Hybrid Intelligent Systems (HIS 2013) (Gammarth, 2013), pp. 173–178. https://doi.org/10.1109/HIS.2013.6920477



- C. Pornpanomchai, B. Lurstwut, P. Leerasakultham, W. Kitiyanan, Shape- and texture-based fish image recognition system. Kasetsart J. Nat. Sci. 47, 624–634 (2013)
- M. Rodrigues, M. Freitas, F. Pádua, R. Gomes, E. Carrano, Evaluating cluster detection algorithms and feature extraction techniques in automatic classification of fish species. Pattern Anal. Appl. (2014). https://doi.org/10.1007/s10044-013-0362-6
- P.X. Huang, B.J. Boom, R.B. Fisher, Hierarchical classification with reject option for live fish recognition. Mach. Vis. Appl. 26, 89–102 (2015)
- M.-C. Chuang, J.-N. Hwang, K. Williams, A feature learning and object recognition framework for underwater fish images. IEEE Trans. Image Process. (2016). https://doi.org/10.1109/tip.2016.2535342
- D. Li, Qi. Wang, X. Li, M. Niu, He. Wang, C. Liu, Recent advances of machine vision technology in fish classification. ICES J. Mar. Sci. 79(2), 263–284 (2022). https://doi.org/10.1093/icesj ms/fsab264
- S.Z.H. Shah, H.T. Rauf, I.U. Lali, S.A.C. Bukhari, M.S. Khalid, M. Farooq, M. Fatima, Fish-Pak: fish species dataset from Pakistan for visual features based classification. Mendeley Data (2019). https://doi.org/10.17632/n3ydw29sbz.3
- H.T. Rauf, M.I. Lali, S. Zahoor, S.Z. Shah, A. Rehman, S.A.C. Bukhari, Visual features based automated identification of fish species using deep convolutional neural networks. Comput. Electron. Agric. (2019). https://doi.org/10.1016/j.compag.2019
- I. Sharmin, N.F. Islam, I. Jahan et al., Machine vision based local fish recognition. SN Appl. Sci. 1, 1529 (2019). https://doi.org/10. 1007/s42452-019-1568-z
- Z. Ju, Y. Xue, Fish species recognition using an improved AlexNet model. Optik 223, 165499 (2020). https://doi.org/10.1016/j.ijleo. 2020.165499
- A.A. dos Santos, W.N. Gonçalves, Improving Pantanal fish species recognition through taxonomic ranks in convolutional neural networks. Ecol. Inform. 53, 100977 (2019). https://doi.org/10.1016/j. ecoinf.2019.100977
- P. Mathew, S. Elizabeth, Fish identification based on geometric robust feature extraction from anchor/landmark points, in *National Conference on Image Processing and Machine Vision* (NCIPMV) 2017 (At University of Kerala, Trivandrum, 2017)
- J. Jäger, E. Rodner, J. Denzler, V. Wolff, K. Fricke-Neuderth, Seaclef 2016: Object proposal classification for fish detection in underwater videos, in *CLEF* (Working Notes) (2016), pp. 481–489
- S.A. Siddiqui, A. Salman, M.I. Malik, F. Shafait, A. Mian, M.R. Shortis, E.S. Harvey, Automatic fish species classification in underwater videos: exploiting pretrained deep neural network models to compensate for limited labelled data. ICES J. Mar. Sci. 75, 374–389 (2017)
- Y. Ma, P. Zhang, Y. Tang, Research on fish image classification based on transfer learning and convolutional neural network model 2018, in 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (2018), pp. 850–855. https://doi.org/10.1109/FSKD.2018.8686892
- A. Salman, S. Maqbool, A.H. Khan, A. Jalal, F. Shafait, Realtime fish detection in complex backgrounds using probabilistic background modelling. Ecol. Inform. 51, 44–51 (2019)
- N.E. Khalifa, M. Taha, A.E. Hassanien, Aquarium family fish species identification system using deep neural networks, in *Proceedings of the International Conference on Advanced Intelligent* Systems and Informatics 2018, (2019), pp. 347–356. https://doi. org/10.1007/978-3-319-99010-1_32
- M. A. Islam, M. R. Howlader, U. Habiba, R. H. Faisal, M. M. Rahman, Indigenous fish classification of Bangladesh using hybrid features with SVM classifier. 2019, in *International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering* (IC4ME2) (2019). https://doi.org/10.1109/ic4me247184.2019.9036679

- A. Jalal, A. Mian, M. Shortis, F. Shafait, Fish detection and species classification in underwater environments using deep learning with temporal information. Ecol. Inform. 57, 101088 (2020). https://doi.org/10.1016/j.ecoinf.2020.101088
- H. Wang, Y. Shi, Y. Yue, H. Zhao, Study on freshwater fish image recognition integrating SPP and DenseNet network. 2020, in *IEEE International Conference on Mechatronics and Automation* (ICMA) (2020). https://doi.org/10.1109/icma49215.2020.923
- A. Banan, A. Nasiri, A. Taheri-Garavand, Deep learning-based appearance features extraction for automated carp species identification. Aquac. Eng. 89, 102053 (2020)
- M.A. Iqbal, Z. Wang, Z.A. Ali et al., Automatic fish species classification using deep convolutional neural networks. Wireless Pers. Commun. 116, 1043–1053 (2021). https://doi.org/10.1007/s11277-019-06634-1
- X. Xu, W. Li, Q. Duan, Transfer learning and SE-ResNet152 networks-based for small-scale unbalanced fish species identification. Comput. Electron. Agric. 180, 105878 (2021). https://doi.org/10.1016/j.compag.2020.105878
- K. Dey, M.M. Hassan, M.M. Rana, M.H. Hena, Bangladeshi indigenous fish classification using convolutional neural networks. Int. Conf. Inf. Technol. (ICIT) 2021, 899–904 (2021). https://doi. org/10.1109/ICIT52682.2021.9491681
- N.S. Abinaya, D. Susan, R. Sidharthan, Naive Bayesian fusion based deep learning networks for multisegmented classification of fishes in aquaculture industries. Ecol. Inform. 61, 101248 (2021). https://doi.org/10.1016/j.ecoinf.2021.101248
- J. Jose, C.S. Dr. Kumar, S. Sureshkumar, Region-based split Octonion networks with channel attention module for tuna classification. Int. J. Pattern Recognit. Artif. Intell. (2022). https://doi.org/10.1142/S0218001422500306
- M.K. Alsmadi, I. Almarashdeh, A survey on fish classification techniques. J. King Saud Univ. Comput. Inf. Sci. 34(5), 1625– 1638 (2022). https://doi.org/10.1016/j.jksuci.2020.07.005
- M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex. Nat. Neurosci. 2, 1019–1025 (1999). https://doi.org/10.1038/14819
- S.J. Pan, Q. Yang, A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22(10), 1345–1359 (2010)
- A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet classification with deep convolutional neural networks. Neural Inf. Process. Syst. (2012). https://doi.org/10.1145/3065386
- K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV, USA, 27–30 June 2016), pp. 770–778
- T. Mitchell, Machine Learning (McGraw-Hill Science/Engineering/Math, Berlin, 1997)
- K. Horak, Introduction to Learning Curves. http://vision.uamt. feec.vutbr.cz/STU/lectures/MachineLearning_LearningCurves. Accessed 08 February (2022)
- https://www.baeldung.com/cs/learning-curve-ml. Accessed 08 Feb 2022

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

