Fish survival prediction in an aquatic environment using random forest model

Md. Monirul Islam[1], Mohammod Abul Kashem[2], Jia Uddin[3]

[1,2]Department of Computer Science and Engineering, Dhaka University of Engineering and Technology, Gazipur, Bangladesh
[3]Technology Studies Department, Endicott College, Woosong University, Daejeon, South Korea

Article Info	ABSTRACT
	In the real world, it is very difficult for fish farmers to select the perfect fish species for aquaculture in a specific aquatic environment. The main goal of this research is to build a machine learning that can predict the perfect fish species in an aquatic environment. In this paper, we have utilized a model using random forest (RF). To validate the model, we have used a dataset of aquatic environment for 11 different fishes. To predict the fish species, we utilized the different characteristics of aquatic environment including pH, temperature, and turbidity. As a performance metrics, we measured accuracy, true positive (TP) rate, and kappa statistics. Experimental results demonstrate that the proposed RF-based prediction model shows accuracy 88.48%, kappa statistic 87.11% and TP rate 88.5% for the tested dataset. In addition, we compare the proposed model with the state-of-art models J48, RF, k-nearest neighbor (k-NN), and classification and regression trees (CART). The proposed model outperforms than the existing models by exhibiting the higher accuracy score, TP rate and kappa statistics.

Corresponding Author:

Jia Uddin
Technology Studies Department
Endicott College, Woosong University
South Korea
Email: jia.uddin@wsu.ac.kr

1. INTRODUCTION

Aquaculture refers to the farming of aquatic animals or plants primarily for food. It contains the breeding, nurture, and reaping of fish, mollusks, crustaceans, and plants in fresh and saltwater environments. The practice was initiated in China about 4,000 years ago and global production remains to be subjugated by China and other Asian countries. Aquaculture is used to harvest food by some of the deprived communities everywhere on the globe as well as by key corporations. Globally, aquaculture by now supplies more than half of all seafood used up by humans, a proportion that continues to rise as the world population produces. According to the Food and Agricultural Organization (FAO) [1], 3 million tons of food were produced by aquaculture in the 1970s, a figure that rose steadily to over 80 million tons in 2017.

Manually fish classification is a very complex and tedious assignment for these who are now not specialists. Fish species are concerned in many industrial and agricultural industries, as nicely as the manufacture of foodstuffs and used as food that is very vital to humans [2]. As marine biologists classify fish from their traits and also used the classification tree in the classification of fish, which led them to use laptops gaining knowledge of and structures in the data, which saved time, effort, and velocity in the classification of fish [3].

Fish classification can be the identification of fish species, depending on their physiognomies or similarities. Also, it can be described as the technique of determining the types of fish [4]. Classification of fish is critical for numerous reasons, inclusive of sample and subsistence matching extraction feature, identification of physical or behavioral characteristics, statistical control and high-quality utilized to fish of all kinds [5]. Moreover, fish classification is regarded as a vital venture for fishing and population assessments [6].

On the other hand, computerized fish classification can speed up the technique and can improve the accuracy of classification or identification of fish species. Several tactics are introduced in the literature for computerized fish species identification. In this paper, we did classification using machine learning model including decision tree classifier (J48), random forest (RF), k-nearest neighbor (k-NN), and classification and regression tree (CART). Classification has used for prediction purposes; traditional rule-based algorithm does not provide any prediction feature for the unknown dataset. Confusion matrix provides various measurement of accuracy in prediction, where rule-based algorithm cannot perform this [7]. CNN is a deep learning model where computation complexity is higher than machine learning models. In this paper, we have considered the machine learning algorithms only due to its less computational complexity. In the CNN, we need much training time than traditional machine learning models.

In this paper, we proposed a fish survival prediction in an aquatic environment based on the RF model. For the rest of the paper, we organize as shown in section 2 states the literature review. In section 3, the proposed model is discussed. Section 4 depicts the experimental setup and result from the analysis. Finally, the findings of this paper are discussed in section 5.

2. LITERATURE REVIEW

The literature states a portion of activities related to decision support systems in aquaculture garden operations. Several decision support systems have been developed. Some of them use machine learning methods and others do not. An automatic fish identification is proposed where shade and texture features are extracted from the fish images [8]. A structure is introduced using the real-time water quality indicators and operational information, where impact on survival rate, biomass, and production failure of aquaculture species are evaluated [9]. A prediction model using one feature of water called DO is presented for the aquatic creature [10]. A hardware is made for monitoring water quality factors including pH, temperature, and dissolved oxygen [11]. An IoT device is proposed for detecting and controlling the water factors including pH, temperature; however, they did not analyze the data [12]. A regression model is utilized for predicting water quality of cultivating fish; however, they did not consider the prediction accuracy [13]. An automated strategy is developed for fish identification primarily based on the use of aid vector desktop and k-means clustering algorithm [14]. A computerized robust Nile-Tilapia fish classification approach is proposed in [15], where the scale-invariant characteristics of fish's change are extracted. Then, these points are used to feed the support vector machine.

Managing hatchery production is focused using rules and calculations of physical, chemical, and biological processes [16]. A scientific model is developed to evaluate environmental impact [17]. A rule is hand-crafted by domain experts [18]. A machine learning method is presented to obtain a balance between the farm closure and the farm opening events [19]. A feature ranking algorithm is displayed to identify the most influential cause of the closure [20]. Time series machine learning approaches is adopted like principal component analysis (PCA) and auto correlation function (ACF) to predict the closure event [21]. A set of rules are extracted from data gathered by sensor networks to find associations between environmental variables and algae growth [22]. An ensemble method is designed to find the relevant environmental variables responsible for algae growth and the growth prediction [23]. A machine learning method is developed to predict the propagation of algae patches along the waterway [24].

3. PROPOSED MODEL

Figure 1 shows a detailed block diagram of the proposed model. At first, we import our dataset. In the preprocessing section, we filter and resample for our dataset. Then we select our model as RF classifiers in the classification section. We classify our various machine learning models here. After classification, classifier output is predicted.

3.1. Description of dataset

The data used in this study involving parameters of an aquatic environment for fish farming taken from the University of Dhaka, Faculty of Fisheries, Dhaka, Bangladesh. There are 191 instances of 4 attributes. Attributes are pH, temperature, turbidity, and fish. We choose pH, temperature, turbidity as feature attributes and fish as target attribute. The dataset is partitioned into two parts. One is aquatic environment

characteristics and another is fish species. The detailed of target attribute is of 11 fish species including katla 14 images, shing 17 images, prawn 14 images, rui 19 images, koi 15 images, pangas 22 images, tilapia 25 images, silver carp 7 images, karpio 33 images, magur 11 images and shrimp 14 images.

Aquatic environment characteristics: We utilized pH, temperature, and turbidity as aquatic environment parameters in our study.

− pH: pH is necessary for aquaculture as a measure of the acidity of the water or soil. The optimal pH for fish is between 6.5 and 9. Fish will grow poorly, and reproduction will be affected at consistently greater or lower pH tiers [25]. The pH level for warm-water pond fish is 4 for acid death point, 4 to 5 for no reproduction, 5 to 6.5 for slow growth, 6.5 to 8.5 for desirable ranges, 9 to 10 for slow growth, and ≥11 for alkaline death point.

− Temperature: The increase and endeavor of the fish rely on their physique temperature. The body temperature of the fish is about the same as the water temperature and varies with it. Each fish species is tailored to develop and reproduce inside well-defined stages of water temperatures, but the most useful boom and replica take area within narrower tiers of temperature. It is important, therefore, to understand the water temperatures reachable at your fish farm nicely to pick out the right species of fish and to graph its management as a result. Table 1 shows the thermal range of some common fish species [26].

− Turbidity: The ability of water to transmit the light that restricts light penetration and limit photosynthesis is termed as turbidity and is the resultant impact of several elements such as suspended clay particles, dispersion of plankton organisms, particulate natural things and also the pigments caused with the aid of the decomposition of organic matter. Acceptable turbidity varies from 30-80 cm is properly for fish health [27].

− Fish species: In our dataset, we utilized a total of 11 fish species as the target variable. The fish species in our dataset are presented in Figure 2; where carpio fish is shown in Figure 2(a), katla fish is in Figure 2(b), rui fish is in Figure 2(c), koi fish is in Figure 2(d), magur fish is in Figure 2(e), pangas fish is in Figure 2(f), prawn fish is in Figure 2(g), silver carp fish is in Figure 2(h), tilapia fish is in Figure 2(i), and shing fish is in Figure 2(j).
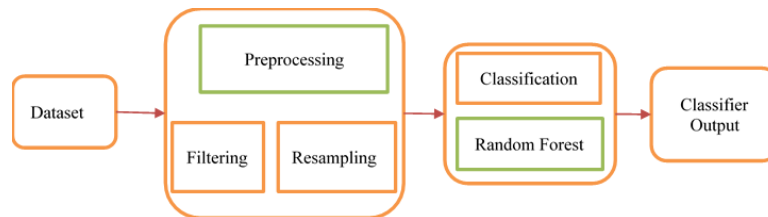

Figure 1. Block diagram of proposed model

Table 1. Thermal range of some common fish species (in °C)

Fish species	Dangerous pond-water temperature lower-upper limit		Optimum thermal range for adults	Thermal range for spawning
Carpio	2	36	23-26 (25)	Above 18
Katla	15	34	26-29	22-28
Bighead carp	5	37	23-31	17-30

(a)   (b)   (c)   (d)   (e)

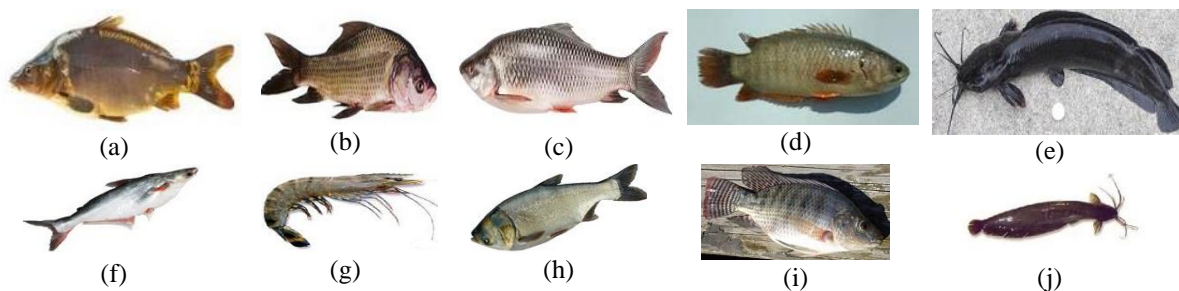(f)   (g)   (h)   (i)   (j)

Figure 2. Sample fishes: (a) carpio fish, (b) katla fish, (c) rui fish, (d) koi fish, (e) magur fish, (f) pangas fish, (g) prawn fish, (h) silver carp fish, (i) tilapia fish, and (j) shing fish

3.2. Preprocessing

In the preprocessing step, we filtered our dataset using a resampling option for observing the current relation of instances and attributes of the dataset. In the attribute selection window, we can check the missing, unique, and distinct value of each attribute. All attributes show 0% missing and pH has 28 unique values, temperature has 22 unique values, turbidity has 56 unique and fish has 11 distinct values.

3.3. Classification

In the classification section, we classified our dataset using 5 various classifiers model. RF outperforms the other described model.

3.3.1. Random forest

RF is a supervised learning method that is a decision tree-based algorithm. As the name proposes as forest the RF classifier is an ensemble of decision trees wherever a random vector sample produce each classifier from the input vector [28] and every tree cast a unit vote for the most popular class to classify an input vector, nearly all of the time trained with a bagging method.

The preparation calculation for RF applies the overall strategy of bootstrap collecting, or packing, to tree students. Given a preparation set $X = x_1, ..., x_n$ with reactions $Y = y_1, ..., y_n$, stowing more than once (A times) chooses an irregular example with substitution of the preparation set and fits trees to these examples. For a=1, ……, A:

−   Test, with substitution, n preparing models from X, Y; call these X_a, Y_a.
−   Train a characterization or relapse tree f_a on X_a, Y_a.

After preparing, expectations for concealed examples x' can be made by averaging the forecasts from all the individual relapse trees on x':

$$\hat{f} = \frac{1}{A}\sum_{a-1}^{A} f_a(x') \tag{1}$$

also, a gauge of the vulnerability of the forecast can be made as the standard deviation of the expectations from all the individual relapse trees on x':

$$\sigma = \sqrt{\frac{\sum_{a-1}^{A}\left(f_a(x') - \hat{f}\right)^2}{A-1}} \tag{2}$$

The universal thought of the bagging method is that the composing of the learning method increases the overall result. The RF is less sensitive than other streamline machine learning classifiers to overfitting and to the quality of training samples [29]. Figure 3 shows the concept of RF model. Tree 1 and Tree 2 belong to Class A. So, predicted output will be Class A. Majority vote is Class A in Figure 3.
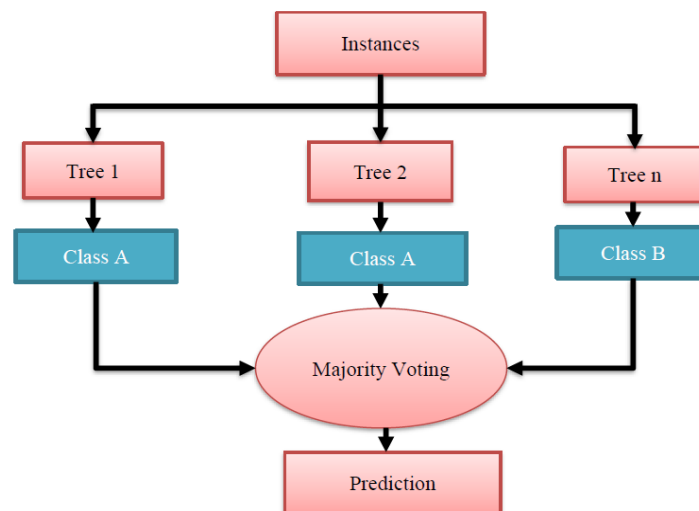

Figure 3. RF model

3.4. Classifier output

In the classifier section, we can see the result performance of our model and other state-of-art models. By choosing our described model, we can check results. In this section, we can see detailed accuracy by class. Figure 4 shows these performance results. We did not find any machine learning model for fish environment monitoring using RF. The dataset we have used in our own dataset. Figure 4 presents average true positive (TP) rate as 0.885, FP rate as 0.013, precision as 0.890, recall as 0.885, F-measure as 0.879, MCC as 0.871, ROC area as 0.981, PRC Area as 0.929, Correctly Classified Instances as 88.48%, Incorrectly classified instances as 11.52%, Kappa statistics as 0.87, mean absolute error as 0.04, root mean squared error as 0.13, relative absolute error as 24.53%, Root relative squared error as 45.46%.

```
Correctly Classified Instances        169              88.4817 %
Incorrectly Classified Instances       22              11.5183 %
Kappa statistic                          0.8711
Mean absolute error                      0.04
Root mean squared error                  0.1298
Relative absolute error                 24.5218 %
Root relative squared error             45.4566 %
Total Number of Instances              191

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
              0.929    0.011    0.867      0.929   0.897      0.889  0.954     0.924     katla
              0.765    0.011    0.867      0.765   0.813      0.797  0.985     0.889     sing
              0.929    0.006    0.929      0.929   0.929      0.923  0.958     0.934     prawn
              0.895    0.012    0.895      0.895   0.895      0.883  0.988     0.953     rui
              0.933    0.023    0.778      0.933   0.848      0.838  0.979     0.893     koi
              0.864    0.030    0.792      0.864   0.826      0.803  0.983     0.920     pangas
              0.920    0.006    0.958      0.920   0.939      0.930  0.996     0.978     tilapia
              0.286    0.000    1.000      0.286   0.444      0.527  0.847     0.406     silverCup
              0.939    0.019    0.912      0.939   0.925      0.910  0.997     0.986     karpio
              0.909    0.006    0.909      0.909   0.909      0.904  0.998     0.981     magur
              1.000    0.006    0.933      1.000   0.966      0.963  1.000     1.000     shrimp
Weighted Avg. 0.885    0.013    0.890      0.885   0.879      0.871  0.981     0.929
```

Figure 4. Classifier output of our model

4. EXPERIMENTAL SETUP AND RESULT ANALYSIS

As data analysis, we have used WEKA tool for classifying the proposed model and described other models. The tool is very helpful to analyze and has various techniques embedded in it. We have used 10% images for testing and 90% images for training in each species for all described model.

4.1. Performance metrics

Performance parameters are the most important metrics to compare among classifier methods to get the best classifier. We have applied 3 performance parameters which are accuracy, true positive (TP) rate and kappa statistics. The parameter is calculated from a confusion matrix which is situated in every step of classification. Accuracy is measured by dividing the total number of correctly classified instances by the total number of instances and also it is measured by confusion matrix which is mathematically counted by (4). TP rate is another performance metric of our study and it is calculated by (3). And kappa statistic is the last metric of our paper which is computed by (5). The higher the kappa statistics, the better the model accuracy level. A general view of the confusion matrix is illustrated in Table 2.

Table 2. Confusion matrix

	Predicted Yes	Predicted No
Actual Yes	TP	FN
Actual No	FP	TN

Here, TP signifies the number of properly classified positive occurrences.

$$TP\ Rate = \frac{TP}{FN + TP} \tag{3}$$

It is also known as the recall. It tells us what percentage of positive instances have been correctly identified.
− FP signifies the number of misclassified positive occurrences.
− FN signifies the number of misclassified negative occurrences.
− TN signifies the number of properly classified negative occurrences.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Accuracy is also represented by total accuracy.

$$Kappa\ statistic = \frac{Total\ accuracy - random\ accuracy}{1 - random\ accuracy} \tag{5}$$

where

$$Random\ accuracy = \frac{(TN + FP) \times (TN + FN) + (FN + TP) \times (FP + TP)}{(TP + TN + FP + FN) \times (TP + TN + FP + FN)} \tag{6}$$

We have used Waikato environment for knowledge analysis (WEKA) for processing data. The proposed model, RF shows the accuracy as the value 88.4817%, the average TP rate as the weight of 88.5% and kappa statistic as the standard of 87.11%. We can say, these three metrics give a better result. We have compared the performance metrics with our proposed model and other state-art-models. We utilized 5 models in our experimental work. They are RF, J48, Naïve Bayes, k-NN, and CART. Table 3 depicted a detailed comparison with all model each other.

Table 3 shows that RF gives the highest score of every metric as accuracy 88.48%, kappa statistic as 87.11%, and TP rate as 88.5%. The second highest score belongs to the k-NN model which tells accuracy as 85.79%, kappa statistic as 84.05% and TP rate as 85.8%. J48 acquires 3rd highest position by achieving an accuracy as 73.16%, kappa statistic as 69.88% and TP rate as 73.2%. CART has 4th place in scoring performance metrics by getting accuracy as 64.21%, kappa statistic as 59.80 and TP rate as 64.2%.

Naïve Bayes (NB) gives the lowest score by acquiring accuracy as 56.84%, kappa statistic as 51.60% and TP rate as 56.88%. NB provides the lowest performance. Because NB classifies only 108 images correctly among 191 images and cannot classify in silver cup fish. We know, NB is probabilistic machine learning algorithm and it studies that the features are free of each other. It also gives lower accuracy than other classifier models. However, in real world, features depend on each other. If we add multiple classifiers in the model, the computational complexity will be higher and for our tested dataset, we already have a significant result for our model.

Table 3. Comparison among classification model based on performance metrics

S.L. No.	Machine learning model	Accuracy (%)	Kappa statistic (KS) (%)	Avg. TP rate (%)	Remarks
1	RF (Proposed Model)	88.48	87.11	88.5	Highest
2	J48 [30]	73.16	69.88	73.2	3rd Highest
3	NB [31]	56.84	51.60	56.8	Lowest
5	k-NN [32]	85.79	84.05	85.8	2nd Highest
6	CART [33]	64.21	59.80	64.2	4th Highest

These performance metrics are shown in Figure 5 graphically. We marked three colored curves for three performance metrics. The blue curve is marked as an accuracy metric. The middle curve is identified for kappa statistic which is maroon color and the green curve is noticed for TP rate. We can see from this, the proposed model, RF gives the highest score in all categories of performance metrics. All circle point for RF model has the top position in performance metrics as accuracy 88.48%, KS as 87.11% and TP rate as 88.5%.
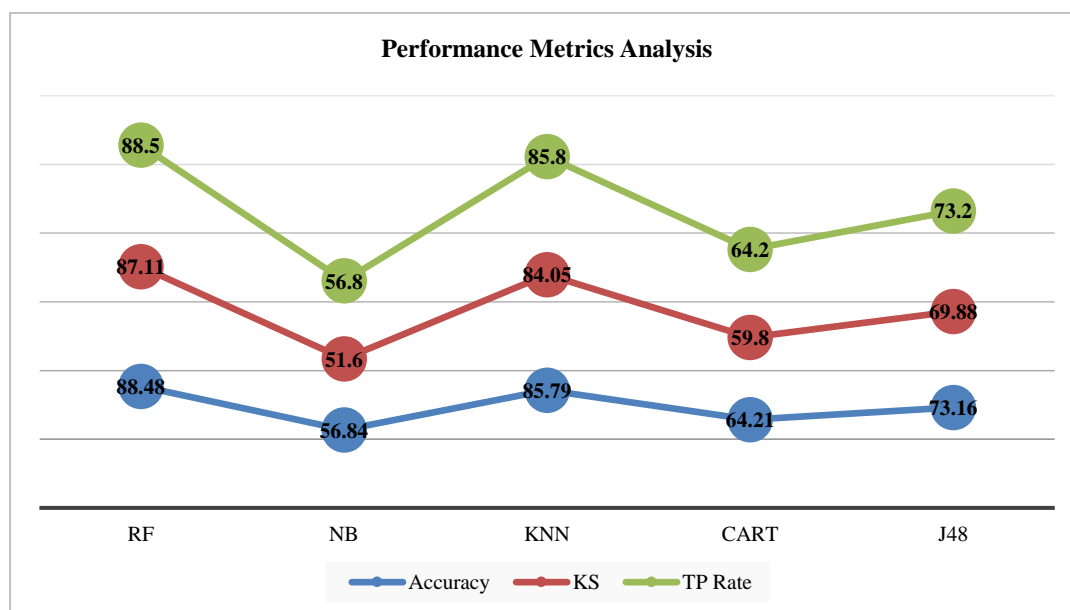
Figure 5. Accuracy analysis

5. CONCLUSION

We conducted this research to find out the best prediction model for fish farmers in an aquatic environment using various aquatic parameters. We used pH, temperature, turbidity, and fish as parameters of the dataset where we marked temperature, pH, turbidity as feature variables and fish as the target variable. We used total 11 types of fish. They are katla, shing, prawn, rui, koi, pangas, tilapia, silver carp, carpio, magur and shrimp. We find out the accuracy, kappa statistic and TP rate as performance metrics. We analyzed a total of five supervised machine models. They are RF, NB, k-NN, CART and J48. Among these models, our proposed model, RF shows the best accuracy, kappa statistic and TP rate as performance metrics that can predict the most fish species in an aquatic environment. RF provides accuracy 88.48%, KS 87.11% and TP rate 88.5%. Further, the research scope can be defined by enriching the dataset by more observation and testing with artificial neural network.

REFERENCES

[1] FAO, The State of World Fisheries and Aquaculture 2020-sustainability in action, 2020, FAO, Italy, 2020, pp. 0-244, http://www.fao.org/3/ca9231en/CA9231EN.pdf.

[2] M. K. Alsmadi, K. B. Omar, S. A. Noah, and A. I. Almarashdeh, "Fish Recognition Based on Robust Features Extraction from Size and Shape Measurements Using Neural Network," *Journal of Computer Science*, vol. 6, no. 10, pp. 1088-1094, 2010, doi: 10.3844/jcssp.2010.1088.1094.

[3] T. H. Hoang, K. Lock, A. Mouton, and P. L.M.Goethals, "Application of classification trees and support vector machines to model the presence of macroinvertebrates in rivers in Vietnam," *Ecological Informatics*, vol. 5, no. 2, 2010, pp. 140-146, 2010, doi: 10.1016/j.ecoinf.2009.12.001.

[4] B. Benson, J. Cho, D. Goshorn, and R. kastner, "Field programmable gate array (FPGA) based fish detection using haar classifiers," *American Academy of Underwater Sciences*, Georgia, USA, 2009, pp. 1-8.

[5] S. Bermejo, "Fish age classification based on length, weight, sex and otolith morphological features," *Fisheries Research*, vol. 84, no. 2, pp. 270-274, 2007, doi: 10.1016/j.fishres.2006.12.007.

[6] A. G. Cabreira, M. Tripode, and A. Madirolas, "Artificial neural networks for fish-species identification," *ICES Journal of Marine Science*, vol. 66, no. 6, pp. 1119-1129, 2009, doi: 10.1093/icesjms/fsp009.

[7] I. Khan, X. Zhang, M. Rehman, and R. Ali, "A Literature Survey and Empirical Study of Meta-Learning for Classifier Selection," *IEEE Access*, vol. 8, pp. 10262-10281, 2020, doi: 10.1109/ACCESS.2020.2964726.

[8] J. Hu, D. Li, Q. Duan, Y. Han, G. Chen, and X. Si, "Fish species classification by color, texture and multi-class support vector machine using computer vision," *Computers and Electronics in Agriculture*, vol. 88, pp. 133-140, 2012, doi: 10.1016/j.compag.2012.07.008.

[9]   G. Bourke, F. Stagnitti, and B. Mitchell, "A decision support system for aquaculture research and management," *Aquacultural Engineering*, vol. 12, no. 2, pp. 111-123, 1993, doi: 10.1016/0144-8609(93)90020-C.

[10]   W. Li, H. Wu, N. Zhu, Y. Jiang, J. Tan, and Y. Guo, "Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU)," *Information Processing in Agriculture*, vol. 8, no. 1, pp. 185-193, 2021, doi: j.inpa.2020.02.002.

[11]   G. A. Defe and A. Z. C. Antonio, "Multi-parameter Water Quality Monitoring Device for Grouper Aquaculture," in *IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management*, 2018, pp. 1-5, doi: 10.1109/HNICEM.2018.8666414.

[12]   T. Abinaya, J. Ishwarya, and M. Maheswari, "A Novel Methodology for Monitoring and Controlling of Water Quality in Aquaculture using Internet of Things (IoT)," in *2019 International Conference on Computer Communication and Informatics*, 2019, pp. 1-4, doi: 10.1109/ICCCI.2019.8821988.

[13]   P.-Y. Yang, J.-T. Tsai, J.-H. Chou, W.-H. Ho, and Y.-Y. Lai, "Prediction of water quality evaluation for fish ponds of aquaculture," in *2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2017, pp. 545-546, doi: 10.23919/SICE.2017.8105455.

[14]   S. O. Ogunlana, O. Olabode, S. Oluwadare, and G. Iwasokun, "Fish classification using support vector machine," *African Journal of Computing and ICT*, vol. 8, no. 2, 2006, pp. 1-8.

[15]   M. M. M. Fouad, H. M. Zawbaa, N. El-Bendary, and A. E. Hassanien, "Automatic Nile Tilapia fish classification approach using machine learning techniques," in *13th International Conference on Hybrid Intelligent Systems*, 2013, pp. 173-178, doi: 10.1109/HIS.2013.6920477.

[16]   D. H. Ernst, J. P. Bolte, and S. S. Nath, "AquaFarm: simulation and decision support for aquaculture facility design and management planning," *Aquacultural Engineering*, vol. 23, no. 1-3, pp. 121-179, 2000, doi: 10.1016/S0144-8609(00)00045-5.

[17]   W. Silvert, "Decision support systems for aquaculture licensing," *Journal of Applied Ichthyology*, vol. 10, no. 4, pp. 307-311, 1994, doi: 10.1111/j.1439-0426.1994.tb00170.x.

[18]   H. Halide, A. Stigebrandt, M. Rehbein, and A. D. McKinnon, "Developing a decision support system for sustainable cage aquaculture," *Environmental Modelling & Software*, vol. 24, no. 6, pp. 694-702, 2009, doi: 10.1016/j.envsoft.2008.10.013.

[19]   C. D'Este, A. Rahman, and A. Turnbull, "Predicting Shellfish Farm Closures with Class Balancing Methods," *Lecture Notes in Computer Science*, pp. 39-48, 2012, doi: 10.1007/978-3-642-35101-3_4.

[20]   A. Rahman, C. D'Este, and J. McCulloch, "Ensemble Feature Ranking for Shellfish Farm Closure Cause Identification," in *Workshop on Machine Learning for Sensory Data Analysis*, 2013, pp. 13-18, doi: 10.1145/2542652.2542655.

[21]   Md. S. Shahriar, A. Rahman, and J. McCulloch, "Predicting shellfish farm closures using time series classification for aquaculture decision support," *Computers and Electronics in Agriculture*, vol. 102, pp. 85-97, 2014, doi: 10.1016/j.compag.2014.01.011.

[22]   M. S. Shahriar, C. D'Este, and A. Rahman, "On detecting and predicting harmful algal blooms in coastal information systems," in *2012 Oceans - Yeosu*, 2012, pp. 1-3, doi: 10.1109/OCEANS-Yeosu.2012.6263406.

[23]   A. Rahman and M. S. Shahriar, "Algae growth prediction through identification of influential environmental variables: A machine learning approach," *International Journal of Computational Intelligence and Applications*, vol. 12, no. 2, pp. 1350008-19, 2013, doi: 10.1142/S1469026813500089.

[24]   M. S. Shahriar and A. Rahman, "Spatial-temporal prediction of algal bloom," in *Ninth International Conference on Natural Computation*, 2013, pp. 973-977, doi: 10.1109/ICNC.2013.6818117.

[25]   M. M. Islam, M. A. Kashem, and F. I. Jui, "Aqua fishing monitoring system using IoT devices," *International Journal of Innovative Science, Engineering and Technology*, vol. 6, no. 11, 2019, pp. 108-114.

[26]   [online].   http://www.fao.org/fishery/docs/CDrom/FAO_Training/FAO_Training/General/x6709e/x6709e02.htm Accessed 4 April 2021.

[27]   A. Bhatnagar and P. Devi, "Water quality guidelines for the management of pond fish culture," *International Journal of Environmental Sciences*, vol. 3, no. 6, 2008, pp. 1980-2009.

[28]   Z. Ferdoush, B. N. Mahmud, A. Chakrabarty, and J. Uddin, "A short-term hybrid forecasting model for time series electrical-load data using random forest and bidirectional long short-term memory," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 1, pp. 763-771, Feb. 2021, doi: 10.11591/ijece.v11i1.pp763-771.

[29]   M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24-31, 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[30]   G. Kaur and A. Chhabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes," *International Journal of Computer Applications*, vol. 98, no. 22, pp. 13-17, Jul. 2014, doi: 10.5120/17314-7433.

[31]   T. Kim, B. D. Chung, and J.-S. Lee, "Incorporating receiver operating characteristics into naive Bayes for unbalanced data classification," *Computing*, vol. 99, no. 3, pp. 203-218, 2016, doi: 10.1007/s00607-016-0483-z.

[32]   M. M. Islam, J. Uddin, M. A. Kashem, F. Rabbi, and Md. W. Hasnat, "Design and Implementation of an IoT System for Predicting Aqua Fisheries Using Arduino and KNN," *Intelligent Human Computer Interaction*, pp. 108-118, 2021, doi: 10.1007/978-3-030-68452-5_11.

[33]   L. Rutkowski, L. Pietruczuk, P. Duda, and M. Jaworski, "Decision Trees for Mining Data Streams Based on the McDiarmid's Bound," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1272-1279, 2013, doi: 10.1109/TKDE.2012.66.

BIOGRAPHIES OF AUTHORS

Md. Monirul Islam is an M.Sc. Student, Department of Computer Science and Engineering (CSE), Dhaka University of Engineering and Technology (DUET), Gazipur-1700, Bangladesh. He has completed B.Sc. in engineering degree in CSE in the year 2017 from Pabna University of Science and Technology (PUST), Pabna-6600, Bangladesh. Currently, he is at Uttara University (UU) as a full-time lecturer since 3rd February 2020. Field of interest: Machine learning/Deep Learning, Data Science.

Mohammod Abul Kashem has been serving as a professor, Department of Computer Science and Engineering (CSE) and Additional Director, Planning & Development, Dhaka University of Engineering & Technology (DUET), Gazipur-1700, Bangladesh. He got Post Doctorate (ULL, France), PhD (LPNU, Ukraine), M.Eng. & B.Eng. (SULP, Ukraine). Field of interest: Internet of Things, Cyber Physical System, The Challenges, Approaches and Used Techniques of CPS for Manufacturing in Industry 4.0, CPS of collaborative value network for manufacturing, multi agent system for industry and critical mission.

Jia Uddin has been serving as an Assistant Professor, Department of Technology studies, Endicott College, Woosong University, Daejeon, South Korea. He received Ph.D. in Computer Engineering from University of Ulsan, South Korea, M.Sc. in Electrical Engineering (Specialization: Telecommunications), Blekinge Institute of Technology, Sweden, and B.Sc. in Computer and Communication Engineering, International Islamic University Chittagong, Bangladesh. He was a visiting faculty at School of computing, Staffordshire University, United Kingdom. He is an Associate Professor (now on leave), Department of Computer Science and Engineering, Brac University, Dhaka, Bangladesh. His research interests are Industrial Fault Diagnosis, Machine Learning/Deep Learning based prediction and detection using multimedia signals.