International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

STRATIFICATION OF FISH SPECIES USING COMPARATIVE MACHINE LEARNING ALGORITHMS

Mr. R.P.Selvam, Dr. R. Devi

Research Scholar, Department of Computer Science,

Vels Institute of Science, Technology & Advanced Studies (VISTAS)

Pallavaram, Chennai-600 117. E-mail id : selvam.periyasamy@gmail.com

Orcid: 0009-0002-9342-383X

Professor & Head, Department of Computer Science,

Vels Institute of Science, Technology & Advanced Studies (VISTAS)

Pallavaram, Chennai-600 117. E-Mail id: devi.scs@vistas.ac.in

Orcid: 0000-0002-8951-2242

Abstract

The fisheries industry and aqua-ecology necessitates the precise stratification of fish species, and endeavours to agnize each type of fish based on certain characteristics. Conventional methods engage in manual characterisation which may not be optimal interms of time and impeccable results. While there have been various strategies to analyse and identify the diverse fish species, the flaws of traditional approaches are surpassed through different domains such as machine learning, deep learning and Internet of Things (IoT). The proposed research pivots on stratifying fish species using comparative machine learning algorithms to overcome the problems related to classifying similar species, while significantly ensuring accurate classification results in optimal time. This indagation utilizes the image viewer tool to unsheathe the characteristics of the procured fish type, while building the database to effectuate classification of the different species. The research incorporates the classification juxtapose through three machine learning algorithms such as the Cubic Support Vector Machines (CSVM), Cosine K-Nearest Neighbour (CosKNN) and the ensemble method using bagged trees and Ada boost optimizer to accurately classify fish species based on morphological features extracted from images. The proposed methodology involves the collection of a comprehensive dataset comprising images of different fish species, while sufficiently extracting relevant features, and further entail the application of machine learning algorithms for classification. This research thus contributes to the advancement of automated fish species identification, which can profoundly augment the proficiency and verity of fisheries management, along with enhancing the conservation of aquatic ecosystems. The simulations are carried out in MATLAB, and the results evince the ensemble method yields the zenith of classification accuracy as compared to the other two algorithms.

Received: August 08, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Keywords: *Fish classification, Ado Boost Optimizer, Ensemble Method, K-Nearest Neighbour, Support Vector Machine, MATLAB.*

I. Introduction

The classification of fish species is paramount in various fields including ecology, conservation, and fisheries management [1]. Accurate species identification serves as the cornerstone for understanding ecosystems, tracking population dynamics, and implementing effective conservation strategies [2], [3]. Traditional methods reliant on manual identification by experts are often time-consuming, subjective, and labour-intensive. Moreover, with the increasing volumes of data, there emerges a pressing need for automated solutions [4] capable of handling large datasets efficiently while maintaining high accuracy.

The advent of machine learning (ML) [11] has revolutionized the landscape of species classification, offering a promising avenue for automating and enhancing the precision of fish species identification [5]. By harnessing the computational power of ML algorithms, the analysis of vast amounts of data and extraction of intricate patterns [17] that may elude human observers are procured with optimal efficacy [6], [7]. This shift towards automation not only accelerates the pace of research but also ensures consistency and scalability in species identification efforts.

This paper aims to explore the application of ML algorithms for the classification of fish species, thereby addressing the growing demand for efficient and accurate identification methods. Through the progression of this research, elucidating the adeptness of ML techniques in categorizing fish species based on morphological features, genetic markers, or ecological attributes is implemented [8]. The progression of this study entails a comprehensive dataset comprising of images from diverse fish species across various habitats. This dataset thus serves as the foundation for training and evaluating the ML models. The assessment of performance on each algorithm through the accuracy metrics, while identifying the most effective ML approach for fish species classification forms to be the crux of this research.

The culmination of this research endeavours to hold significant implication on ecological monitoring, biodiversity conservation, and fisheries management [9], [18]. An accurate and efficient fish species classification system can streamline data collection efforts, facilitate species inventories, and render informed evidence-based decision-making processes. Furthermore, by advancing the state-of-the-art in ML-based species identification, this research contributes to the broader goal of harnessing technology to address pressing environmental challenges and safeguard aquatic ecosystems for future generations [10].

This paper is structured with section II elaborating the literature review in relevance to fish species classification. Section III explicates the proposed methodology entailed in this indagation, followed by the results and conclusion in section IV and V respectively.

II. Empirical Review

Bo Gong et al [7] delineated about the issue of feature extraction in fishes using the novel methodology for multi-water fish classification (Fish-TViT), that was derived from transfer learning and visual transformers. The novel methodology addresses the issues pertaining

Received: August 08, 2025

Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

overfitting of classifiers. The study also utilizes Gradient-weighted Category Activation Mapping (Grad-CAM) that extracts the features, and the boundaries where the optimization is required. Relative data manipulation methods to pre-process the image, and further utilize pre-trained model to crop the fish image to the required attributes is incorporated in this study. The prediction of the fish species type is then effectuated through the multi-layer perceptron. The results obtained for this study shows that the Fish-TViT renders a classification accuracy of 94.33% for low-resolution marine fish data, and 98.34% for high-resolution freshwater fish data, indicating the performance of this algorithm over traditional methods.

Siti Nurulain Mohd Rum et al [19] proposed an application to explicitly identify fish species, by using various image processing techniques. The study pivoted on the prototype system called "FishDeTec" that is used to identify freshwater fish species in Malaysia. The proposed methodology was constructed with VGG-16, that is a type of CNN that efficiently enables classification for large databases. The research entailed a combination of programming languages and databases such as Python to develop the model, Java for the application building, TensorFlow for DL purpose, Keras used as API for the DL, Google Collab that runs the application, Android Studio to integrate mobile working and management, and Firebase database that stores the information about fish species. The results evinced that transfer learning induced VGG-16 implementation yielded best results as compared to the network incorporated without the learning process.

Ceren Atik et al [20] delineated about Support Vector Machine (SVM) algorithms with the use of single SVM classifier, along with explicating the pros and cons of the algorithm. The proposed research explains about support vector machine chains (SVMC), that entails the combination of multiple SVMs, with decrementing attributes at each phase. The voting system called as tournament voting. This method works in groups with each throughput when surpassing will move to the next stage. The prediction results thus obtained is termed as the winner of the tournament. The output of SVMC thus procured rendered was juxtaposed with the basic SVM, with the former giving 88.11% and the latter estimated to 86.71%.

III. Proposed Methodology

The methodology for this indagation incorporates fish image data from Kaggle comprising of 7 types of species, along with their respective weights compiling together 4500 images. The training and testing of the algorithms are stratified into a 75% and 25% ratio. The diverse species of fishes stored in the fish image datastore is effectuated into a batch image processor, through which the characteristics of fish species are further extracted. The height and width of the fish species is extracted through ROI extractor [11]. Once the database is created from the extracted attributes, the classification learner in MATLAB is used for effectively comprehending the classifier accuracy. The overall architecture for this indagation is depicted in the figure below:

Received: August 08, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

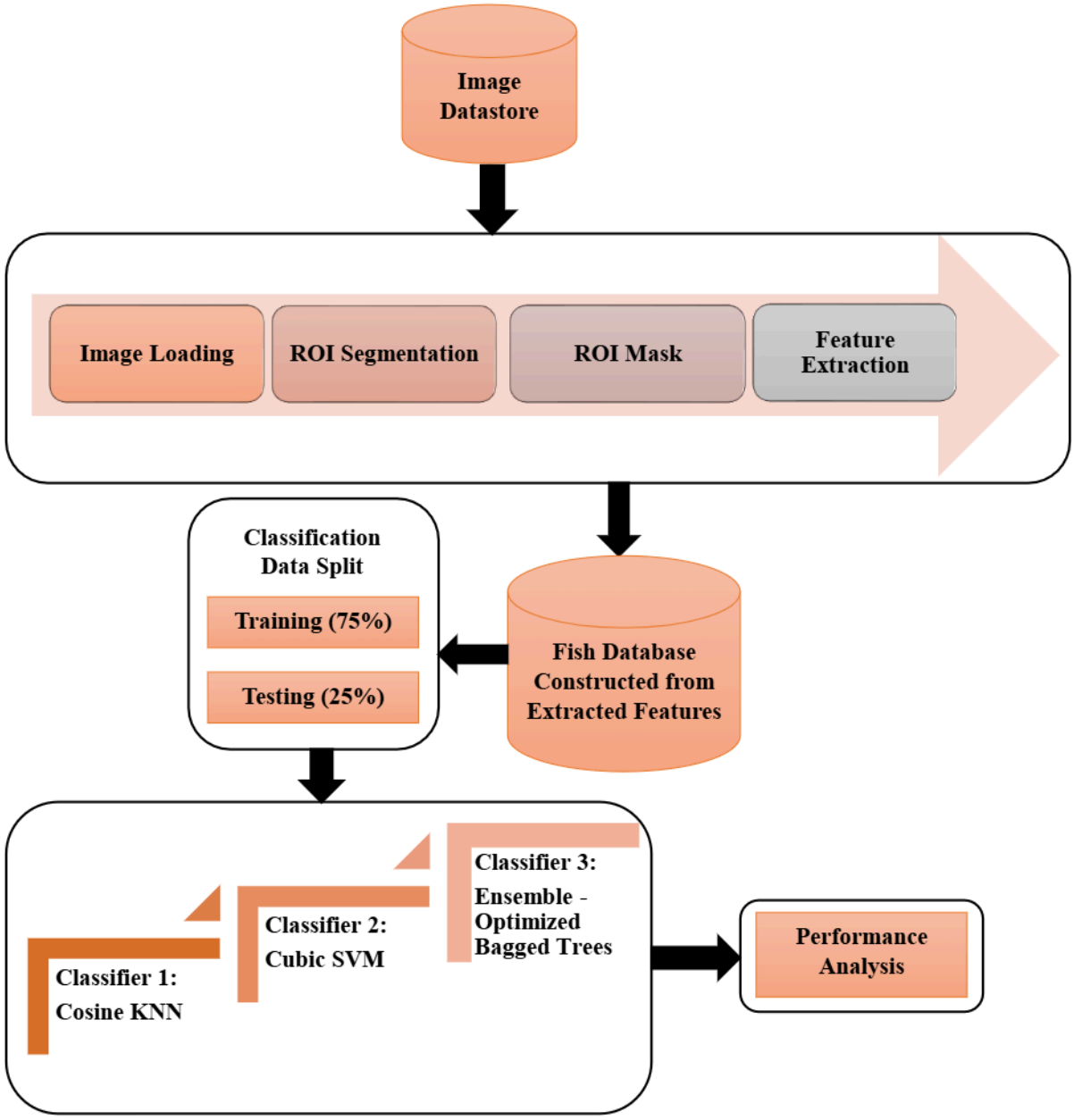

Fig 1. Proposed Architectural flow

The initial step in the proposed indagation entails the process of data collection. The repository of fish database is procured from Kaggle, with 7 diverse types such as the perch, pike, parkki, whitefish, bream, roach, and smelt. The attributes relevant to each type was unsheathed inorder to build a primary database after explicitly effectuating the fish images through a Region of Interest (ROI) extractor [14]. The attributes thus extracted are the weight, length and width of the fish through which the classification can be implemented. For the process of classification, three machine learning algorithms are chosen, and they are delineated as below:

Received: August 08, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

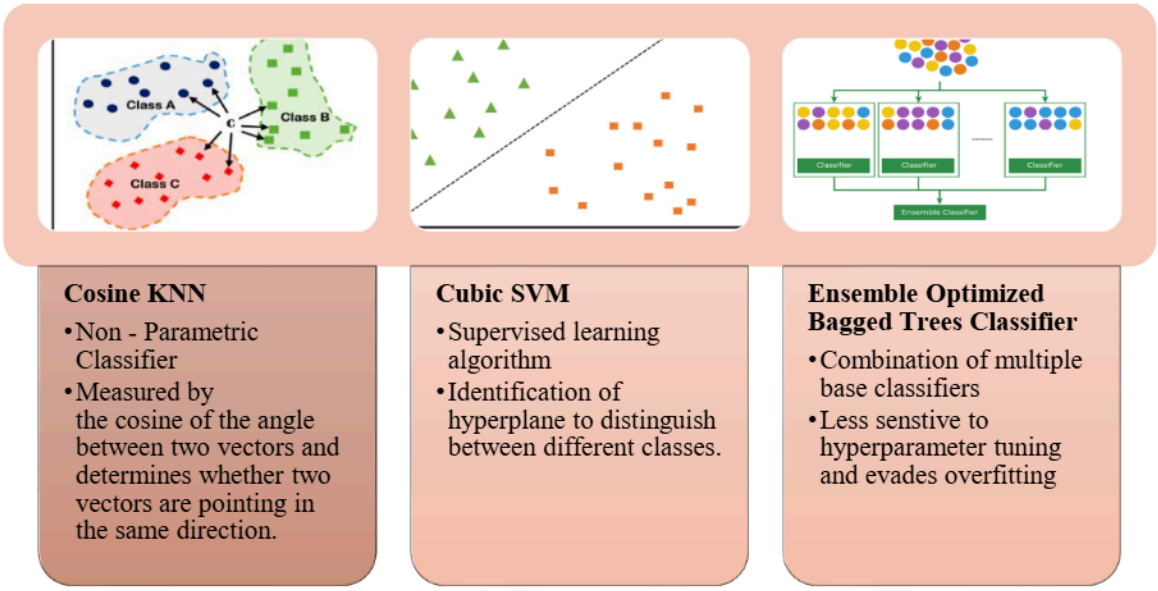

Fig 2. Classifiers used in Proposed Study

Cosine K - Nearest Neighbour Algorithm (KNN):

The cosine KNN algorithm [4] works on the principle of distance metrics to segregate those data that may lie in outlier, and entails the cosine distance metric to segregate the data. The distance metric thus used in this study is the cosine distance metric, and is shown in the equation given below:

$$Cos\ \theta = \frac{\vec{a}.\vec{b}}{\|\vec{a}\|.\|\vec{b}\|} \quad (1)$$

Where a and b are any given points, whose distance similarity is either 0 or 1, with the former referring to their similarity and the latter indicating the dissimilarity.

The algorithm works in three steps, with the first step initiating the calculation of the distances from the unlabelled points to every other point in the training set. The second step is to process the calculated distance similarity, while efficiently selecting the neighbours, and the final step is to effectuate classification based on the similar distances for the unlabelled points in the chosen dataset. The hyperparameters for the cosine KNN is elucidated with the number of neighbours fixed to ten neighbours, and the distance weights equalized. The results thus obtained from this classification is depicted in the subsequent section.

Cubic Support Vector Machine (CSVM):

The Cubic Support Vector Machine (CSVM) [20] is an algorithm that utilizes the hyperplane to segregate different classes. This algorithm relies on the structural risk minimization, and mitigates the generalization error. The objective function entailed in support vectors are normalized by a constraint and an error scale entailing the Lagrangian relaxation method [16], that largely minimizes the classification error for the algorithm. In this study, the cubic kernel function, with a box constraint level of 1, and one-to-one multiclass coding hyperparameters are implemented. The CSVM is represented mathematically through the below formula:

Received: August 08, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$S(m_i, m_j) = (1 + m_i^X m_j)^3 \quad (2)$$

Where m_i, m_j are the feature vectors, with their kernel function raised to the power of three to indicate the cubic capacity for the SVM. The results procured from the algorithmic implementation is elucidated in section IV.

Ensemble Optimized Bagged Tree Classifier:

The Ensemble optimized Bagged trees [12], [13], classifier combines multiple base classifiers to improve classifier accuracy. Bagged tree classifiers, based on the decision tree learner is combined with ada boost optimizer to further enhance the training and subset feature analysis. In the proposed study, the Ensemble model combines multiple classifier trees on a random subset of the training data. The randomization of the chosen values thus aid in overcoming the problems risen due to overfitting, while significantly augmenting the resilience of the algorithm. In this indagation, 30 independent tree classifiers are chosen individually, trained and then combined into an ensemble bagged tree approach [14], [15]. The total number of splits equalled 119, with the base learner adhering to the decision tree preset. The results of this algorithmic implementation yielded maximum accuracy, although the time to train and validate were comparatively higher than the previous two algorithmic models. The throughput of this effectuation is consecutively illustrated in the following section.

IV. Results

The results procured from image processing and classification learning are presented in this section to comprehend the efficiency of the learners, Fig 3 and 4 illustrate the ROI marking and masking. The subsequent figures explicate the classification accuracy and the results procured from the classification learner algorithms incorporated for this study.

Fig 3. ROI Selection

Fig 4. ROI Mask

The unsheathed attributes garnered from the image is then converted to data attributes comprising of the height, width and weight of the fish species. The target attribute of the fish species relies on the specified attributes to explicitly classify them into the right labelled set of class. The results thus obtained from the classification algorithms is depicted in the table below:

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 1. Performance Chart of Classification Algorithms

S.No	Classification Algorithms	Classification Accuracy	Training Time (Sec)	Testing Time (Obs / Sec)
1	Cosine KNN	84.2 %	2.59	20000
2	Cubic SVM	95.8 %	1.97	9700
3	Ensemble Bagged Tree	99.2 %	2.90	3000

The subsequent figures depict the Receiver Operating Curve (ROC) for each of the classifiers to comprehend the obtained accuracy, and the performance metric.

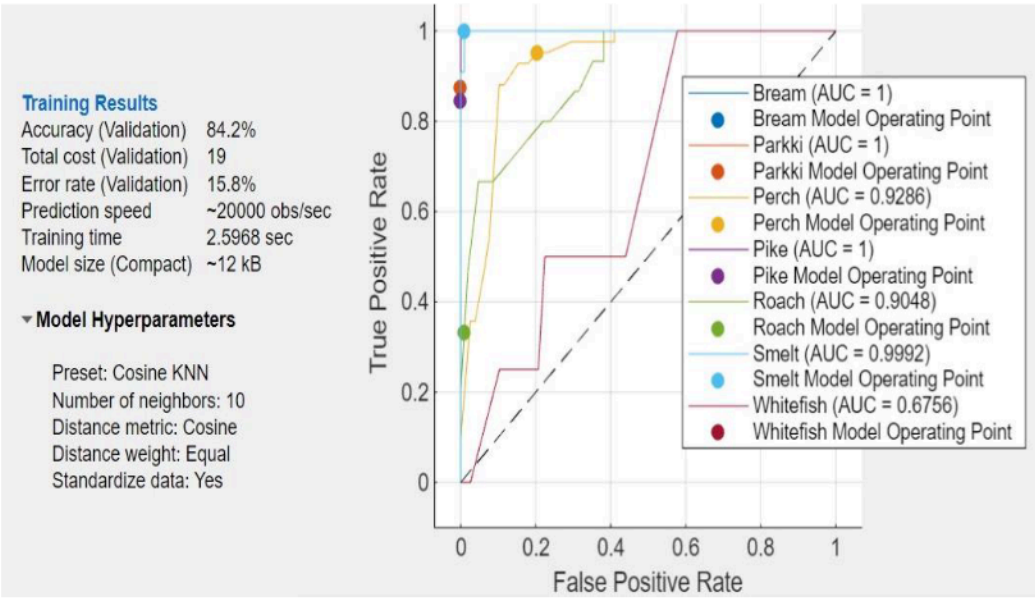

Fig 5. ROC with trainer classifier results for Cosine KNN Algorithm

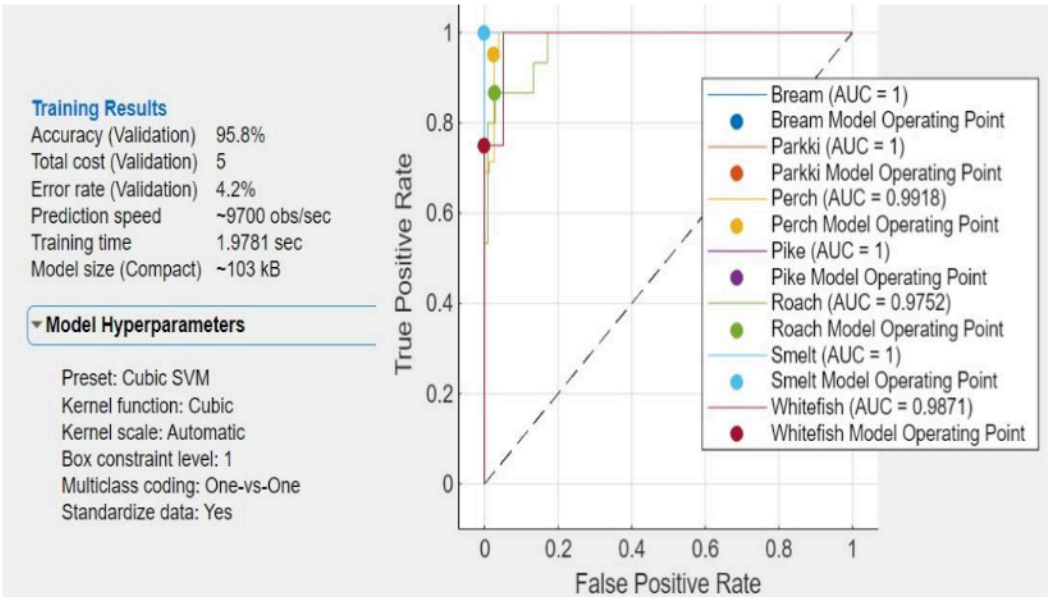

Fig 6. ROC with trainer classifier results for Cubic SVM Algorithm

Received: August 08, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

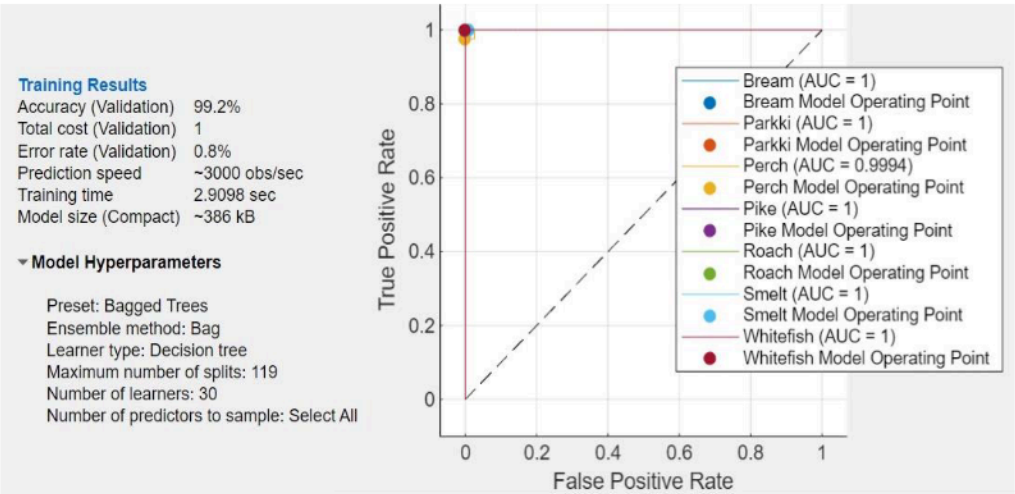

Fig 7. ROC with trainer classifier results for Ensemble Bagged Tree Classifier Algorithm

The results unambiguously evinces that the ensemble method supersedes the classifier accuracy of the other two algorithms, but requires more time for processing due to the bagging of multiple tree classifiers. The extraction of features plays a vital role in accelerating the accuracy in classifiers.

V. Conclusion

Fish species are crucial living organisms constituting the aqua-ecology, and their explicit identification and comprehensive scrutinization of various stratification techniques forms the crux of this research. The incorporation of machine learning algorithms such as CSVM, CosKNN and Ensemble Bagged trees with optimized Ada Boost are incorporated to depict the efficacy of the classification with respect to the unsheathed features. The ROI delineation to extract the height and width of the species plays an imperative aspect of this research in extracting the features from the fish repository. The results of the classifier models demonstrate that the simplistic nature of CosKNN makes it appropriate especially when the decision boundaries of the database are volatile, thereby establishing more scrutiny on the relationship of the local data points, yielding a throughput of 84.2%, while CSVM holds the adeptness to surpass problems beyond complex decision boundaries, along with effectively rendering an overall classifier accuracy of 95.8%. The Ensemble Bagged tree approach is significant interms of delving into the complex patterns, and holds higher efficacy of accuracy equalling 99.2%. This research thus pivots on facilitating prominence in biodiversity monitoring, while adequately aiding the fisheries and aquaculture department in ensuring higher levels of preservation of species. Future work on leveraging advanced hybrid models of classification, and automated real-time data acquisition techniques utilized with sensors could optimize this field of research consistently.

References

1. Souvik Roy, Sayak Mondal, Shreyashree Sarkar, Sumit Kumar Banerjee, Suman Bhattacharya, Mahamuda Sultana, “AI based Framework for Fish Species Identification

Received: August 08, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

and Classification", International Journal of Computer Sciences and Engineering, Vol.11, Special Issue.1, pp.81-88, ISSN: 2347-2693, 2023.

2. J. Dewan, A. Gele, O. Fulari, B. Kabade and A. Joshi, "Fish Detection and Classification," International Conference on Computing, Communication, Control and Automation(ICCUBEA), 2022
3. C. J. C. Z. Z. L. X. &. W. Z. Zhang, "A fish species identification and classification method based on deep learning and random forest," Aquaculture, vol. 531, p. 735962, 2021
4. R. P. F. A. I. S. a. A. J. Z. Abidin, "Betta Fish Image Identification using Feature Extraction GLCM and K-Nearest Neighbour Classification," in IEEE, Jakarta, 2022
5. Bhubneshwar Sharma, "Automatic Fish Detection and Species Classification using Optimal Archimedes Shooty Term Deep Network", HSOA Journal of Aquaculture & Fisheries, 7: 071, DOI: 10.24966/AAF-5523/100071, 2023
6. Jayashree Deka, Shakuntala Laskar, Bikramaditya Baklial, "Automated Freshwater Fish Species Classification using Deep CNN", Journal of Inst. Eng. India Ser. B, 104(3):603–621, doi.org/10.1007/s40031-023-00883-2, 2023
7. Bo Gong, Kanyuan Dai, Ji Shao, Ling Jing, Yingyi Chen, "Fish-TViT: A novel fish species classification method in multiwater areas based on transfer learning and vision transformer", doi.org/10.1016/j.heliyon.2023.e16761, 2023.
8. J.C. Ovalle, C. Vilas, L.T. Antelo, "On the use of deep learning for fish species recognition and quantification on board fishing vessels", doi.org/10.1016/j.marpol.2022.105015, elsevier.com/retrieve/pii/S0308597X22000628, 2022
9. Sudhakara M, Meena M. J, Madhavi K. R, Anjaiah P, & K L. P, "Fish Classification Using Deep Learning on Small Scale and Low-Quality Images", International Journal of Intelligent Systems and Applications in Engineering, 10(1s), 279, https://ijisae.org/index.php/IJISAE/article/view/2292, 2022
10. Zhao, Zhenxi, "Composited FishNet: Fish Detection and Species Recognition from Low-quality Underwater Videos." IEEE Transactions on Image Processing, 2021.
11. Bawa A., Samanta S, Himanshu S.K, Singh J, Kim J.J, Zhang T, Chang A, Jung J, DeLaune P, Bordovsky J, "A Support Vector Machine and Image Processing Based Approach for Counting Open Cotton Bolls and Estimating Lint Yield from UAV Imagery", Smart Agri. Tech. 3, 100140, 2023
12. Noor A, Ucąr M.K, Polat K, Assiri A, Nour R, Masciari E, "A Novel Approach to Ensemble Classifiers: FsBoost-Based Subspace Method", Math. Probl. Eng. 2020, 1–11.
13. Rojarath A, Songpan W, "Cost-sensitive Probability for Weighted Voting in an Ensemble Model for Multi-Class Classification Problems", Appl. Intell., 51, 4908–4932, 2021
14. Shahhosseini M, Hu, G.; Pham, H. Optimizing Ensemble Weights and Hyperparameters of Machine Learning Models for Regression Problems. Mach. Learn. Appl. 2022, 7, 100251.
15. Tuysuzoglu G, Birant D, "Enhanced bagging (eBagging): A novel approach for ensemble learning", Int. Arab J. Inf. Tech., 17, 515–528, 2020
16. M. Gaudioso, E. Gorgone, M. Labbe, A. M. Rodr´Iguez-Ch´Ia, "Lagrangian Relaxation for SVM Feature Selection", Computers & Operations Research, DOI: 10.1016/j.cor.2017.06.001, 2017

Received: August 08, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

17. Soad El Hiak, Xu He, "Automated Fish Measurement and Classification Using Convolutional Neural Networks (CNNs)", Computational Biology and Bioinformatics, 11(2), 33-48. https://doi.org/10.11648/j.cbb.20231102.12, 2023
18. H. Wang, Y. Shi, Y. Yue, and H. Zhao, "Study on Freshwater Fish Image Recognition Integrating SPP and DenseNet Network," IEEE International Conference on Mechatronics and Automation (ICMA), pp. 564-569, 2020
19. Siti Nurulain Mohd Rum, Fariz Az Zuhri Nawawi, "FishDeTec: A Fish Identification Application using Image Recognition Approach", International Journal of Advanced Computer Science and Applications, Vol. 12, No. 3, 2021
20. Ceren Atik, Recep Alp Kut, Reyat Yilmaz and Derya Birant, "Support Vector Machine Chains with a Novel Tournament Voting", Electronics 12(11), 2485; https://doi.org/10.3390/electronics12112485, 2023

Received: August 08, 2025