Publications

Available Online

J. Sci. Res. **17** (2), 619-635 (2025)

JOURNAL OF SCIENTIFIC RESEARCH

www.banglajol.info/index.php/JSR

Prediction of Regression Models for Haematological Parameters of Common Carp *Cyprinus carpio* Fed with Probiotic Feed using Machine Learning Algorithm

M. S. Rama[1], R. Sudharsanan[2], P. V. Gopirajan[3], K. Premalatha[4], K. Sivakumar[5*]

[1]Department of Biotechnology, St. Josephs' College of Engineering, Chennai, India

[2]Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, India

[3]Department of Computational Intelligence, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

[4]Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India

[5]Department of Biotechnology, Karpaga Vinayaga College of Engineering and Technology, Chengalpattu India

Received 22 December 2024, accepted in final revised form 27 February 2025

Abstract

The aim of the study is prediction of regression model to forecast hematological profiles of *Cyprinus carpio* (common carp) using machine learning algorithm. The fish fed with control and probiotic *Lactobacillus macrolides* (LM) enriched feed (2 % LM, 4 % LM, 6 % LM, and 8 % LM pelletized feed) duration of 60 days. The blood samples were drawn and subjected to hematological profiles assessed post-experiment. The regression models like gradient booster regressor, random forest regressor, linear regression, and decision tree regressor were employed to determine the most accurate predictive model, followed by validated through voting regressor methods. Significant variations in hematological profiles were observed among the different feeding regimes. The gradient booster regressor emerged as the most effective model, achieving a coefficient of determination (R^2) of 1.00, while the decision tree regressor exhibited R^2 values ranging from 0.99 to 1.00 across different hematological parameters except MCV and MCH. Notably, the voting regressor method confirmed the superiority of the gradient booster regressor, indicating a robust predictive capacity. These findings underscore the potential of machine learning techniques to enhance nutritional strategies in aquaculture by predicting fish health outcomes, thus contributing to more sustainable and effective aquaculture practices. For accuracy and predictive power, Gradient Boosting may be the best choice, while Random Forest offers a similarly strong alternative.

Keywords: Machine Learning; *Cyprinus carpio;* Haematological Parameters; Blood serum; Regression model.

doi: https://dx.doi.org/10.3329/jsr.v17i2.78599 J. Sci. Res. **17** (2), 619-635 (2025)

---

[*] *Corresponding author*: ksivakumar76@gmail.com

1. Introduction

The global demand for fish and seafood is met largely by the aquaculture industry. As aquaculture operations have grown, the need for efficient, data-driven management approaches has become apparent. Machine learning (ML) techniques offer a promising solution for enhancing aquaculture productivity and sustainability. These algorithms have demonstrated effectiveness in various aspects of intelligent aquaculture, including fish biomass estimation, species identification and classification, behavior analysis, and water quality prediction. Probiotic feed usage in aquaculture has grown due to its potential benefits for fish health, growth, and disease resistance [1-3]. Understanding how probiotics affect fish physiology, particularly haematological and serum biochemical profiles, is crucial for optimizing their use [4]. Current research explores the potential of machine learning to predict these profiles in *Cyprinus carpio* (common carp) fed probiotic-supplemented diets [3]. Hematological and serum biochemical markers serve as important indicators of fish health, reflecting immune function, nutritional status, and overall physiological condition [5]. Evaluating these parameters provides insights into the effects of dietary interventions, including probiotic supplements [6].

ML algorithms excel at analyzing complex data and identifying patterns, making them highly suitable for predicting biological responses [7]. By leveraging datasets of *C. carpio* haematological and serum biochemical characteristics following various probiotic dosages, ML models can uncover relationships between probiotic intake and physiological responses [8]. Several ML techniques, such as Random Forests, have shown success in classifying haematological data of *C. carpio* given different probiotic regimens, achieving high accuracy. This suggests Random Forests' ability to predict hematological profiles based on probiotic consumption [8]. Support Vector Machines (SVM) are proficient in classification and regression tasks [9]. While no specific studies on *C. carpio* and probiotics were identified, SVMs have been successfully applied in aquaculture to predict fish growth performance based on environmental factors, indicating their potential for predicting haematological and serum biochemical profiles [10]. Neural networks are adept at recognizing complex non-linear data relationships. Their application in predicting hematological and serum biochemical profiles of probiotic-fed *C. carpio* requires further investigation [11]. Developing accurate ML models for predicting these profiles in probiotic-fed *C. carpio* offers numerous benefits, including optimized probiotic dosing, the ability to forecast physiological responses to probiotics for determining ideal dosages to improve fish health and growth [4], early disease detection [12], and personalized feeding strategies [13]. The use of machine learning to predict hematological profiles of probiotic-fed *C. carpio* represents a promising avenue for improving aquaculture practices.

2. Materials and Methods

2.1. *Experimental design*

The experiments were conducted in accordance with the guidelines and regulations established by the Committee for Control and Supervision of Experiments on Animals (CCSEA), Department of Animal Husbandry and Dairying, Ministry of Fisheries, Animal Husbandry and Dairying, Government of India and the experimental protocol was approved by Institutional Animal Ethical Committee (Karpaga Vinayaga Institute of Medical Sciences and Research Institute, Tamil Nadu, India) (No: 181GO/ERE/S/15/CPCSEA dated 04.12.2018). *C. carpio* (12.19 ± 1.12 cm and 25.93 ± 1.48 g) was obtained from a fish farm in Kolathur, Chennai, Tamil Nadu. They were promptly transferred to the laboratory, where they were adequately aerated and acclimatized to the laboratory conditions for 48 hours with standard pelletized diet. Following acclimation, fish were categorized into five groups: Trial I (control feed), Trial II (2 % *Lactobacillus macrolides* (LM) probiotic feed), Trial III (4 % LM probiotic pelletized feed), Trial IV (6 % LM probiotic feed), and Trial V (8 % LM probiotic pelletized feed) for a duration of 60 days. The experiments were conducted in triplicate and each tank inoculated 10 fishes. Daily, the fish excrement and three-quarters of the aquarium's water were expelled and replaced with fresh water. Immunological parameters of *C. carpio* subjected to various feeding regimes were examined. They were sustained under standard culture conditions and provided with ad libitum access to the appropriate diet.

2.2. *Collection of blood*

At the end of the trial period, fish from each treatment group were anesthetized with benzocaine (50 µL/L). Blood was extracted from the caudal vein with a 1 ml syringe and a 25-gauge needle that had been pre-rinsed with a 2.7 % ethylenediaminetetraacetic acid (EDTA) solution. Subsequently, the blood was promptly transferred to a test tube containing EDTA powder, utilized as an anticoagulant, and meticulously stirred to avert blood lysis. Blood was utilized for subsequent analysis.

Hematological assays were conducted using the methodology [14]. Red blood cell (RBC, $\times 10^6$ cells/µL) and white blood cell (WBC, $\times 10^3$ cells/µL) counts were ascertained by adhering to a standardized methodology utilizing a Neubauer hemocytometer. The hematocrit (HCT, %) was assessed using the micro-hematocrit technique. Hemoglobin (Hb, g/dL) was measured using a UV-1800 spectrophotometer (Shimadzu, Japan) at a wavelength of 450 nm. The microhematocrit method was employed for HCT determination in fish blood analysis [15]. The derived haematometric indices were Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), and Mean Corpuscular Hemoglobin Concentration (MCHC) [16].

2.3. *Prediction of regression model using ML algorithm*

The dataset comprising RBC, WBC, Hb, PCV, HCT, MCV, MCH, and MCHC was preprocessed to eliminate noise, redundancy, incompleteness, and missing values, after which the refined data, 80 % and 20 % data was considered as training and test data, respectively. The data was analyzed using a Decision tree regressor Gradient boosting regressor, Linear regressor, and Random Forest Regressor (Fig. 1).

The Decision tree regressor is the most effective machine learning model for predicting target variables by segmenting them [17-19], and it assesses the adequacy of the data while pruning any overfitting. The gradient boosting regressor enhances prediction accuracy and efficiency in the proposed work through iterative verification and validation of the achieved and expected outcomes [20,21]. The linear regressor identifies the correlation between the dependent and independent variables and formulates a linear equation based on the trained data [22]. The Random Forest Regressor is an ensemble of decision trees used to predict continuous outcomes. The linear characteristics derived from the linear regressor are subsequently utilized as input for the Random Forest Regressor model to discern the nonlinear relationships among the target variables. This amalgamation of linear and nonlinear components yields exceptional predictive accuracy.

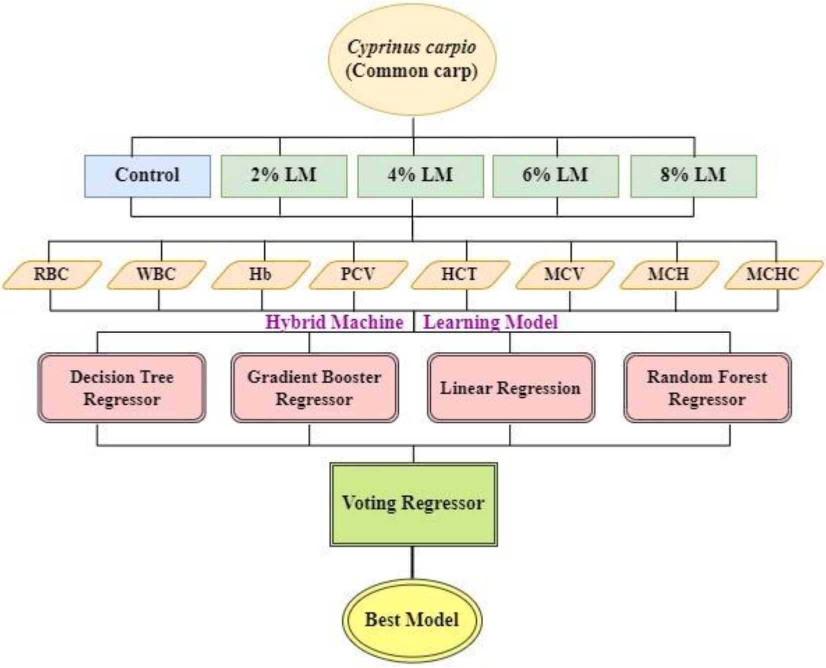

Fig. 1. Data analysis flow chart for prediction of best regression model.

To study the impact of probiotic alterations of L. macrolides on RBC count, WBC count, Hb, HCT, packed cell volume (PCV), MCV, MCH, and MCHC, a hybrid ML model was applied. It included different regression algorithms: Decision Tree Regressor, Gradient Booster Regressor, Linear Regression, Random Forest Regressor, and others. The data obtained from these models is processed by means of a Voting Regressor that chooses the Best Model for analyzing the effects of lipid meal concentration on the physiological parameters of *C. carpio*.

2.4. *Voting regressor*

To enhance prediction accuracy, we utilized a voting regressor that amalgamates various machine learning models by the averaging of their predictions. This ensemble methodology relies on the premise that aggregating the predictions of multiple models generally yields superior generalization performance compared to any individual model's capabilities [10,23].

The voting regressor integrates basic models that separately excel in various facets of the data, thus mitigating the shortcomings of each model. This study used a Random Forest Regressor, Gradient boosting regressor, Linear regressor, and Decision tree regressor as the foundational models for the voting regressor. These algorithms were selected for their diversity in learning paradigms bagging, boosting, and kernel-based methods—thereby guaranteeing different decision bounds. Each model was trained autonomously on the training set, and their respective forecasts were averaged to generate the final prediction. The voting regressor employed a soft voting strategy, where each model's prediction contribution was proportional to its confidence level. We assessed the ensemble model's performance using standard regression metrics, including mean absolute error (MAE), and R-squared (R^2) on the test set. Furthermore, to mitigate overfitting, 5-fold cross-validation was utilized throughout the training phase. The aggregated forecasts consistently produced reduced error metrics relative to the singular models, illustrating the efficacy of the ensemble approach. All models were executed with the scikit-learn toolkit in Python [24], and hyperparameters for each regressor were optimized through grid search.

2.5. *Statistical analysis*

The experimental data were reported as mean ± SD. All the experimental data such as haematological profile data were determined using one-way ANOVA ($P<0.05$) followed by Duncan Multiple Range Test (DMRT) performed. The data was analyzed using SPSS 22.0 ver.

3. Results and Discussion

3.1. *Hematology profile*

The hematological indices of *Cyprinus carpio,* namely RBC, WBC, Hb, PCV, HCT, MCV, MCH, and MCHC, exhibited substantial fluctuations in response to varying feeding regimes, especially with elevated concentrations of *Lysinibacillus macroides* (LM). The observed rise in RBC counts from 1.01×10^6 cells/mm in the control group to 2.94×10^6 cells/mm in the 8 % LM group (Table 1) indicates a stimulatory effect of LM on erythropoiesis, aligning with findings in other fish species where dietary supplements have been demonstrated to enhance red blood cell production and overall hematological health [8,25]. The elevation in WBC counts suggests a possible augmentation of the immunological response, as heightened leukocyte levels are frequently correlated with enhanced health and disease resistance in fish [26,27].

The increase in hemoglobin levels from 9.56 (g/dL) in the control group to 14.21 (g/dL) in the 8 % LM group (Table 1) indicates enhanced oxygen-carrying ability, essential for the metabolic requirements of the fish, especially under fluctuating environmental conditions [28]. This increase in hemoglobin concentration corresponds with research indicating that dietary factors affect hemoglobin levels, with certain nutrients or supplements resulting in notable enhancements in oxygen transport efficiency in aquatic organisms [29,30]. Moreover, elevations in PCV and HCT values indicate an augmented ratio of red blood cells, substantiating the assertion that LM supplementation beneficially influences the comprehensive blood profile of *C. carpio*.

Fluctuations in MCV, MCH, and MCHC values signify changes in red blood cell shape and hemoglobin concentration per cell, attributable to the nutritional makeup of the diets [31,32]. Although MCH exhibited no significant difference between the 6 % and 8 % LM groups, the prevailing trend indicated that elevated LM concentrations may result in more marked alterations in red blood cell characteristics, a phenomenon frequently observed in fish subjected to diverse dietary regimes [33]. The statistical significance of these characteristics, demonstrated by the one-way ANOVA and subsequent DMRT test, showed a significant difference ($P<0.05$) between the different feeding regimes (Table 1). The reliability of these findings and their implications for aquaculture operations have focused on enhancing fish health through nutritional treatments [34].

Table 1. Hematological indices of *C. carpio* fed in the different feeding regimes.

	RBC (10^6 cells/mm^3)	WBC (10^3 cells/mm^3)	Hb (g/dL)	PCV (%)	HCT (%)	MCV	MCH	MCHC
Trail I	$1.01±0.07^a$	$11.68±0.08^a$	$9.56±0.07^a$	$28.42±0.07^a$	$32.36±0.07^a$	$283.60±18.65^a$	$95.32±5.79^a$	$20.17±0.10^a$
Trail II	$1.72±0.09^b$	$12.03±0.70^b$	$11.28±0.08^b$	$35.15±0.08^b$	$38.58±0.13^b$	$205.41±10.05^b$	$65.89±2.95^b$	$19.25±0.09^b$
Trail III	$1.99±0.10^c$	$12.90±0.10^c$	$12.01±0.10^c$	$33.16±1.10^c$	$37.27±1.84^c$	$166.93±3.17^c$	$60.52±2.57^c$	$21.74±0.55^c$
Trail IV	$2.27±0.35^d$	$13.40±0.09^d$	$13.02±0.09^d$	$34.11±0.26^d$	$38.85±0.09^d$	$153.41±21.24^d$	$58.52±7.99^{cd}$	$22.89±0.12^b$
Trail V	$2.94±0.08^e$	$14.26±0.08^e$	$14.21±0.31^e$	$35.36±0.08^b$	$39.93±0.05^b$	$120.53±2.91^e$	$48.43±1.10^d$	$24.11±0.50^d$

Values are represented as Mean ± SD; Anova followed by Duncan's Multiple Range Test performed; Different alphabets in the same columns indicates a significant difference ($P < 0.05$)

3.2. *Regression models*

The regression models depicted several hematological parameters (RBC, WBC, Hb, PCV, HCT, MCV, MCH, and MCHC) in relation to the target variable. Every graphic features a fitted regression line accompanied by statistical metrics that assess the efficacy of the model. This value indicates the degree to which the regression line corresponds to data. It spans from -1 to 0 and from 0 to 1. This statistic quantifies the average magnitude of the mean absolute errors (MAE) in a series of predictions, disregarding their directionality. The reduced MAE values signify superior model performance.

3.2.1. *Linear Regression*

Examination of blood parameters in *C. carpio* demonstrated substantial relationships with the goal variable, notably emphasizing the importance of Hb, PCV, and HCT. The Hb parameter had a robust correlation with the target variable, displaying an R^2 value of 0.97 and a mean absolute error (MAE) of 0.38 (Fig. 2). This signifies a strong predictive ability, implying that Hb levels are a vital indicator of the physiological conditions in *C. carpio*. The elevated R^2 value indicates that fluctuations in Hb can account for a significant percentage of the variability in the target variable, corroborating prior research that has associated Hb levels with environmental stress and overall health in fish populations [33]. Conversely, both PCV and HCT exhibited moderate associations, with R^2 values of 0.50 and MAE values of 1.62. These measures are crucial for evaluating the oxygen-carrying capacity and total blood volume in fish, which can be affected by multiple factors such as water quality and salinity [35]. The moderate connection indicates that, although PCV and HCT are significant, they may be more vulnerable to variations caused by environmental stresses or clinical states than Hb. This corresponds with the concept that blood parameters act as indices of health state, mirroring the physiological reactions of *C. carpio's* response to external stresses, including variations in salinity and exposure to pollutants [35,36]. MAE values offer an additional understanding of the forecasting precision of these factors. Reduced MAE values, especially for Hb, signify a more dependable model for clinical evaluation and decision-making in aquaculture and fisheries management. Accurate prediction of health outcomes from blood factors can improve care methods for *C. carpio*, particularly in settings where they encounter pollution or other stressors [37,38]. These findings emphasize the necessity of monitoring these blood parameters as an integral component of a complete health assessment plan for *C. carpio*, enhancing management methods in aquaculture, and conservation initiatives.

3.2.2. *Gradient Booster Regressor*

The utilization of Gradient Boosting regression analysis for forecasting the blood parameters of *C. carpio* exhibited significant predictive efficacy, as indicated by the attained R^2 value of 1.00 (Fig. 3). This signifies an optimal alignment of the model with the data, implying that the model can account for all variability of the target variable based on the chosen blood measurements. Elevated R^2 values signify a strong model, consistent with prior research demonstrating that gradient boosting surpasses conventional regression methods across multiple fields, including healthcare [39,40].

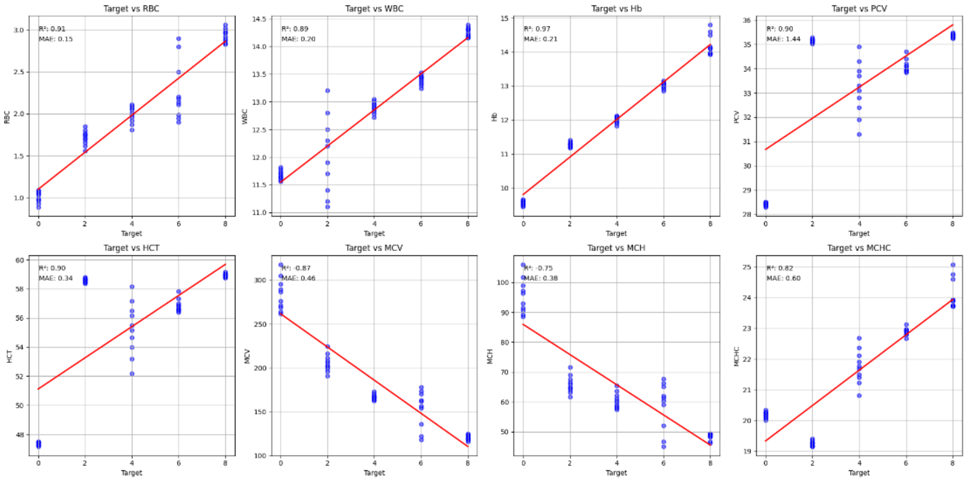

Fig. 2. Hematological profile prediction and actual value of *C. carpio* fed with different feeding regimes using Linear Regression. This figure shows scatter plots of the relationship between the probiotic feed concentrations and the different hematological parameters of *C. carpio* that were measured. Each graph shows the relationship between the probiotic feed and the corresponding parameter of RBC, WBC, Hb, HCT, PCV, MCV, MCH, and MCHC (the value of R^2 describes how well the data fit a model of this form; larger values mean better fit. Mean Absolute Error (MAE) measures the average magnitude of the errors in a set of predictions, not the accuracy of the model.).

The minimal MAE across all parameters further substantiated the trustworthiness of the model. The MAE is an essential metric in regression analysis because it measures the average magnitude of errors in a collection of predictions. A persistently low MAE signifies that the model's predictions closely align with the actual values, which is crucial for practical applications where accuracy is vital (Nepal and Ghimire 2023). In fish physiology, prediction accuracy is vital for evaluating the health and stress responses of fish populations because blood parameters serve as essential markers of physiological circumstances [41].

The graphical depiction of the model's predictions facilitates intuitive comprehension of the correlations between the blood parameters and the target variable. This is especially advantageous in scientific studies because intricate datasets can frequently conceal fundamental trends. The capacity to rapidly evaluate these associations improves the model's interpretability, facilitating academics and practitioners to extract actionable insights from the data [42]. The iterative process of gradient boosting, which amalgamates several weak learners to create a robust predictive model, enhances its efficacy in identifying intricate patterns within data [43].

The Gradient boosting regression analysis utilized on the blood parameters of *C. carpio* not only exhibited remarkable prediction efficacy, but also underscored the promise of machine learning methodologies in biological research. These results highlight the necessity of employing sophisticated statistical techniques to improve our comprehension of physiological metrics in aquatic organisms, facilitating future investigations into the ramifications of these forecasts in ecological and environmental frameworks [44].

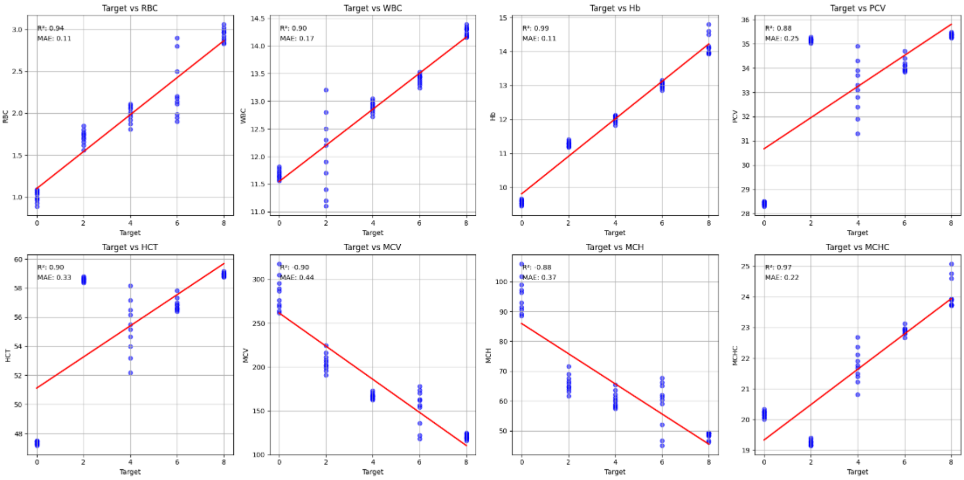

Fig. 3. Hematological profile prediction and actual value of *C. carpio* fed with different feeding regimes using Gradient Booster Regressor. The figure shows a series of scatter plots testing the correlation of the target variable with a number of physiological parameters of C. carpio. Each subplot corresponds to one particular parameter: RBC, WBC, Hb, HCT, PCV, MCV, MCH, and MCHC (R^2 Values represent the strength of linear association, which in between 0.88 to 0.99 signifies a generally strong association clearly visible for Hb (R^2 = 0.99). Mean Absolute Error (MAE) provides insight into accuracy in prediction, and the values range from 0.11 to 0.44, where lower MAE indicates superior model performance.

3.2.3. *Decision tree regressor*

Decision tree regression is a robust nonparametric supervised learning method that effectively predicts continuous values by segmenting the data into subsets according to the input feature values. This system is notably effective owing to its intuitive framework, wherein each node symbolizes a feature, branches indicate decision rules, and leaf nodes reflect the outcomes. Examination of many hematological markers of *C. carpio*, including RBC, WBC, Hb, HCT, MCV, MCH, and MCHC, via decision tree regression revealed predictive efficacy, as indicated by elevated R^2 values and minimal MAE metrics.

The findings demonstrate that the model attained a R^2 value of 0.99 for RBC and a R^2 of 1.00 for WBC, Hb, HCT, MCV, MCH, and MCHC, with associated MAE values of 0.08 and 0.00, respectively (Fig. 4). The metrics indicate that the model's predictions closely align with the actual observed values, underscoring the effectiveness of the decision tree regression in this situation. This result corresponds with prior research highlighting the efficacy of decision-tree techniques in regression tasks, demonstrating their capacity to manage intricate datasets and achieve high accuracy [45]. The elevated prediction accuracy shown in this investigation can be ascribed to the intrinsic properties of Decision Tree algorithms, which facilitate the examination of feature interactions and recognition of key predictors. The hierarchical organization of decision trees enhances the visualization of decision-making processes, rendering them interpretable and beneficial for subsequent study [46,47]. This interpretability is essential in biological situations, as understanding the correlations among various blood indicators can yield insights into the health and physiology of *C. carpio*.

The investigation indicated potential methods for improving model efficacy, including feature engineering and hyperparameter adjustment, along with robust predictive performance. These methodologies can enhance the decision tree's architecture and augment its generalization skills, as evidenced by the literature endorsing optimization tactics in machine learning models [48,49]. Investigating these strategies may enhance the accuracy and reliability of predictions, thus reinforcing decision tree regression as an essential instrument in biological data processing. The implementation of decision tree regression to forecast blood parameters in *C. carpio* has produced remarkable outcomes, as evidenced by the elevated R^2 values and low MAE. The results highlight the method's robust prediction powers and its promise for directing future investigations into feature interactions and model optimization techniques.

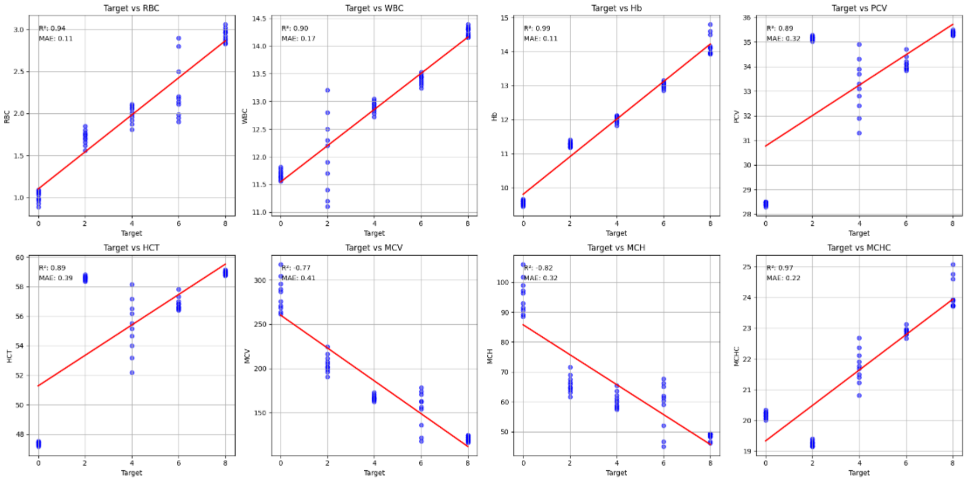

Fig. 4. Hematological profile prediction and actual value of *C. carpio* fed with different feeding regimes using Decision Tree Regressor. The figure depicts the association between a probiotic concentrations and some hematological parameters: RBC, WBC, Hb, HCT, PCV, MCV, MCH, and MCHC. The graphs indicating the strength and direction of the association between the probiotic concentrations and each hematological parameter, as shown by the R^2 and MAE values. The high R^2 values of RBC, Hb, and MCHC indicate strong positive correlations, whereas the negative R^2 of MCV and MCH indicates inverse correlation with the target variable.

3.2.4. *Random forest regressor*

The Random Forest Regression analysis results demonstrated a predictive correlation between the different blood parameters of *C. carpio* and the target variable, as indicated by the elevated R^2 values. The associations identified between hemoglobin (Hb) and mean corpuscular hemoglobin concentration (MCHC) indicate that these parameters act as reliable predictors in this setting. The R^2 values of 1.00 for Hb and 0.97 for MCHC, coupled with mean absolute errors (MAE) of 0.00 and 0.03 respectively, highlight the dependability of these measures in forecasting the target variable (Fig. 5).

The importance of Hb as a predictor corresponds with findings from other studies that highlight the impact of hematological indicators in evaluating fish health and metabolic conditions. The hematological parameters are intricately connected to the metabolic levels related to fish size, with larger fish exhibiting elevated RBC and Hb values, potentially indicative of their heightened metabolic requirements [50]. The correlation between hematological parameters and metabolic activity, suggesting that larger fish typically exhibit elevated metabolic rates, thus elucidating the significant predictive capacity of hemoglobin in the present analysis [51]. Moreover, the seasonal fluctuations in blood parameters indicate that physiological alterations associated with reproduction and environmental conditions can markedly affect hematological indices, thereby reinforcing the significance of Hb and MCHC as indicators of fish health [52].

Furthermore, the results concerning MCHC align with the extensive literature on fish physiology, suggesting that changes in blood parameters may function as biomarkers of environmental stress and general health status. Erythrocyte indices for detecting anemia and evaluating the health of fish populations under various environmental conditions. This underscores that MCHC, in conjunction with Hb, is a vital metric for assessing the physiological condition of *C. carpio* [53].

The utilization of random forest regression in this investigation is particularly significant, as it has demonstrated superior accuracy and dependability compared with other modeling techniques. The capacity of this method to manage intricate variable interactions without significant preprocessing renders it an optimal selection for biological data analysis [54].

This is especially pertinent to fish blood parameters, as several factors can affect the results. The high accuracy of the RF model in predicting the target variable from blood parameters suggests its potential application in aquaculture and fisheries management, where monitoring fish health is essential for sustainable practices.

The robust predictive correlations exhibited by the random forest regression analysis for Hb and MCHC in *C. carpio* emphasize the significance of these hematological indicators in evaluating fish health. The results are validated by the existing literature that highlights the metabolic and physiological importance of blood parameters in fish, indicating that they can function as dependable indicators for assessing the health and welfare of aquatic animals.

3.3. ***Voting regressor***

Examination of hematological parameters in fish is essential for understanding their health and the influence of environmental factors. The outcome of the voting regressor analysis emphasizes the heterogeneity in the predictions generated by various regression models for these parameters, especially when the quantity of training data increases. This variability is substantial as it indicates the impact of multiple factors on fish health, which may be evaluated via their blood profiles. Environmental factors, including water quality, temperature, and pollution levels, have been demonstrated to influence hematological parameters in fish species, rendering these parameters valuable indicators of their physiological condition and overall health [55-58].

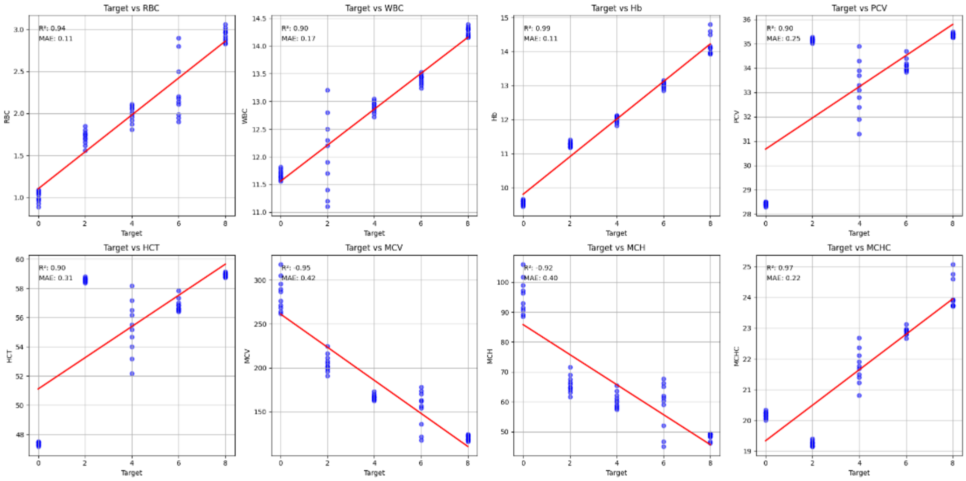

Fig. 5. Hematological profile prediction and actual value of *C. carpio* fed with different feeding regimes using Random Forest Regressor. This set of scatter plots shows the correlations between a target variable and seven hematological parameters: RBC, WBC, Hb, HCT, PCV, MCV, MCH, and MCHC. A fitted linear regression line is added to each plot to show the correlation between the probiotic concentration and the corresponding parameters. The plots illustrate different levels of correlation between these parameters, with RBC, Hb, and MCHC having particularly high positive correlations, and MCV and MCH having strong negative correlations with the target variable.

The voting regressor research indicated that models, such as the gradient boosting regressor and random forest regressor, yielded superior predicted values relative to the decision tree regressor (Fig. 6). This gap may be ascribed to the sophistication and flexibility of the more advanced models, which can capture the subtleties of the data more effectively. The significance of model selection in forecasting hematological parameters is highlighted by research indicating that various environmental stressors, including heavy metals and temperature variations, can result in substantial alterations in blood parameters such as hemoglobin concentration and red blood cell counts [45,59-61].

The anticipated values, spanning roughly 0.925 to 1.1, indicate variability in the models' outputs, consistent with research demonstrating that hematological parameters can vary due to multiple factors, such as stress, age, and environmental conditions [43,55,56,62]. Research indicates that exposure to pollutants such as lead and cadmium can result in anemia in fish, as evidenced by reduced red blood cell counts and hemoglobin levels [62,63]. Seasonal oscillations in temperature and water quality have been associated with alterations in hematological parameters, highlighting the necessity for comprehensive predictive models that can accommodate these variations [64,65]. The Gradient Boosting Regressor's performance indicates that it may be more proficient in managing the intricacies of fish hematological data.

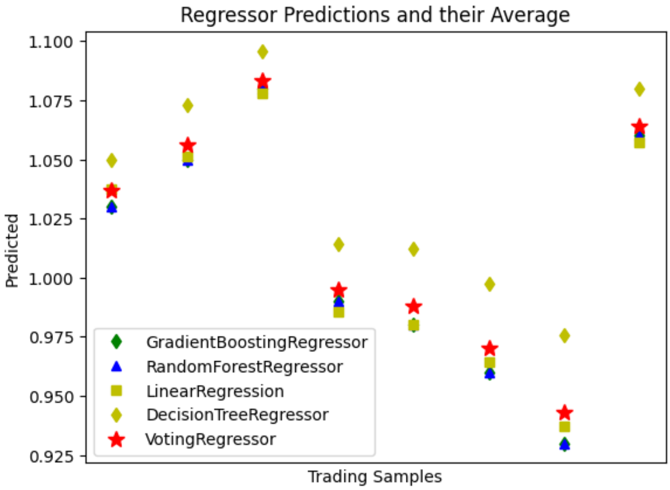

Fig. 6. Comparision of different machine learning prediction. The plot displays the predicted values from five different regression models applied to samples, alongside their average predictions. Each symbol represents a distinct regression model, facilitating comparisons between their predictions across various samples. **Green Diamonds:** Gradient Boosting Regressor; **Blue Triangles:** Random Forest Regressor; **Yellow Squares:** Linear Regression; **Brown Diamonds:** Decision Tree Regressor; **Red Stars:** Voting Regressor (average prediction). The red stars indicate the Voting Regressor, which consolidates predictions from all other models to produce a final averaged output. The plot illustrates that while most models track similarly, the Voting Regressor tends to stabilize predictions across the trading samples, suggesting its effectiveness in capturing an averaged performance. This visualization aids in assessing model performance and variability, guiding decision-making in selecting appropriate predictive models

This corresponds with studies demonstrating that sophisticated machine-learning methodologies can proficiently predict biological data, encapsulating interactions among numerous variables that more simplistic models may neglect [58,66]. The capacity of these models to adjust to expanding training samples underscores the significance of possessing a thorough dataset that includes a diverse array of environmental conditions and fish species, as this variety can improve the predictive precision [55,56,66]. Furthermore, the results of the vote regressor analysis align with the extensive literature on fish health evaluations. Hematological parameters are essential biomarkers for assessing the physiological impact of environmental stress on fish populations. Research indicates that alterations in blood parameters can reflect the health status of fish in contaminated habitats, rendering them crucial for environmental monitoring and conservation initiatives [56,57,67]. The use of machine learning models in this domain may enhance the precision and promptness of fish health evaluations, thus improving management methods in aquaculture and natural ecosystems [21,66]. The voting regressor analysis offers significant insights into the prediction efficacy of several regression models regarding hematological characteristics of fish. The disparities in model efficacy underscore the intricacy of biological data and the impact of environmental variables on fish health. Advancements in aquatic biology necessitate the incorporation of sophisticated statistical approaches and machine learning techniques to improve our understanding of fish physiology and the effects of environmental changes on aquatic ecosystems.

These data indicate that elevated levels of *L. macrolides* positively influence the hematological parameters of *C. carpio*, improving both erythropoiesis and immunological responses. This study enhances the understanding of the strategic application of nutritional supplements to augment fish health and productivity in aquaculture environments.

From these findings, it is clear that machine learning holds promise in predicting fish health and optimizing nutritional approaches in aquaculture. The analysis emphasizes the importance of predictive modeling for aquaculture practices that are both effective and sustainable. Although Gradient Boosting showed the highest accuracy, Random Forest performed well as a substitute. These results should and could be further investigated to justify the choice of specific regression models tailored for aquaculture. To increase the effectiveness of nutritional approaches in aquaculture, future research should pursue the following suggestions- The generalisability of the outcomes is limited due to the narrow range of conditions and fish populations. Environmental factors should be included in future studies to gain a clearer understanding of the interactions between diet and hematological parameters. Genomic, microbiome, and metabolomic data could enhance understanding of how dietary changes affect fish health. Broadening the scope to include other aquaculture species may shed more light on how different fish are treated nutritionally from a sustainability standpoint.

4. Conclusion

In conclusion, this analysis illustrates the strength of the Gradient boosting and Random forest models in predicting the target variable from blood parameters of *C. carpio* fed with different feed, as well as the importance of specific factors like Hb and MCHC. As a result, these models and metrics may be useful aids in fish physiological assessments and hematological decision-making processes. The Voting Regressor analysis explains why the predictive models behave differently. While the Gradient boosting and Random forest models provide the highest prediction accuracy, than the Decision tree appears to be less successful. The variability in predictions as training samples increase implies that model performance is influenced by sample size as well as data attributes. more investigation of feature relationships and potential model modifications using approaches like as hyperparameter tweaking may optimize predictive performance even more.

Acknowledgments

The authors are thankful to the Management, Principal, and Head of the Department of Biotechnology, Karpaga Vinayaga College of Engineering and Technology, India for providing the necessary resources to carry out the work.

References

1. M. A. O. Dawood and S. Koshio, Aquaculture **454**, 243 (2015). https://doi.org/10.1016/j.aquaculture.2015.12.033
2. S. K. Brindha, P. Charkarborthy, S. Chenkual, J. Zohmingthanga, J. L. Pautu, P. Nath, A. Maitra, N. S. Kumar, and L. Hmingliana, J. Sci. Res. **14**, 943 (2022). https://doi.org/10.3329/jsr.v14i3.58331
3. M. Calcagnile, S. M. Tredici, and P. Alifano, Heliyon **10**, ID e40892 (2024). https://doi.org/10.1016/j.heliyon.2024.e40892
4. E. Bocioc, V. Cristea, N. Patriche, I. Grecu, A. Antache, and M. C. Mocanu, Bull. University Agricult. Sci. Veterinary Med. Cluj-Napoca Animal Sci. Biotechnol. **72** (2015). https://doi.org/10.15835/buasvmcn-asb:10739
5. S. E. Michael, E. D. Abarike, and J. Cai, J. Fisheries Sci.Com. **13**, 25 (2019). https://doi.org/10.36648/1307-234X.13.3.166
6. D. L. Merrifield, A. Dimitroglou, A. Foey, S. J. Davies, R. T. M. Baker, J. Bøgwald, M. Castex, and E. Ringø, Aquaculture **302**, 1 (2010). https://doi.org/10.1016/j.aquaculture.2010.02.007
7. D. Kirk, E. Kok, M. Tufano, B. Tekinerdogan, E. J. M. Feskens, and G. Camps, Adv. Nutrit. **13**, 2573 (2022). https://doi.org/10.1093/advances/nmac103
8. S. R. Mani, P. Kandhasamy, K. Vijayan, M. Palanichamy, J. P. Jacob, and S. Kandhasamy, Israeli J. Aquaculture - Bamidgeh **73** (2021). https://doi.org/10.46989/001c.25146
9. J. R. Florence, S. S. Priyadharsini, and G. S. Chandran, J. Sci. Res. **14**, 189 (2022). https://doi.org/10.3329/jsr.v14i1.54739
10. L. F. Rahman, M. Marufuzzaman, L. Alam, M. A. Bari, U. R. Sumaila, and L. M. Sidek, Sustainability **13**, 9124 (2021). https://doi.org/10.3390/su13169124
11. Y. Gültepe and N. Gültepe, The Israeli J. Aquaculture - Bamidgeh **72** (2020).
12. X. Yang, S. Zhang, J. Liu, Q. Gao, S. Dong, and C. Zhou, Rev. Aquaculture **13**, 66 (2020). https://doi.org/10.1111/raq.12464
13. F. Aljehani, I. N'Doye, and T.-M. Laleg-Kirati, Aquacult. Eng. **109**, ID 102511 (2025). https://doi.org/10.1016/j.aquaeng.2024.102511
14. P. C. Blaxhall and K. W. Daisley, J. Fish Biol. **5**, 771 (1973). https://doi.org/10.1111/j.1095-8649.1973.tb04510.x
15. B. Velmurugan, E. I. Cengiz, P. Senthilkumaar, E. Uysal, and A. Satar, Environ. Poll. Protection **1**, 32 (2016). https://dx.doi.org/10.22606/epp.2016.11004
16. Dacie and Lewis, Practical Haematology, 9th Edition (Churchill Livingstone, London, 2001).
17. M. Alrbai, S. Al-Dahidi, H. Alahmer, L. Al-Ghussain, H. Hayajneh, B. Shboul, M. Abusorra, and A. Alahmer, Thermal Sci. Eng. Prog. **54**, ID 102784 (2024). https://doi.org/10.1016/j.tsep.2024.102784
18. C. Zhang, A. Kotz, K. Kelly, and L. Rippelmeyer, Transportation Res. Part D Transport Environ. **95**, ID 102843 (2021). https://doi.org/10.1016/j.trd.2021.102843
19. S. Zhao, S. Zhang, J. Liu, H. Wang, J. Zhu, D. Li, and R. Zhao, Aquaculture **540**, ID 736724 (2021). https://doi.org/10.1016/j.aquaculture.2021.736724
20. S. Y. Kuzu, J. Mol. Struct. **1277**, ID 134834 (2022). https://doi.org/10.1016/j.molstruc.2022.134834
21. M. S. Iqbal, T. Khan, S. Kausar, and L. Bin, SN Appl. Sci. **1**, ID 1225 (2019). https://doi.org/10.1007/s42452-019-1270-1
22. N. R Dennis and N.D. Keerthana, Int. J. Data Informat. Intelligent Comput. **3**, 34 (2024). https://doi.org/10.59461/ijdiic.v3i2.114
23. J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, Appl. Sci. **10**, 1042 (2020). https://doi.org/10.3390/app10031042
24. N. Soundhariya, S. Arunkumar, and M. R. Rajan, J. Appl. Nat. Sci. **15**, 1587 (2023). https://doi.org/10.31018/jans.v15i4.4994

25. M. E. Yonar and K. Koprucu, Cellular Molecular Biol. **62**, 49 (2016). https://doi.org/10.14715/cmb/2016.62.13.9
26. A. Şahan, J. Aquacult. Eng. Fish. Res. **1**, 133 (2015). https://doi.org/10.3153/JAEFR15014
27. F. Ejraei, M. Ghiasi, and H. Khara, Archives of Polish Fish. **23**, 163 (2015). https://doi.org/10.1515/aopf-2015-0019
28. F. Fadeifard, M. Raissy, M. Jafarian, H. R. Boroujeni, M. Rahimi, and M. Faghani, Arquivo Brasileiro De Medicina Veterinária E Zootecnia **70**, 199 (2018). https://doi.org/10.1590/1678-4162-8489
29. O. H. Jorum and N. M. Piero, J. Hemat. Thromb. Dis. **04**, ID 1000232 (2016). https://doi.org/10.4172/2329-8790.1000232
30. A. E. Mrong, Md. R. Islam, K. Kole, N. N. Neepa, Md. J. Alam, Md. R. Haque, U. O. Rahman, and G. M. Mostakim, J. Toxicol. **2021**, ID 9417380 (2021). https://doi.org/10.1155/2021/9417380
31. S. Duman, J. Anatolian Environ. Animal Sci. **5**, 212 (2020). https://doi.org/10.35229/jaes.714366
32. P. Satheeshkumar, G. Ananthan, D. S. Kumar, and L. Jagadeesan, Comparative Clinic. Pathol. **21**, 1187 (2011). https://doi.org/10.1007/s00580-011-1259-7
33. O. I. Azeez and S. F. Braimah, Eur. J. Med. Health Sci. **2** (2020). https://doi.org/10.24018/ejmed.2020.2.4.411
34. H. A. A. Hilali, Int. J. Appl. Agricult. Sci. **2**, 17 (2016). https://doi.org/10.11648/j.ijaas.20160201.13
35. A. J. Ali, N. J. Akbar, A. Kumar, S. Vijayakumar, and B. A. John, J. Clean WAS 10 (2018).
36. A. M. Yousafzai, F. Ullah, F. Bari, S. Raziq, M. Riaz, K. Khan, U. Nishan, I. A. Sthanadar, B. Shaheen, M. Shaheen, and H. Ahmad, BioMed. Res. Int. **2017**, ID 5801432 (2017). https://doi.org/10.1155/2017/5801432
37. S. Rajeshwari, and S. P. Sevarkodiyone, Int. J. Aquac. Fish. Sci. **4**, 018 (2018). http://doi.org/10.17352/2455-8400.000038
38. J. A. Hernesniemi, S. Mahdiani, J. A. Tynkkynen, L.-P. Lyytikäinen, P. P. Mishra, T. Lehtimäki, M. Eskola, K. Nikus, K. Antila, and N. Oksala, Ann. Medicine **51**, 156 (2019). https://doi.org/10.1080/07853890.2019.1596302
39. L. Z. Xuan and H. Yang, J. Informat. Commun. Convergence Eng. **21**, 225 (2023). https://doi.org/10.56977/jicce.2023.21.3.225
40. S. A. Wakkas, M. I. Haque, M. Ashaf-Ud-Doulah, and M. Shahjahan, Asian J. Med. Biol. Res. **9**, 23 (2023). https://doi.org/10.56977/jicce.2023.21.3.225
41. A. K. Sahin, SN Appl. Sci. **2,** ID 1308 (2020). https://doi.org/10.1007/s42452-020-3060-1
42. S. Ghimire, P. Guéguen, A. Pothon, and D. Schorlemmer, Nat. Haz. Earth Syst. Sci. **23**, 3199 (2023). https://doi.org/10.5194/nhess-23-3199-2023
43. M. S. Azad, S. S. Khan, R. Hossain, R. Rahman, and S. Momen, PLoS ONE **18**, ID e0296336 (2023). https://doi.org/10.1371/journal.pone.0296336
44. J. Moumouh, S. Benjelloun, A. Latifi, and L. Khamar, MATEC Web Conf. **379**, ID 07008 (2023). https://doi.org/10.1051/matecconf/202337907008
45. U. Sa'adah, M. Y. Rochayani, and A. B. Astuti, Indonesian J. Electr. Eng. Comput. Sci. **21**, 1151 (2021). https://doi.org/10.11591/ijeecs.v21.i2.pp1151-1159
46. P. Gulati, A. Sharma, and M. Gupta, Int. J. Comput. Applicat. **141**, 19 (2016). https://doi.org/10.5120/ijca2016909926
47. F. Gasir, Z. Bandar, and K. Crockett – *Proc. on Int. Conf. on Systems, Man and Cybernetics* (2010).
48. H. Luo, F. Cheng, H. Yu, and Y. Yi, IEEE Access **9**, 55999 (2021). https://doi.org/10.1109/ACCESS.2021.3070575
49. N. Nabi, I. Ahmed, and G. B. Wani, Saudi J. Biol. Sci. **29**, 2942 (2022). https://doi.org/10.1016/j.sjbs.2022.01.019
50. M. Witeska, E. Kondera, and B. Bojarski, Animals **13**, 2625 (2023). https://doi.org/10.3390/ani13162625

51. A. Bani and A. H. Vayghan, Ichthyological Res. **58**, 126 (2011). https://doi.org/10.1007/s10228-010-0199-6
52. P. Uiuiu, C. Lațiu, T. Păpuc, C. Craioveanu, A. Ihuț, A. Sava, C. Răducu, C. Șonea, R. Constantinescu, D. Cocan, and V. Mireșan, Animals **11**, 1810 (2021). https://doi.org/10.3390/ani11061810
53. M. Schonlau and R. Y. Zou, The Stata J. Promot. Commun. Stat. Stata **20**, 3 (2020). https://doi.org/10.1177/1536867X20909688
54. S. Duman, Turkish J. Vet. Animal Sci. **44**, 624 (2020). https://doi.org/10.3906/vet-1910-41
55. A. M. S. Hertika, D. Arfiati, E. D. Lusiana, and R. B. D. S. Putra, F1000Res. **10**, 1169 (2022). https://doi.org/10.12688/f1000research.74117.2
56. H. Sahiti, K. Bislimi, E. Dalo, and K. Murati, Nat. Eng. Sci. **3**, 323 (2018). https://doi.org/10.28978/nesciences.468987
57. I. Ahmed, Q. M. Reshi, and F. Fazio, Aquac. Int. **28**, 869 (2020). https://doi.org/10.1007/s10499-019-00501-3
58. A. Amsath, Curr. Trends Biomed. Eng. Biosci. **7**, ID 555702 (2017). https://doi.org/10.19080/CTBEB.2017.07.555702
59. S. A. Khan, P. Zhou, X. Liu, H. Li, J. Li, Z. U. Rehman, and I. Ahmad, Acta Biochimica Polonica **62**, 581 (2015). https://doi.org/10.18388/abp.2015_1036
60. A. Küçükgül, M. Yüngül, Z. Kahraman, and M. Dörücü, Braz. Archives Biol. Technol. **62**, (2019). https://doi.org/10.1590/1678-4324-2019180400
61. I. Ahmed and Z. A. Sheikh, Europ. Zoological J. **87**, 12 (2020). https://doi.org/10.1080/24750263.2019.1705647
62. N. A. Solaiappan and N. P. S. Leon, Int. J. Life Sci. Pharma Res. **11**, 130 (2021). https://doi.org/10.22376/ijpbs/lpr.2021.11.6.L130-136
63. B. Deivasigamani and A. Q. Elizabeth, Asian J. Fish. Aquatic Res. **19**, 20 (2022). https://doi.org/10.9734/ajfar/2022/v19i230471
64. M. Shahjahan, M. H. Uddin, V. Bain, and M. M. Haque, Fish Phys. Biochem. **44**, 1309 (2018). https://doi.org/10.1007/s10695-018-0522-0
65. S. Duman and A. Şahan, Brazilian Archives Biol. Technol. **61**, ID e18180283 (2018). https://doi.org/10.1590/1678-4324-2018180283
66. M. A. Ogundiran and O. O. Fawole, J. Fish. Sci. Com. **12** (2018).