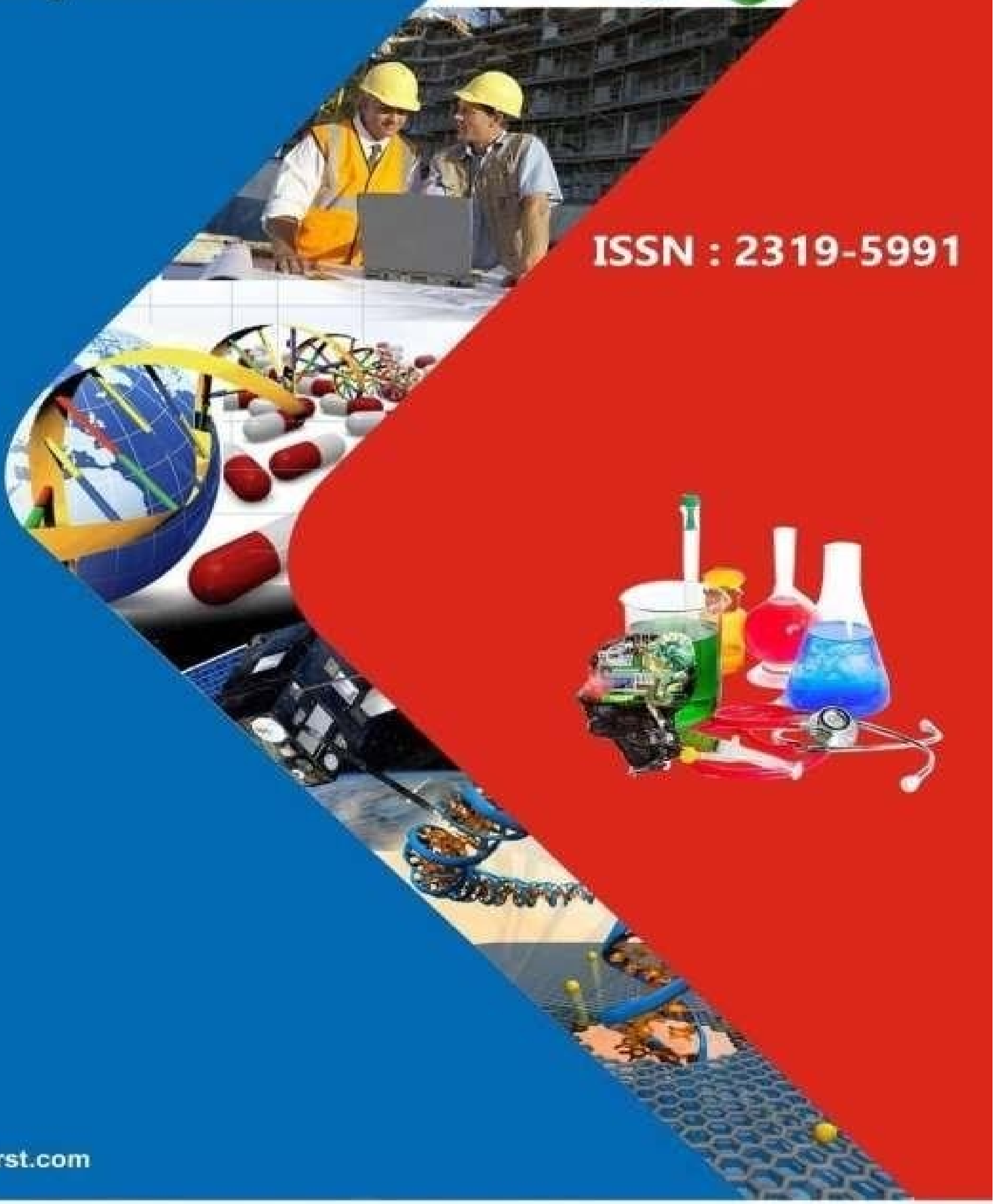
International Journal of Engineering Research and Science & Technology



ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

Int. J. Eng. Res. & Sci. & Tech. 2025

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

Integrating Machine Learning Models forAccurate Water Quality Assessment in Sustainable Fish Farming

T.Vamshi sai[1], K.Sai Krishna Kaushal [2], P.VimalKumar [3], Mrs.CH.Anusha [4]
[1,2,3] UG Scholar, Dept. of ECE, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[4]Assistant Professor, Dept. of ECE, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
vamshisai228@gmail.com

Abstract:

Water quality assessment has been a critical factor in aquaculture for centuries, with early methods relying on direct observation and manual testing. Traditional practices included monitoring parameters like temperature, pH, and dissolved oxygen using basic chemical kits and visual inspections. With technological advancements, digital sensors became available, enabling more precise measurements, though they required manual data analysis and interpretation. The objective of this research is to develop an integrated approach that utilizes machine learning models for accurate water quality assessment, promoting sustainable fish farming practices through real-time monitoring and predictive analysis of water parameters. Before machine learning, water quality in fish farming was assessed manually using chemical kits, physical sensors, and visual inspection of water conditions. Traditional water quality assessment methods in fish farming are labor-intensive, time-consuming, and prone to human error, often resulting in delayed responses to changing water conditions, which can adversely impact fish health and productivity. The proposed system involves deploying digital sensors in fish farming setups to continuously collect data on water quality parameters like pH, temperature, and dissolved oxygen. This data is then processed using machine learning models that predict water quality trends, enabling proactive adjustments to maintain optimal conditions for fish growth. An integrated dashboard provides real-time visualization and alerts for timely interventions.

Keywords: Machine Learning, Water Quality Assessment, Sustainable Fish Farming, Integration, Prediction Models, Aquaculture, Environmental Monitoring, Data-Driven Decision Making, Artificial Intelligence, Water Parameters, Real-Time Monitoring, Sensor Data, Predictive Analytics, Automation, Big Data, Neural Networks, Deep Learning, Support Vector Machines, Random Forest Algorithm, Sustainability.

1.INTRODUCTION

Machine learning-based water quality assessment in fish farming allows for real-time monitoring and analysis of critical parameters like pH, dissolved oxygen, and temperature. This ensures timely intervention, leading to better fish health and yield. Applications include early detection of water contamination, optimization of feeding practices, and automation of water treatment systems. India, one of the largest producers of fish globally, has a thriving aquaculture sector that contributes significantly to the country's economy and food security. The country produced over 14 million metric tons of fish in 2022, with aquaculture accounting for more than 60% of this output.

Despite the growth, the sector faces challenges related to water quality management, which directly affects fish health and yield. Traditionally, fish farmers in India have relied on manual methods, such as chemical kits and physical sensors, to monitor water quality. However, these methods are time-consuming, labor-intensive, and prone to human error, leading to delayed responses to critical changes in water parameters. To improve the efficiency and sustainability of fish farming, there is a need to integrate advanced technologies like machine learning that can provide accurate, real-time water quality assessments and predictive insights.

2. LITERATURE SURVEY

Ma, Long, et.al. (2022) [5] They develop a new Self-Calibrated Illumination (SCI) learning framework for fast, flexible, and robust brightening images in real-world low-light scenarios. To be specific, they establish a cascaded illumination learning process with weight sharing to handle this task. Considering the computational burden of the cascaded pattern, they construct the self-calibrated module which realizes the convergence between results of each stage, producing the gains that only use the single basic block for inference (yet has not been exploited in previous works), which drastically diminishes computation cost. They then define the unsupervised training loss to elevate the model capability that can adapt general scenes. Further, they make comprehensive explorations to excavate SCI's inherent properties (lacking in existing works) including operation-insensitive adaptability (acquiring stable performance under the settings of different simple operations) and model-irrelevant generality (can be applied to illumination-based existing works to improve performance). Finally, plenty of experiments and ablation studies fully indicate our superiority in both quality and efficiency. Applications on low-light face detection and nighttime semantic segmentation fully reveal the latent practical values for SCI.

Wang, Yufei, et.al. (2022) [6] They investigate to model this one-to-many relationship via a proposed normalizing flow model. An invertible network that takes the low-light images/features as the condition and learns to map the distribution of normally exposed images into a Gaussian distribution. In this way, the conditional distribution of the normally exposed images can be well modelled, and the enhancement process, i.e., the other inference direction of the invertible network, is equivalent to being constrained by a loss function that better describes the manifold structure of natural images during the training. The experimental results on the existing benchmark datasets show our method achieves better quantitative and qualitative results, obtaining better-exposed illumination, less noise and artifact, and richer colors.

Hai, Jiang, et.al. (2023) [7] A novel Retinex-based Real-low to Real-normal Network (R2RNet) is proposed for low-light image enhancement, which includes three subnets: a Decom-Net, a Denoise-Net, and a Relight-Net. These three subnets are used for decomposing, denoising, contrast enhancement and detail preservation, respectively. Our R2RNet not only uses the spatial information of the image to improve the contrast but also uses the frequency information to preserve the details. Therefore, our model achieved more robust results for all degraded images. Unlike most previous methods that were trained on synthetic images, they collected the first Large-Scale Real-World paired low/normal-light images dataset (LSRW dataset) to satisfy the training requirements and make our model have better generalization performance in real-world scenes. Extensive experiments on publicly available datasets demonstrated that our method outperforms the existing state-of-the-art methods both quantitatively and visually. In addition, our results showed that the performance of the high-level visual task (i.e., face detection) can be effectively improved by using the enhanced results obtained by our method in low-light conditions.

Xiong, Wei, et.al. (2022) [8] tackle the problem of enhancing real-world low-light images with significant noise in an unsupervised fashion. Conventional unsupervised approaches focus primarily on illumination or contrast enhancement but fail to suppress the noise in real-world low-light images. To address this issue, they decoupled this task into two sub-tasks: illumination enhancement and noise suppression. They proposed a two-stage, fully unsupervised model to handle these tasks separately. In the noise suppression stage, they propose an illumination-aware denoising model so that real noise at different locations is removed with the guidance of the illumination conditions. To facilitate the unsupervised training, they constructed pseudo triplet samples and propose an adaptive content loss correspondingly to preserve contextual details. To thoroughly evaluate the performance of the enhancement models, they build a new unpaired real-world low-light enhancement dataset. Extensive experiments show that our proposed method outperforms the state-of-the-art unsupervised methods concerning both illumination enhancement and noise reduction.

Zheng, Shen, et.al. (2022) [9] proposed a semantic-guided zero-shot low-light enhancement network (SGZ) which is trained in the absence of paired images, unpaired datasets, and segmentation annotation. Firstly, they design an enhancement factor extraction network using depthwise separable convolution for an efficient estimate of the pixel-wise light deficiency of a low-light image. Secondly, we propose a recurrent image enhancement network to progressively enhance the low-light image with affordable model size. Finally, we introduce an unsupervised semantic segmentation network for preserving the semantic information during intensive enhancement. Extensive experiments on benchmark datasets and a low-light video demonstrate that our model outperforms the previous state-of-the-art. They further discuss the benefits of the proposed method for low-light detection and segmentation.

Wu, Yirui, et.al. (2022) [10] proposed an edge computing and multi-task driven framework to complete tasks of image enhancement and object detection with fast response. The proposed framework consists of two stages, namely cloud-based enhancement stage and edge-based detection stage. In cloud-based enhancement stage, they establish connection between mobile users and cloud servers to input rescaled and small-size illumination parts of lowlight images, where enhancement subnetworks are dynamically combined to output several enhanced illumination parts and corresponding weights based on low-light context of input images. During edge-based detection stage, cloud-computed weights offers informativeness information on extracted feature maps to enhance their representation abilities, which results in accurate predictions on labels and positions for objects. By applying the proposed framework in cloud computing system, experimental results show it significantly improves detection performance in mobile multimedia and low-light environment.

Sun, Ying, et.al. (2022) [11] proposed a low-light image enhancement algorithm based on improved multi-scale Retinex and Artificial Bee Colony (ABC) algorithm optimization in this paper. First of all, the algorithm makes two copies of the original image, afterwards, the irradiation component of the original image is obtained by used the structure extraction from texture via relative total variation for the first image, and combines it with the multi-scale Retinex algorithm to obtain the reflection component of the original image, which are simultaneously enhanced using histogram equalization, bilateral gamma function correction and bilateral filtering. In the next part, the second image is enhanced by histogram equalization and edge-preserving with Weighted Guided Image Filtering (WGIF). Finally, the weight-optimized image fusion is performed by ABC algorithm. The mean values of Information Entropy (IE), Average Gradient (AG) and Standard Deviation (SD) of the enhanced images are respectively 7.7878, 7.5560 and 67.0154, and the improvement compared to original image is respectively 2.4916, 5.8599 and 52.7553. The results of experiment show that the algorithm improves the light loss problem in the image enhancement process, enhances the image sharpness, highlights the image details, restores the color of the image, and also reduces image noise with good edge preservation which enables a better visual perception of the image.

3. PROPOSED METHODOLOGY

The hybrid DNN+KNN model, trained on the processed dataset, is utilized to predict water quality parameters on unseen test data. The predictions are added to the test dataset, demonstrating the model's ability to generalize and provide accurate insights into water quality, supporting sustainable fish farming practices. As shown in fig 1.

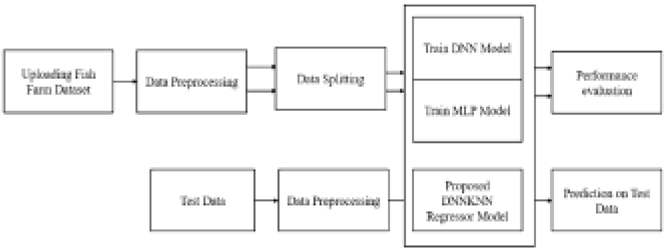


Figure 1: Architectural Block Diagram of the Proposed System.

The proposed methodology typically includes the following key components:

- Step 1: Water Quality Fish Farm Dataset: The research begins with the selection and loading of the dataset containing water quality parameters related to fish farming. This dataset typically includes attributes like temperature, pH, salinity, dissolved oxygen, and other critical factors influencing fish farming environments. These data points form the foundation for the subsequent analysis and modelling steps.
- Step 2: Dataset Preprocessing The raw dataset undergoes preprocessing to ensure it is clean and ready for analysis. This step includes handling null values by imputing missing data using statistical methods like the mode or mean. Outliers are identified and removed using statistical techniques such as Z-score. The data is then scaled using a Standard Scaler for uniformity, ensuring that all features contribute equally during model training. Additionally, the dataset is resampled to balance any class imbalances and to maintain an adequate sample size.
- Step 3: Exploratory Data Analysis (EDA) EDA is conducted to understand the dataset better and identify patterns, trends, and correlations among variables. Techniques like correlation heatmaps, histograms, and boxplots are employed to visualize the relationships between features and detect anomalies. This step provides valuable insights, aiding in feature selection and the refinement of preprocessing steps.
- Step 4: Existing DNN Classifier Algorithm An existing Deep Neural Network (DNN) model is implemented to predict water quality parameters. The DNN architecture comprises multiple layers with activation functions like ReLU and dropout layers for regularization. The model is trained on the processed dataset, optimizing for loss using algorithms like Adam. The performance metrics, such as

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Mean Absolute Error (MAE) and R-squared, are evaluated on the test data.

- Step 5: Existing MLP Classifier Algorithm A Multilayer Perceptron (MLP) model is also employed as a baseline regression algorithm. This feedforward neural network consists of one or more hidden layers and uses the backpropagation technique for optimization. The MLP is trained and tested on the dataset, and its performance is assessed to provide a comparative baseline against the DNN model.
- Step 6: Proposed DNN + KNN Classifier (Hybrid Algorithm) A novel hybrid algorithm combining the strengths of DNN and K-Nearest Neighbors (KNN) is proposed. The DNN is used to extract high-level features from the input data, which are then fed into the KNN model for final prediction. This combination leverages the feature extraction capabilities of DNN and the localized prediction strengths of KNN, providing an advanced and robust approach to water quality prediction.
- Step 7: Performance Comparison The performances of the DNN, MLP, and hybrid DNN+KNN models are compared using metrics like MAE, MSE, RMSE, and R-squared. The comparison is visualized through bar charts and scatter plots, highlighting the hybrid model's improvements in predictive accuracy and efficiency over the standalone models.
- Step 8: Prediction of Output from Test Data with the Trained Model The hybrid DNN+KNN model, trained on the processed dataset, is utilized to predict water quality parameters on unseen test data. The predictions are added to the test dataset, demonstrating the model's ability to generalize and provide accurate insights into water quality, supporting sustainable fish farming practices

Applications:

- **Real-Time Water Quality Monitoring** – ML models analyze sensor data (pH, dissolved oxygen, temperature, turbidity) to provide instant insights on water conditions.
- **Early Detection of Water Contamination** – AI-driven systems identify anomalies in water parameters and predict contamination risks, preventing fish mortality.
- **Optimized Feeding Strategies** – ML-based predictive models adjust feeding schedules based on water quality and fish behavior, reducing waste and improving efficiency.
- **Disease Prediction and Prevention** – ML algorithms detect early signs of disease outbreaks by analyzing water conditions, fish health data, and environmental changes.

Advantages:

1. **Enhanced Performance:** The DNN extracts robust features, making the KNN classifier more effective.
2. **Flexibility:** Works well with small datasets as KNN does not rely on parameter optimization.
3. **Simplicity in Implementation:** The modular approach integrates two well-known algorithms.
4. **Improved Interpretability:** The KNN layer can provide instance-based explanations for predictions.

4. EXPERIMENTAL ANALYSIS

This figure showcases the graphical user interface (GUI) where the water quality dataset specific to a fish farm is uploaded for analysis. The dataset includes key parameters like salinity, dissolved oxygen, pH, water depth, and temperature. Users can seamlessly upload the dataset, ensuring its availability for further processing and model training within the application. As shown in fig 2: Upload of Water Quality Fish Farm Dataset in the GUI Interface.

Figure 2: Upload of Water Quality Fish Farm Dataset in the GUI Interface

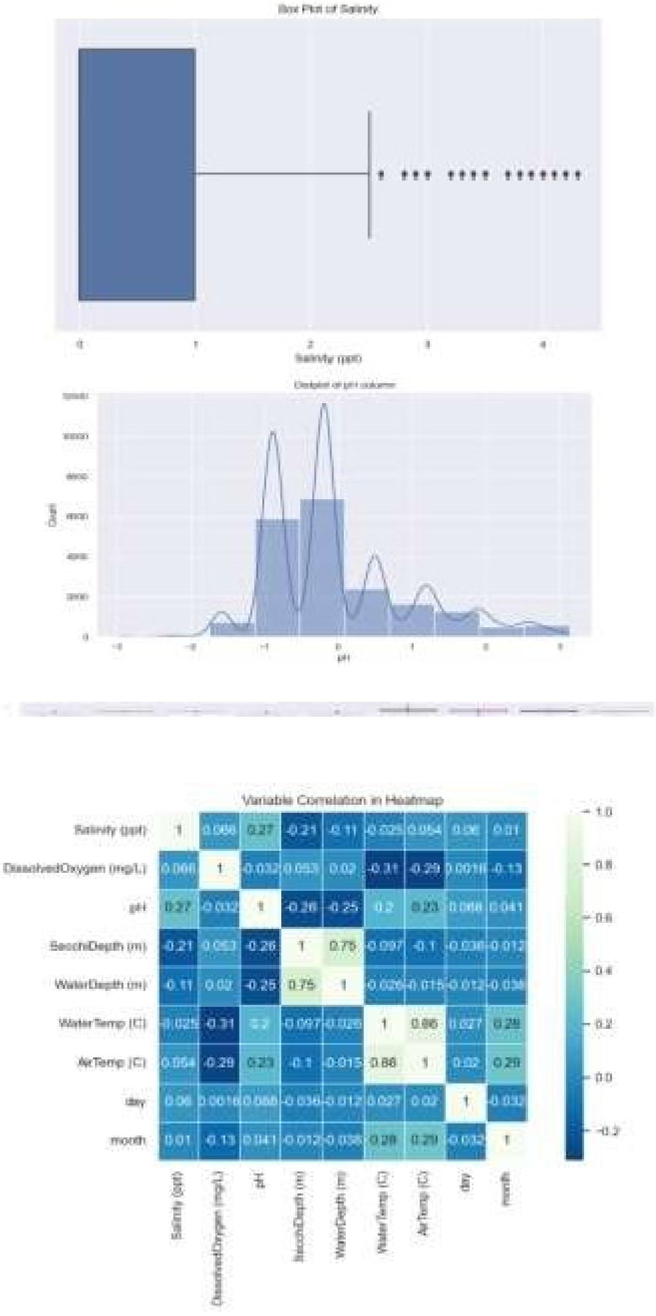


Figure 3: All EDA Plots of the Dataset.

The fig 3: All EDA Plots of the Dataset presents a set of exploratory data analysis (EDA) plots that visualize the distribution and relationships of the dataset's features. It includes histograms, scatter plots, and box plots, providing insights into the data's structure,

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

trends, and possible outliers. EDA is essential to understand the dataset before applying machine learning algorithms, helping identify patterns or areas for preprocessing.

Figure 4: Data Preprocessing in the GUI

This fig 4: Data Preprocessing in the GUI illustrates the data preprocessing steps performed within the GUI. Preprocessing tasks such as handling missing values, feature scaling, and encoding categorical variables are executed here. It ensures that the dataset is clean and standardized for efficient model training, contributing to improved prediction accuracy insubsequent machine learning algorithms.

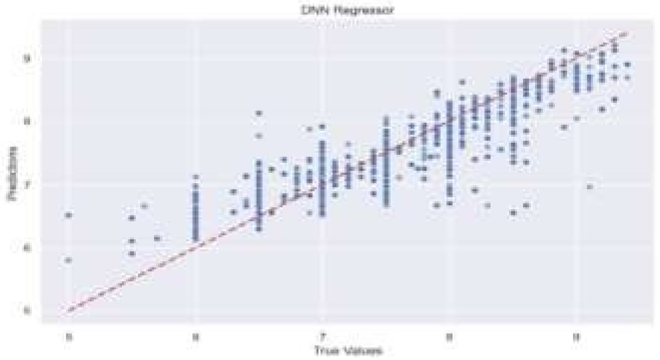


Figure 4(a): Actual vs Predict Plot of DNN Regressor Model

Figure 4(b): Performance Metrics

This fig 4(a): Actual vs Predict Plot of DNN Regressor Model and fig 4(b): Performance Metrics displays the performance metrics of the Deep Neural Network (DNN) regressor model. It includes error measures like Mean Squared Error (MSE) and R-squared values, alongside an "Actual vs. Predicted" plot. The plot compares the model's predicted values against actual values, demonstrating how well the DNN regressor performs in predicting the water quality parameters for the test data.

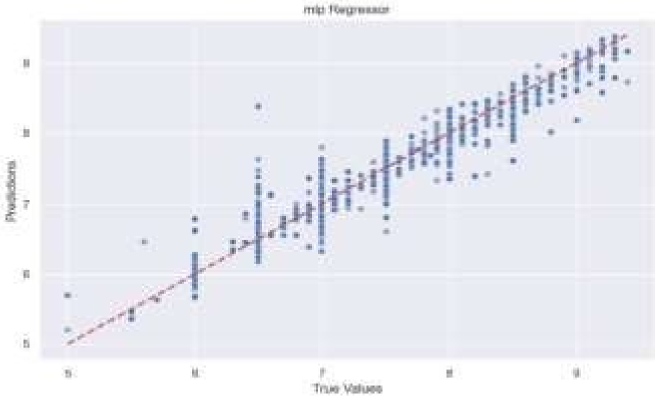


Figure 5(a): Actual vs Predict Plot of MLP Regressor Model

Figure 5(b): Performance Metrics

Similar to fig 4, this figures fig 5(a) and fig 5(b) shows the performance evaluation of the Multilayer Perceptron (MLP) regressor model. It highlights the accuracy of the model in predicting water quality parameters by comparing actual values to predictions. The performance metrics, such as RMSE and R-squared, are displayed to assess the model's precision.

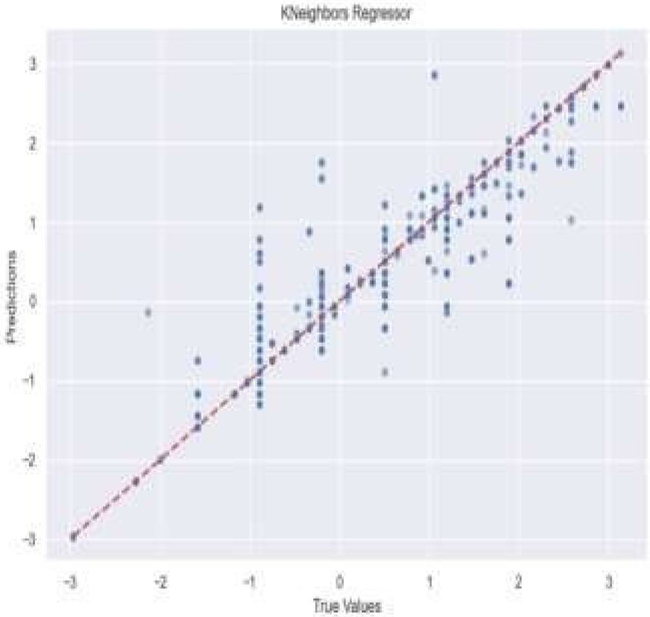


Figure 6(a): Actual vs Predict Plot of DNN+KNN Hybrid Regressor Model

This figure presents the performance of the hybrid model combining DNN and KNN classifiers.

Figure 6(b): Performance Metrics

As shown in figures fig 6(a) and fig 6(b) It compares the predicted water quality values to actual observations in a visual plot, alongside performance metrics. The hybrid approach aims to enhance the predictive power by leveraging both DNN's deep learning capabilities and KNN's instance-based learning.

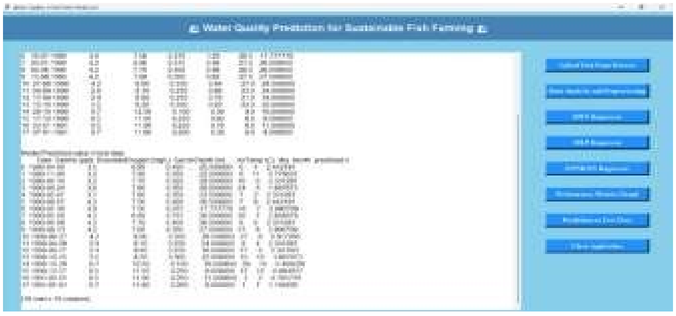


Figure 7: Model Prediction on the Test Data

This fig 7: Model Prediction on the Test Data demonstrates the model's performance on unseen test data, validating its generalization ability. It presents predictions for water quality parameters, comparing them to actual test data values. The accurate predictions validate the effectiveness of the trained models in real-world scenarios, particularly in assessing fish farm water quality.

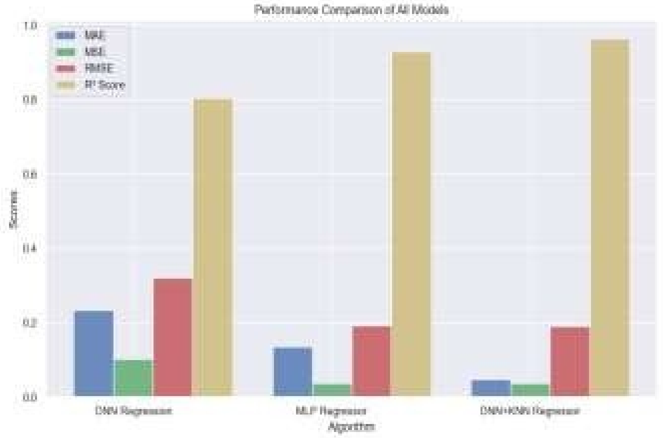


Figure 8: Performance Comparison Graph of All Models

This fig 8: Performance Comparison Graph of All Models shows a comparative analysis of the performance of all the models used in the research: DNN, MLP, and the DNNKNN hybrid model. The performance comparison graph displays metrics such as accuracy, RMSE, or R-squared for each model, helping to identify the best-performing algorithm for water quality prediction in the fish farm dataset.

5. CONCLUSION

The research aimed at analysing water quality and environmental parameters through machine learning models such as DNN (Deep Neural Network) and MLP (Multilayer Perceptron). Various features, such as salinity, dissolved oxygen, pH, and temperature, were considered to predict the quality of the water or its condition over time. After preprocessing and splitting the data into training and testing datasets, several classification models were tested and compared to assess their performance in predicting the water quality. The integration of DNN with KNN classifiers showed promising results in improving prediction accuracy. The research demonstrates the utility of machine learning in environmental monitoring and can significantly contribute to early detection of changes in aquatic ecosystems, guiding sustainable management of water bodies.

REFERENCES

[1] The capacity to foresee is essential for developing and executing proactive water management plans. An important benefit of machine learning in water quality prediction is its capacity to effectively process extensive amounts of data from many sources.

[2] The algorithm is provided with prepared data during the training phase, which may include labels in supervised learning situations or may not have labels in unsupervised learning settings.

[3] The effectiveness of ML models is heavily dependent on the quality of the data used. Inaccuracies, inconsistencies, or gaps in data can lead to poor model performance (Rahat et al., 2023).

[4] In the context of water quality (Zhu et al., 2022), gathering sufficient data that covers various pollutants, different water bodies, and a range of environmental conditions can be challenging.

[5] Rivers, which serve as the primary water source for numerous purposes including industrial and irrigation, are particularly susceptible to environmental degradation due to their constantly evolving characteristics and interaction with waste disposal. Water quality pertains to the overall state or condition of water, including its chemical, physical, and biological aspects. Hence, it is crucial to accurately predict and control the quality of surface water.

[6] Weighted Arithmetic Water Quality Index (WAWQI).

[7] This comprehensive study attempts to achieve the following objectives: 1 to assess and predict water quality in relation to the WQI using three machine learning models: RF, XGBoost, and LightGBM; 2 to employ the SHAP method to interpret the black-box nature of these machine learning models and understand the impact of each water quality parameter on the corresponding predictions.

[8] The WAWQI was employed to compute the WQI utilizing the above specified water quality parameters. Several studies made use of this method to categorize water quality according to its level of purity, using widely identified indicators of water quality.

[9] With the increasing human activities associated with industrialization and urbaniza-tion development, the water quality of coastal rivers is facing escalating and severe threats and degradation.

[10] This approach allows for a comprehensive assessment of the model's per-formance across diverse experiments, reducing the potential influence of random errors associated with a single

experiment.

[11] The entropy weight method is a technique employed to determine the weights of multiple indicators. It utilizes the concept of entropy to measure the uncertainty or diversity of indicators by computing their information entropy.

[12] The discharge of sewage into the receiving water body will significantly increase turbidity and organic and inorganic substances, thereby changing the living environment of marine organisms.

[13] A significant volume of wastewater generated by local residents is often discharged into the sea after undergoing rudimentary water treatment.

[14] Numerical models, which derive results from rules and data.

[15] Despite the feasibility of machine learning-based dissolved oxygen prediction, acquiring water quality data in challenging environments remains a hurdle. Seongsik Park et al. introduced redox potential as a preferred input variable, using machine learning to predict dissolved oxygen—a cost-effective method.