INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 12, Issue 5 - V12I5-1210)
Available online at: https://www.ijariit.com

A Low-Cost Dual-Layer Wastewater Pollution and Toxicity Monitor for MSMEs

Reshma B
reshma.ec23@krct.ac.in
K. Ramakrishnan College of Technology, Tiruchirappalli (Trichy), Tamil Nadu,

Selvalakshmi U
selvalakshmi.ec23@krct.ac.in
K. Ramakrishnan College of Technology, Tiruchirappalli (Trichy), Tamil Nadu,

Megaa S R
megaa.ec23@krct.ac.in
K. Ramakrishnan College of Technology, Tiruchirappalli (Trichy), Tamil Nadu,

Ranjani A
ranjani.ec23@krct.ac.in
K. Ramakrishnan College of Technology, Tiruchirappalli (Trichy), Tamil Nadu,

Keerthana G
gkeerthana@gmail.com
K. Ramakrishnan College of Technology, Tiruchirappalli (Trichy), Tamil Nadu,

ABSTRACT

Biochemical Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) testing is required to monitor pollution load for all medium scale industrial units which generate process wastewater legally. The incubation period for BOD can be up to five days which makes it impractical to perform routine testing for micro, small and medium enterprises (MSMEs), leading to the practice of more frequent pollution monitoring being neglected. In this paper, a cost field-deployable wastewater screening architecture is proposed, which indirectly estimates BOD and COD by two low cost proxy measurements. Turbidity, Electrical conductivity. The proposed system assumes a 2-sensor input space to enable the full system from acquisition to inference to run on a low-power microcontroller without needing external connectivity. This paper outlines the system architecture, data-collection protocol, candidate regression algorithms, and an evaluation plan using regression metrics compared against statutory discharge limits, placing the system in a position to be a cost-effective early-warning screening device as opposed to a certified laboratory replacement.

Keywords: *Wastewater Monitoring, Soft Sensor, Bod Estimation, COD Estimation, Machine Learning, Turbidity, Conductivity, Low-Cost IoT, MSME Pollution Screening, Edge Inference*

1. INTRODUCTION

Industrial and small-commercial water use, such as dyeing, food processing, textile washing, tannery operations and general workshop cleaning creates wastewater that is polluted with chemical matter. To measure that pollution we use BOD, which tells us how much oxygen is eaten by bacteria as they break down material. We also use COD, which tells us the amount of chemicals that can be oxidized even chemicals that cannot be broken down by bacteria. Regulations say that before anyone can dump this water into a drain a water body or, onto land they must keep BOD and COD below limits.

It is normal in an industry with a special environment cell for the cost and time of performing BOD/COD testing. The same is not equitable for a small dyeing operation, or a local food stand. The real result is that no testing is done all. According to this paper, the primary challenge is the cost of sensing and the time required to provide a result. It's not that people do not want to play by the rules. The key to the solution is to develop a low cost and rapid-response system rather than a scaled down version of a lab instrument.

In this investigation we will employ a sensor, a computational model used to estimate a parameter which is difficult or expensive to measure directly (e.g., BOD/COD) based on values which are easily and cheaply measured (e.g., turbidity and conductivity). We only consider the two most common analog sensors (which are the cheapest).

2. RELATED WORK

The previous research on this topic can be categorized into three groups. The first group develops machine-learning sensors to forecast BOD and COD based on information from the treatment plants. These approaches usually have input variables and use ensemble or boosting-based regressors like Random Forest, Gradient Boosting and XGBoost [1]–[4] [8] [9] [13] [14]. These studies demonstrate that these parameters (BOD and COD) can be accurately predicted by indirect measurements. These models are generally for monitoring at the main treatment plant only, however. They are not usually made for use, in industries that are spread out.

The second stream explores IoT prototypes for the monitoring of wastewater. These prototypes log measurements like pH, temperature and turbidity. Send the data to a dashboard to view. They don't attempt to determine BOD. Cod themselves [12]. The third (smallest) stream is about the ability of turbidity and conductivity to predict outcomes. It uses them as proxy variables [7] [11]. Typically, these will be a collection of features rather than just a 2 sensor system for embedded applications. Analyses of soft-sensor development have revealed that this is a developing area. Also, they highlight that there are still gaps. The gaps include difficulty, in interpreting the sensors and trouble generalising the results across types of wastewater [6] [15].

The problem that emerges in each of the three streams is a system that restricts itself to only the possible set of sensors so as not to spend money. It does all its processing right, on the device with no outside help. This paper fills in this joint gap.

Table -1: Comparison of existing approaches vs. proposed system

Approach	Typical Basis	Limitation for MSMEs
Lab BOD/COD testing	Accredited lab, 5-day BOD	Slow, costly, needs technicians
Plant-scale soft sensors	5+ process variables	Needs plant-grade instrumentation
IoT dashboards	pH, temp., turbidity logging	Reports trends, not BOD/COD
Proposed system	Turbidity + conductivity only	Category-specific retraining needed

3. RESEARCH GAP AND OBJECTIVES

No existing low-cost, embedded system that estimates BOD/COD solely from turbidity and electrical conductivity combined with a machine-learning model that is specific to a defined type of wastewater. This work aims at doing:

- Design an end with two sensors that can be immersed in wastewater and is inexpensive.
- Define a protocol for collecting data that pairs sensor readings with certified laboratory BOD and COD values.
- Compare regression algorithms to see how accurately they estimate BOD and COD.
- Evaluate the system honestly against values and report its accuracy limits instead of claiming it is a laboratory-equivalent instrument.

4. PROPOSED SYSTEM ARCHITECTURE

This occurs only once in the development. It is carried out for a type of wastewater. The second one is the inference phase. This is always running following system deployment in the field.

4.1 Hardware Components

Table -2: Hardware components and function

Component	Function
Turbidity sensor	Suspended-particle cloudiness
Conductivity / TDS	Dissolved-ion concentration
MCU (ESP32)	Acquisition + on-device inference
Alert module	LED / buzzer / app notice
Housing	IP-rated waterproof enclosure

4.2 System Block Diagram

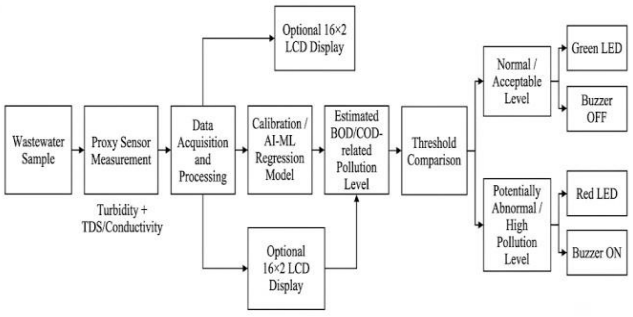


Figure -1: End-to-end system block diagram from sensing to alert.

4.3 Data Acquisition and Training Pipeline

This helps to include the changes that happen in the level of pollution. Each sample is separated into two parts. One part is sent to a certified laboratory to get the BOD and COD values. The other part is tested then and there using a sensor probe to measure turbidity and conductivity. The data that comes from these two tests together makes up the training set.

4.4 Machine Learning Model Selection

The pipeline, Cross-validates several candidate regressors. Multivariate Linear Regression as a baseline, Random Forest and XGBoost as ensemble tree-based methods and Support Vector Regression, as a kernel-based alternative. And chooses the model that works best on a held-out validation split while staying small enough to run on a microcontroller.

4.5 Embedded Deployment

The chosen model is turned into a light-weight inference process. It gets built into the microcontroller's firmware. This way the main estimation function works without needing any connection. If the ESP32 has Wi-Fi it can be used optionally. That's for logging data and sending alerts when needed.

4.6 Indicative Cost Analysis

A rough list of parts for one unit that can be deployed. Is expected to be affordable for a business very different from laboratory-grade online BOD/COD analyzers, which usually cost much more and still need chemicals and regular maintenance. The one-time cost of testing in a lab during the training period is seen as a development expense.

5. METHODOLOGY

5.1 Problem Formulation

The estimation task is set up as a regression problem. We are given an input feature vector x = [turbidity conductivity]. The goal is to find a function f that can predict the reference laboratory values y = [BOD, COD], for the sample. This function f is learned by minimizing the squared-error loss across the training set.

$$L(\theta) = (1/N)\ \Sigma_i \| f(x_i;\ \theta) - y_i \|^2 \qquad (1)$$

where theta denotes the model parameters and N is the number of training samples with regularization applied to avoid overfitting on a modest field dataset on the order of 100, to 200 samples.

5.2 Feature Engineering

In addition to the two raw sensor readings we also evaluate features, such as the turbidity-to-conductivity ratio and the temperature-compensated conductivity as possible inputs. On the trained models. This process confirms that both raw sensor readings add value to the prediction.

5.3 Model Training and Validation

The dataset is divided using k-fold cross-validation with k equal to 5. This approach helps get an accuracy estimate because the sample size is small which is common in single-category field studies. For the tree-based models hyperparameters are optimized using grid search. This tuning happens inside the -validation loop to avoid overfitting. Once the best hyperparameters are found the final model is trained again on the dataset. After that the model is prepared for deployment, on embedded systems.

5.4 Embedded Inference Procedure

The on-device inference routine follows a fixed sequence at each measurement cycle: (i) sample the turbidity and conductivity ADC channels and average, over readings to reduce sensor noise; (ii) apply the stored calibration curve and temperature adjustment; (iii) evaluate the trained regression function on the calibrated feature vector to get BOD and COD; and (iv) compare both estimates against the category-specific discharge limits stored in firmware. I designed this sequence to finish in a seconds so that the on-device inference routine can be used interactively right at the point of discharge without a separate sampling-and-wait step.

6. SYSTEM WORKFLOW

- Step 1 — Immerse the turbidity probe and the conductivity probe into the wastewater sample.
- Step 2 — Collect raw analog readings. Then apply calibration factors and temperature compensation to those readings.
- Step 3 — Use the on-device regression model to calculate the estimated BOD value and the estimated COD value.
- Step 4 — Compare the estimated BOD value and the estimated COD value to the discharge limit that we have on hand.
- Step 5 — Trigger an alert if the estimate exceeds the limit.
- Step 6 — Optionally save the readings locally. Then send them to a dashboard, for trend monitoring.

7. EXPERIMENTAL DESIGN AND EVALUATION METRICS

The evaluation plan requires collecting least one hundred to one hundred fifty paired samples from a single clearly defined wastewater category. These samples must come from production batches from various times of day and, from different seasonal conditions so that the plan does not overfit to one narrow operating regime. A held-out test set, twenty percent of the samples and stratified by pollution-load range is kept separate to give an unbiased final accuracy estimate.

I measure model performance by checking the coefficient of determination (R^2) the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) that compare predicted values with laboratory-measured BOD/COD. I also evaluate the system by looking at the percentage of test samples where the estimated pollution level falls on the side of the regulatory discharge limit as the certified laboratory value because this binary screening decision is the main real-world function of the device.

Table -3: Target evaluation metrics for field viability

Metric	Target Threshold
R^2 (BOD, COD)	≥ 0.80 (indicative)
RMSE	Within category noise band
Binary screening agreement	≥ 85%
Inference latency	< 5 seconds on-device

8. NOVELTY AND CONTRIBUTION

The contribution of this work is not in discovering that turbidity and conductivity relate to pollution load. That connection is already well known in sensing studies. The real contribution lies in limiting the sensing and computation setup, to the smallest possible configuration that still enables continuous wastewater screening. This makes the system affordable and practical for a small business owner. It also includes an evaluation framework that clearly presents the system as a screening aid not as a certified or equivalent measurement instrument.

9. LIMITATIONS

- Turbidity and conductivity are not measurements and may not show all the types of pollution that can be present.
- This estimate is not a measurement that can be used for rules and regulations and should not take the place of tests done in a certified lab when it is required by law.
- Sensors can lose accuracy, over time and living things can get on them so there needs to be a plan to check and fix them to make sure they work properly in the field for a long time.

10. DISCUSSION

The proposed architecture gives up continuous-parameter accuracy to make the system more accessible. A soft sensor that runs on a treatment-plant scale and uses five or more process variables will usually beat a two-sensor system when you look at raw regression accuracy. However that comparison is not the one that matters to the target user. The comparison that matters is between the proposed system and the current situation, for an MSME, which's no monitoring at all. Viewed this way even a screening tool that has moderate continuous accuracy but a very good binary screening agreement is a big improvement. It turns a discharge event into a flagged event that triggers human review.

Another implication is about how big the deployment can be. Since the model is trained for each type of wastewater of for each treatment plant the same overall process. Including the sensors how data is collected, comparing different models and converting the data. Can be used for many small businesses by just training it with data that fits the category. This ability to reuse the system is why the low-cost method works economically when expanding across a group of small businesses. It is different, from making a tool just for one customer.

11. CONCLUSION AND FUTURE SCOPE

This project has presented the design of a cost machine-learning-based proxy-sensor system. The system is meant to help scale industrial wastewater generators. It offers a way to get near-instant screening for BOD and COD. The system uses turbidity and conductivity as inputs. It runs on a microcontroller for on-device inference. Future work includes expanding the training dataset to cover types of wastewater. This could reduce the number of paired laboratory samples needed for each wastewater category. Another goal is to integrate a self- sensor-drift detection routine. Finally we aim to conduct a field pilot with a group of MSMEs. The purpose is to validate how well the system performs in real-world conditions. We will compare its screening results, with accredited laboratory testing.

REFERENCES

[1] B. S. Pattnaik, A. S. Pattanayak, S. K. Udgata, and A. K. Panda, “Machine learning based soft sensor model for BOD estimation using intelligence at edge,” Complex & Intelligent Systems, vol. 7, pp. 961–976, 2021, doi: 10.1007/s40747-020-00259-9.

[2] P. M. L. Ching, X. Zou, D. Wu, R. H. Y. So, and G. H. Chen, “Development of a wide-range soft sensor for predicting wastewater BOD_5 using an eXtreme gradient boosting (XGBoost) machine,” Environmental Research, vol. 210, Art. no. 112953, 2022, doi: 10.1016/j.envres.2022.112953.

[3] I. K. Mutai, K. Van Laerhoven, N. W. Karuri, and R. K. Tewo, “Using Multivariate Linear Regression for Biochemical Oxygen Demand Prediction in Waste Water,” arXiv preprint arXiv:2209.14297, 2022, doi: 10.48550/arXiv.2209.14297.

[4] A. S. Qambar and M. M. Al Khalidy, “Prediction of municipal wastewater biochemical oxygen demand using machine learning techniques: A sustainable approach,” Process Safety and Environmental Protection, vol. 168, pp. 833–845, 2022, doi: 10.1016/j.psep.2022.10.033.

[5] Y. Xie, Y. Chen, Q. Lian, H. Yin, J. Peng, M. Sheng, and Y. Wang, “Enhancing Real-Time Prediction of Effluent Water Quality of Wastewater Treatment Plant Based on Improved Feedforward Neural Network Coupled with Optimization Algorithm,” Water, vol. 14, no. 7, Art. no. 1053, 2022, doi: 10.3390/w14071053.

[6] H. -Y. Shyu, C. J. Castro, R. A. Bair, Q. Lu, and D. H. Yeh, “Development of a Soft Sensor Using Machine Learning Algorithms for Predicting the Water Quality of an Onsite Wastewater Treatment System,” ACS Environmental Au, vol. 3, no. 5, pp. 308–318, 2023, doi: 10.1021/acsenvironau.2c00072.

[7] E. Aghdam, S. R. Mohandes, P. Manu, C. Cheung, A. Yunusa-Kaltungo, and T. Zayed, "Predicting quality parameters of wastewater treatment plants using artificial intelligence techniques," Journal of Cleaner Production, vol. 405, Art. no. 137019, 2023, doi: 10.1016/j.jclepro.2023.137019.

[8] P. M. L. Ching, R. H. Y. So, and T. Morck, "Advances in soft sensors for wastewater treatment plants: A systematic review," Journal of Water Process Engineering, 2022.

[9] R. Mathur, M. K. Sharma, K. Loganathan, M. Abbas, S. Hussain, G. Kataria, M. S. Alqahtani, and K. S. Rao, "Modeling of two-stage anaerobic onsite wastewater sanitation system to predict effluent soluble chemical oxygen demand through machine learning," Scientific Reports, vol. 14, Art. no. 1835, 2024, doi: 10.1038/s41598-023-50805-x.

[10] O. Inbar, M. Shahar, and D. Avisar, "Predictive modeling of BOD throughout wastewater treatment: A generalizable machine learning approach for improved effluent quality," Environmental Science: Water Research & Technology, vol. 10, pp. 2577–2588, 2024, doi: 10.1039/D4EW00111G.

[11] M. A. P. Cechinel, J. Neves, J. V. R. Fuck, R. C. de Andrade, N. Spogis, H. G. Riella, N. Padoin, and C. Soares, "Enhancing wastewater treatment efficiency through machine learning-driven effluent quality prediction: A plant-level analysis," Journal of Water Process Engineering, vol. 58, Art. no. 104758, 2024, doi: 10.1016/j.jwpe.2023.104758.

[12] D. Costa, Y. Bayissa, K. V. Barbosa, M. Dias Villas-Boas, A. Bawa, J. Lugon, A. J. Silva Neto, and R. Srinivasan, "Water quality estimates using machine learning techniques in an experimental watershed," Journal of Hydroinformatics, vol. 26, no. 11, pp. 2798– 2814, 2024, doi: 10.2166/hydro.2024.132.

[13] L. Y. K. Wang, J. C. H. Loo, Y. X. Wong, and Y. Y. Lai, "IoT Based Wastewater Temperature, pH and Turbidity Monitoring with Data Analytics and Visualisation," in Proceedings of the 12th International Conference on Information Technology, 2024, pp. 213–218, doi: 10.1145/3718391.3718428.

[14] M. Hassnain, S. M. W. Lee, and M. R. Azhar, "Soft sensing of biological oxygen demand in industrial wastewater using machine learning models," Journal of Water Process Engineering, vol. 78, Art. no. 108699, 2025, doi: 10.1016/j.jwpe.2025.108699.

[15] Gahrouei et al., "Enhancing the interpretability of spatially variable N_2O model predictions with soft sensors during wastewater treatment," 2026.