

IMAGE PROCESSING AND MACHINE LEARNING APPROACH TO ESTIMATE TURBIDITY IN RAW WATER FOR TREATMENT PLANTS IN SRI LANKA

J.M.R.N. Ambagolla^{1*}, K.H.P. Madusanka¹, H.P.A.M. Siriwardana¹, Akarshani Amarasinghe²

¹ Department of Civil Engineering, University of Sri Jayewardenepura, Mattegoda, Sri Lanka

² Department of Computer Engineering, University of Sri Jayewardenepura, Mattegoda, Sri Lanka

**Corresponding author: rovimango@gmail.com*

Abstract

Conventional turbidity monitoring in water treatment plants requires frequent manual sampling and laboratory-based measurements, which can limit the ability to achieve continuous and cost-effective monitoring. This research developed an image-based machine learning approach for estimating raw water turbidity using digital images, evaluated as a proof-of-concept alternative monitoring method. A total of 200 raw water samples were collected from the Ambatale Water Treatment Plant laboratory, and corresponding digital images were acquired using a DSLR camera under controlled conditions. The measured turbidity ranged between 6.59 and 10.73 nephelometric turbidity units (NTU). Convolutional neural network (CNN) models were developed and evaluated using image data paired with laboratory-measured turbidity values. The results demonstrated that the improved augmented model achieved the highest prediction performance with a root mean square error (RMSE) of 0.51 and the lowest mean absolute error (MAE) of 0.374, a coefficient of determination of 0.517 as well as a Pearson correlation of 0.740, showing a moderate-to-strong agreement with measured turbidity values while reducing estimation errors compared with the initial models. Furthermore, a conceptual hardware prototype and software monitoring dashboard were developed to evaluate the practical implementation potential of the proposed approach. The study concludes that image-based machine learning can serve as a potential supporting tool for real-time turbidity monitoring in water treatment facilities; however, given the narrow turbidity range and single-plant dataset, these results can only validate a successful early proof of concept; broader datasets must be developed and tested using different plants across wider operational turbidity ranges for island-wide real-time deployment.

Keywords: Water quality monitoring, Computer vision, Real-time monitoring, Convolutional neural networks, Water infrastructure development

1. Introduction

Turbidity is a key water quality parameter that is monitored throughout the raw water treatment process including during the water intake, disinfection and coagulation-flocculation control phases [2]. Conventional turbidity monitoring in Sri Lanka relies on manual sampling and laboratory-based instruments, which are costly, time-consuming, and prone to delayed response to rapid fluctuations in water quality [1]. Automated inline sensors offer real-time monitoring, but their high cost continues to limit widespread adoption in many treatment facilities [1].

Image-based monitoring has emerged as a new and promising low-cost alternative, using digital cameras to capture variations in colour intensity that can be correlated with known corresponding turbidity levels [2]. Recent advances in machine learning, particularly

convolutional neural networks (CNNs), have further improved the accuracy of such image-based estimation methods by automatically extracting complex visual features from sample images [3,4].

Building on these advances, this study develops and evaluates the suitability of using an image-based machine learning approach for estimating water turbidity for raw water treatment plants in Sri Lanka. Specifically, it assesses the practical feasibility of implementing such a system with existing plant infrastructure and evaluates the prediction accuracy achievable using geo-specific, real-time turbidity sample images collected in a local raw water treatment plant.

2. Methodology

Site investigations and pilot testing were conducted at the Ambatale Water Treatment Plant, which established a controlled, repeatable image acquisition protocol. A total of 200 raw water samples were collected from the laboratory raw water line across 2 days of testing. For each sample, turbidity was measured three times using a calibrated Hach 2100Q portable turbidimeter and averaged to an NTU value, and a corresponding image was captured using a Nikon D3500 DSLR camera under standardised conditions (matte white background, 30 cm camera distance, 115–120 lux illumination, fixed exposure) which were determined through the pilot testing phase.

The dataset was split 160/40 for training and testing. A regression-based convolutional neural network (CNN) was developed with PyTorch and evaluated across three configurations: ML1 (a pre-trained baseline model), ML2 (retrained on the new dataset), and ML3 (retrained with augmentation consisting of rotation, brightness/contrast adjustment, and cropping), which artificially expanded the training set to 640 images. Each CNN consisted of stacked convolutional and pooling layers for feature extraction followed by fully connected layers producing a continuous NTU output. For ML2 and ML3, multiple epoch settings (25, 50, 100, 200 and 500 for ML2; 25, 50 and 100 for ML3) were tested, and the best-performing configuration was selected by comparing MAE, RMSE, R^2 , Pearson correlation and residual analysis on the 40 held-out test images. A preliminary hardware and dashboard feasibility concept was also created.

3. Results and Discussion

The final dataset consisted of 200 raw water samples drawn from the Kelani River intake, spanning a turbidity range of 6.59–10.73 NTU.

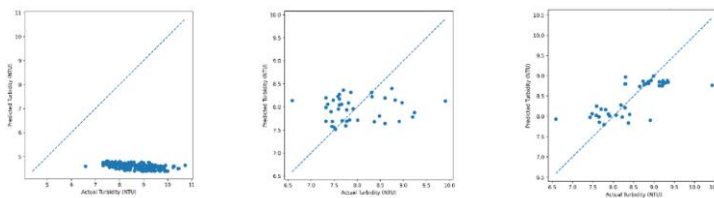


Figure 1: Predicted versus actual turbidity for ML1 (left), ML2 (middle) and ML3 (right).

As seen in Figure 1, when applied to this dataset, the pre-trained baseline model (ML1) generalised poorly as predictions fell substantially below measured values, reflecting a

systematic bias and producing a root mean square error (RMSE) of 4.174. Re-training a model directly on the newly collected samples (ML2, 25 epochs) brought the RMSE down to 0.697, though this and most other ML2 configurations produced negative R^2 values and weak Pearson correlations, indicating that predictions tend to be compressed toward the mean. The strongest

performance came from the augmented model (ML3) as shown in Figure 1, which achieved the lowest RMSE (0.510) and mean absolute error (0.374), together with a positive coefficient of determination (0.517) and moderate-to-strong Pearson correlation (0.740). These results indicate that augmenting a limited image dataset substantially improves CNN robustness and predictive accuracy.

Beyond model performance, a conceptual monitoring system was designed to translate these results into a deployable tool. A preliminary hardware enclosure was modelled using Fusion CAD software. This hardware concept was paired with a prototype software dashboard for plant operators, which converts the CNN model's raw output into a live, interpretable data stream showing real-time turbidity estimates, running averages, and historical trend graphs as seen in Figure 2.



Figure 2: Conceptual hardware prototype (left) and software monitoring dashboard (right) for real-time turbidity estimation.

4. Conclusion

This proof-of-concept study demonstrates that CNN models, enhanced through data augmentation, can estimate raw water turbidity from digital images, achieving promising prediction accuracy with minimal systematic bias under tested conditions. The conceptualised hardware and software prototypes further support the practical feasibility of deploying this approach as a low-cost supporting tool, though not yet a validated replacement for conventional methods. Future work should expand data collection across multiple plants and operational turbidity ranges as well as validate the hardware prototype in the field.

References

- [1] Kadaruddin, K. and Zainuddin, M. (2020), "A brief review on low-cost turbidimeter designs," *IOP Conference Series: Earth and Environmental Science*.
- [2] Mohammed, A. (2024), "Determine water turbidity by image processing technology," *International Journal of Intelligent Systems and Applications in Engineering*.
- [3] Matos, T., Martins, M.S., Henriques, R. and Goncalves, L.M. (2024), "A review of methods and instruments to monitor turbidity and suspended sediment concentration," *Journal of Water Process Engineering*.
- [4] Nie, Y., Chen, Y., Guo, J., Li, S., Xiao, Y., Gong, W. and Lan, R. (2024), "An improved CNN model in image classification application on water turbidity," *Research Square*.

Image Processing and Machine Learning Approach to Estimate Turbidity in Raw Water for Treatment Plants in Sri Lanka

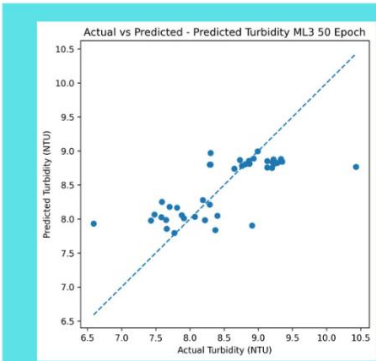
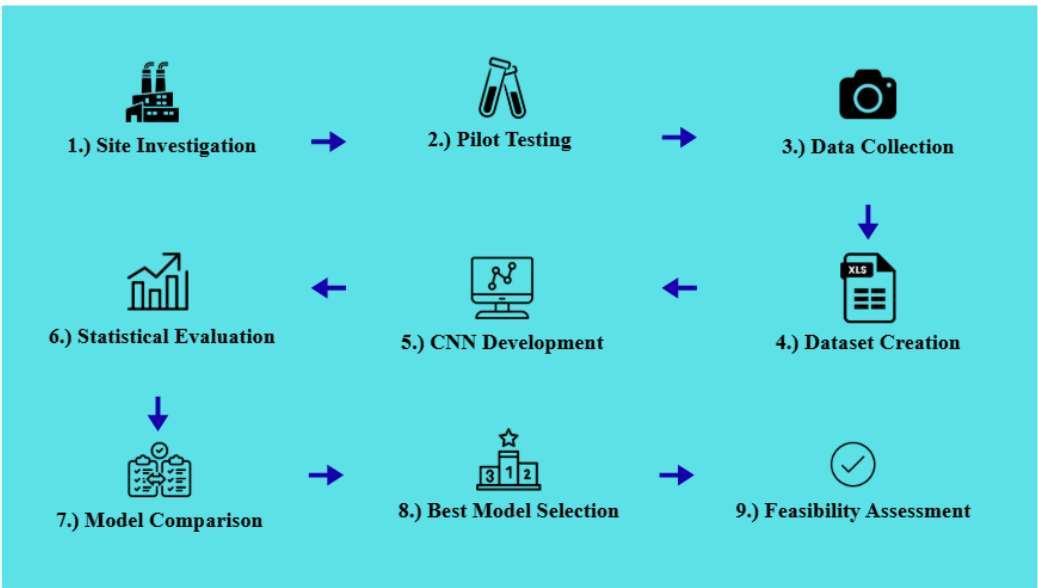


Image-based CNN model estimates raw water turbidity with accuracy, supported by a conceptual hardware and dashboard design for low-cost continuous monitoring