BIBECHANA
Vol. 20, No. 3, December 2023, 248–258

ISSN 2091-0762 (Print), 2382-5340 (Online)
Journal homepage: http://nepjol.info/index.php/BIBECHANA
Publisher:Dept. of Phys., Mahendra Morang A. M. Campus (Tribhuvan University)Biratnagar

Time and Space Domain Prediction of Water Quality Parameters of Bagmati River Using Deep Learning Methods

Pujan Bashyal[1], **Mandira Pradhananga Adhikari**[2,*], **Nanda Bikram Adhikari**[1]

[1] *Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering (IOE),Tribhuvan University, Lalitpur, Nepal*
[2] *Central Department of Chemistry, Tribhuvan University, Kirtipur, Kathmandu, Nepal*

[*]Corresponding author. Email: mandira43@hotmail.com;adhikari@ioe.edu.np

Abstract

Bagmati river is biologically, geologically, religiously and historically significant among the river systems of the Kathmandu Valley. The river is affected by five major tributaries, including Manohara, Dhobi Khola, Tukucha, Bishnumati, and Balkhu Khola, which significantly impact the water chemistry inside the Kathmandu Valley. The data of water quality parameters pH, dissolved oxygen, turbidity, temperature, oxygen reduction potential, conductivity, total dissolved solids, salinity among others was collected using fixed sensors (in period of 5 seconds) and mobile sensors (with latitude and longitude) along the river. The observation is important for two reasons, one because it was collected in real-time and fine scale, which is not normally possible with traditional ways, and next such observation was done for the first time in Bagmati River. The aim of this study was to predict water quality parameters of the Bagmati River using machine learning time series models, specifically ARIMA and LSTM. The LSTM model was designed with one input layer, one encoder layer, one repeat layer, one decoder layer, and one output dense layer to separate the output into temporal slices. Additionally, a DNN model was employed for location-based prediction, utilizing two input layers for latitude and longitude and seven output layers for the seven water quality parameters considered for study. The models demonstrated promising performance, but further data collection and parameter variation are recommended for continued optimization.

Keywords
ARIMA, DNN, LSTM, spatial prediction, temporal prediction, time series models.

Article information
Manuscript received: August 18, 2023; Accepted: September 14, 2023
DOI https://doi.org/10.3126/bibechana.v20i3.57736

1 Introduction

The Bagmati River holds tremendous religious, historical, and ecological significance in Nepal. Originating from the Shivapuri Hills, this river passes through the culturally important Pashupatinath Temple and other areas in the Kathmandu Valley. Unfortunately, the remarkable growth of population and unplanned urbanization in the Kathmandu valley has led to severe deterioration of the Bagmati's water quality [1]. The water pollution poses major environmental and health hazards, demanding urgent assessment and remediation. Periodic measurement of critical water quality parameters like pH, dissolved oxygen, temperature, turbidity etc. provides vital insights into the pollution levels and overall river health. However, the conventional methods of manual sample collection and lab-based analysis are extremely time-consuming, labor-intensive and expensive. The sparse, delayed data obtained through traditional monitoring is insufficient to capture the high-resolution spatio-temporal dynamics of a complex river system [1]. Recent technological advances have enabled real-time, automatic and fine-scale sensing of water quality through fixed and mobile probes . Additionally, machine learning models like long short-term memory (LSTM) networks and deep neural networks (DNN) have shown promise in effectively analyzing and predicting water quality data . However, these modern techniques are yet to be implemented or evaluated for Bagmati River. Therefore, this study aims to implement machine learning models on this data to predict the water quality parameters in time and space domain. Successful implementation of this approach can provide an efficient alternative to traditional techniques for river monitoring in developing regions facing resource constraints. Furthermore, the spatio-temporal insights obtained from data-driven modelling can strengthen pollution control policies and remediation efforts for the Bagmati River.

The Bagmati River in Nepal emerges as a critical focus for management due to its alarming pollution levels. Stretching approximately 51 kilometers through the culturally significant Kathmandu Valley and covering a catchment area of around 678 square kilometers, this river holds significant biological, geological, and historical importance. Unfortunately, it has been significantly impacted by the influx of pollutants from five major tributaries – Manohara, Dhobi Khola, Tukucha, Bishnumati, and Balkhu Khola – which substantially alter its water chemistry. The uncontrolled urban growth within the Kathmandu Valley has led to the deterioration of the river's water quality, with untreated sewage and waste being directly discharged into its waters. The river has devolved into a repository for solid waste, untreated domestic, industrial, and agricultural effluents. Consequently, accurate prediction of water quality parameters at various GPS locations along the Bagmati River is imperative for effective pollution management and mitigation.

The research conducted by Adhikari et al. [1] endeavors to comprehensively profile and characterize pollutants in real-time and space. Water quality encompasses the physical, chemical, biological, and radiological attributes that define water's appropriateness for specific applications. Parameters such as temperature, pH, dissolved oxygen, nutrients, metals, and pollutants collectively define water quality and significantly impact aquatic life, ecosystems, and human well-being. This quality is often evaluated against established standards for various parameters, including temperature, pH, oxidation-reduction potential (ORP), electrical conductivity (EC), resistivity (RES), total dissolved solids (TDS), salinity (Sal), dissolved oxygen (DO), turbidity (Turb), biochemical oxygen demand (BOD), chemical oxygen demand (COD), nitrogen, phosphorus, and pollutants like PM2.5, CO2, formaldehyde, and volatile organic compounds (VOCs). Each parameter holds a defined range of values deemed safe for both human and aquatic life. Temperature influences water's physical and chemical properties, while pH indicates its acidity or alkalinity. EC and RES reflect water's conductivity, correlating with dissolved ion concentration. TDS and Sal measure dissolved solids and salts, while DO is crucial for aquatic survival. Turbidity gauges water clarity, with pollutants like PM2.5, CO2, formaldehyde, and VOCs emerging as additional concerns. The present study focuses on key parameters including Temperature, pH, ORP, EC, RES, TDS, Sal, DO, and Turbidity. Utilizing sophisticated mobile tracers and analyzers, data collection was conducted along the Bagmati River using a rafting boat, as well as through a fixed sensor system stationed within the Kathmandu Valley. These sensors recorded real-time data of various physical parameters, including pH, conductivity, salinity, total dissolved solids (TDS), dissolved oxygen (DO), temperature, and turbidity.

Among the 14 designated data collection sites, the upstream Gokarna site (B-1) represents the most remote point, situated approximately 8 kilometers from the entry point (Sundarijal) of the Bagmati River into the Kathmandu Valley. In this rural setting, the sources of pollutants are less apparent. Moving along the river, the B-2 and B-3 sites are positioned just upstream and downstream of the Guheshwori Wastewater Treatment Plant, respectively. These sites encompass a blend of residential and industrial areas, including wool dying companies, medical colleges, and hotels. While the wastewater treatment plant treats sewage before discharge, pollutants from various sources still affect the river. The Guheshwori temple site (B-4) is impacted by activities such as bathing, washing, and picnicking along the riverbanks. The Gaurighat site (B-5) captures the influence of local residential areas before the river reaches the revered Pashupatinath temple. This temple, of great cultural and religious significance, witnesses various practices such as bathing and rituals, including cremation. Unfortunately, the ash from cremations is directly released into the river. Pashupatinath temple (Aryaghat) (B-6) observes the discharge of untreated wastewater and effluents from the treatment plant into the Bagmati river. Tilganga (B-7), just downstream of the Guheshwori Wastewater Treatment Plant, portrays the aftermath of these discharges. Tinkune (B-8) reveals a disturbing sight of solid waste and sewer lines leading directly to the river, resulting in visibly dark and turbid water. The subsequent sites (B-9 to B-14) serve to assess the effects of tributaries on the Bagmati River, with the Thapathali site (R-11) illustrating the water quality above the highly polluted Tukucha tributary. Observations at the Shankhamul site provide insight into daily and diurnal variations before tributaries influence the river.

Adhikari et al.'s research focused on the analysis of water quality parameters, highlighting the limitations of conventional fixed-point measurements in capturing spatial and temporal profiles. Collecting data along the Bagmati River frequently proves to be both time-consuming and costly. Despite these constraints, the application of time and space domain data modeling could facilitate insightful conclusions and support the research objectives. By leveraging mathematical and statistical techniques, machine learning analyzes data, identifies patterns, and enables predictions or decisions based on the analysis. In a context marked by limited data availability, models grounded in statistical and machine learning techniques offer invaluable insights into the water quality domain.

In this research, insights from the high-resolution spatiotemporal data are extracted via statistical and machine learning models. Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) networks are implemented for accurate time series forecasting of water quality at fixed locations. Additionally, Deep Neural Networks (DNNs) are leveraged to elucidate water quality variations along the river's spatial profile.

2 Research Background

Water quality parameter research offers insights into water resource quality and its implications for human, aquatic life, and environmental health. By analyzing parameters such as pH, dissolved oxygen, turbidity, and nutrients, researchers can pinpoint contamination sources and assess aquatic ecosystem health [1]. However, data collection is time-consuming and costly, compounded by complex influencing factors. Predictive algorithms, like artificial neural networks and decision trees, are developed to forecast water quality parameters with limited data sets [2].

One study employs multivariate analysis (CA, PCA, DA) to classify river water, demonstrating the effectiveness of statistical techniques [3]. Another study highlights the success of EMD-LSTM models in predicting various parameters [4]. A deep learning framework employing LSTM networks forecasts water quality parameters based on historical data [5], and Bi-S-SRU models showcase higher accuracy [6]. A CNN-LSTM-SVR hybrid model outperforms traditional methods [7], while hybrid techniques and preprocessing enhance freshwater quality prediction [8].

The application of a BPNN-Kalman filter model for river water temperature, pH, and DO concentration prediction proves accurate [9]. NARNET and LSTM models predict WQI, while SVM, KNN, and Naive Bayes classify WQI data [10]. Bayesian Uncertainty Processor enhances deep learning-based ANNs [11], and a novel MVD-based framework expands DNN application even with limited data [12]. DeepST models spatiotemporal data, outperforming baselines [13]. Flow's influence on water quality parameters is studied, showing impacts on dissolved oxygen, turbidity, pH, and ORP [14–18]. These studies emphasize considering river flow in water quality assessments, acknowledging its complex relationship.

It is noteworthy that the traditional forecasting methods have lots of problems, such as low accuracy, poor generalization, and high time complexity. To solve these shortcomings, these days some machine learning based novel water quality parameters (WQP) prediction methods, like deep LSTM learnings and Deep Neural Networks are in practices [4,6,9,13,19–21]. In the case of Bagmati river, to our knowledge so far, since there exist no such sequential time series-based ML modelling studies except a scenario-based so-called Water Evolution and Planning (WEAP) [22], this research considers ML modelling as of its main scope using the following methodology.

3 Methodology

The first step involved collecting time series data related to various water quality parameters of interest. Once the raw data is obtained, it is loaded and basic exploratory analysis is performed to understand trends, distributions, relationships etc. Next, data cleaning tasks are undertaken to prepare the data for modelling. This includes handling missing values, removing outliers, and dropping irrelevant columns or features. With clean data in place, models like ARIMA and LSTM can be initialized with suitable parameters and configurations. The models are then trained on historic data and validated by testing on a holdout dataset. Training loops through iterations to minimize the loss function and update model weights and biases. Validation provides insight into how well the models generalize. Once the models are trained and validated, they can be used for one-step or multi-step ahead prediction on new data. The predictions are compared to actual values to evaluate model accuracy.

In summary, the key steps in the methodology involve data collection, exploratory analysis, cleaning, model development and training, and final prediction. This provides a structured approach to apply time series modelling for forecasting water quality indicators. The process aims to build highly accurate and robust models.

3.1 Environmental Setup

To run the models and other necessary programs, a dedicated GPU based resource Colab provided by Google Research was used. Excel was also used to save and open csv files, and process data. The libraries keras, tensorflow, statsmodels, pandas profiling, matplotlib, sklearn, pandas, numpy, and datetime were used.

3.2 Data Collection and Analysis

Water quality data was collected by Adhikari et al. at 14 fixed stations along the Bagmati River using a multi-parameter sensor system. Key parameters included pH, dissolved oxygen (DO), and temperature. On which, descriptive statistics were generated using Excel and Pandas. Data was visualized using Pandas profiling. Data was cleaned by removing outliers and parameters with high levels of missing data. The final dataset contained pH, DO and temperature.

3.3 Model Development

An ARIMA time series model was developed using a grid search to select optimal p, d, q parameters based on lowest MSE and MAE. An LSTM model was developed for time series forecasting, using encoder-decoder based architecture. A deep neural network (DNN) model was developed for spatial prediction at different locations. The model had four fully connected layers, ADAM optimizer, MAE as loss function and ReLU activation function for each layer.

3.4 ARIMA

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. That is:

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.

A nonseasonal ARIMA model is classified as an "ARIMA (p,d,q)" model, where:

- p is the number of autoregressive terms,
- d is the number of nonseasonal differences needed for stationarity, and
- q is the number of lagged forecast errors in the prediction equation.

$$\hat{y_t} = \mu + \phi_1 y_{t-1} + \phi_p y_{t-p} - \theta_1 y_{t-1} - ... - \theta_q y_{t-q} \quad (1)$$

Equation 1 General Equation for ARIMA Model. $\phi_p$represents the auto-regressive parameters; y represents the difference terms; i.e. if d = 0: $y_t = Y_t$, if d = 1 $y_t = Y_t - Y_{t-1}$ and so on θ_p represents the moving average parameters; e represents the error term at each timestamp; μ is the constant. The parameters (p,d,q) imply the order of the model. While selecting the suitable order for the model, AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are used to evaluate the quality of statistical models by balancing their complexity and goodness of fit. The AIC is a measure of the relative quality of a statistical model for a given set of data. It is defined as:

$$AIC = 2k - 2ln(L) \quad (2)$$

where, k is the number of parameters in the model and L is the maximum likelihood estimate of the likelihood function of the model. The AIC balances the trade-off between the goodness of fit and the complexity of the model, and the model with the lowest AIC is preferred.

The BIC is similar to the AIC but places a stronger penalty on model complexity. It is defined as:

$$BIC = kln(n) - 2ln(L) \quad (3)$$

where, n is the sample size, k is the number of parameters in the model and L is the maximum likelihood estimate of the likelihood function of the model. The BIC is based on the Bayesian approach, where the model with the highest posterior probability is preferred. The BIC penalizes complex models more than the AIC, and the model with the lowest BIC is preferred.

Table 1: Order of ARIMA model for various parameters

Parameter	p	q	d
DO	1	1	1
pH	3	1	0
Temperature	0	1	0
ORP	0	1	0
EC	0	1	1
TDS	0	1	1

The first step converts the dataset into a time series format by setting the time column as the index and defining the water quality parameters as feature columns. This structures the data for temporal modelling. Next, the stats-models library's Auto ARIMA capability is leveraged to automatically optimize the ARIMA parameters p, d, and q based on the data characteristics. The optimized model order provides the best starting point for the ARIMA modelling. The dataset is then partitioned into separate training and testing sets to validate the model's performance. The ARIMA model is trained on the training data, and the testing data is used to evaluate the model's accuracy. The model's predictions are compared with the actual data in the testing set to assess its performance. The mean squared error (MSE) and mean absolute error (MAE) are calculated as metrics to quantify the model's accuracy. These metrics provide insights into how well the ARIMA model predicts the water quality parameters.

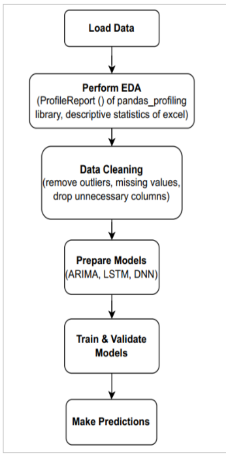

Figure 1: Methodology followed in the research.

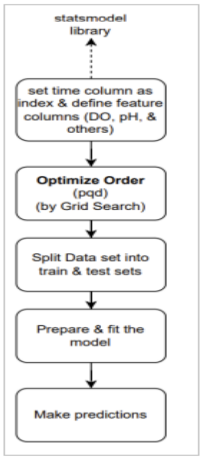

Figure 2: Steps for implementation of ARIMA Model.

3.5 LSTM

The LSTM has an input x(t) which can be the output of a CNN or the input sequence directly. h(t-1) and c(t-1) are the inputs from the previous timestep LSTM. o(t) is the output of the LSTM for this timestep. The LSTM also generates the c(t) and h(t) for the consumption of the next time step LSTM.

The first step in the LSTM involves converting the dataset into a time series format by setting the time column as the index and the water quality parameters as feature columns. This prepares the data for temporal modelling. Next, MinMax scaling is applied to normalize all features to a common 0-1 range, which aids model optimization. The dataset is then split into training and validation/testing sets in a ratio suitable for the problem, such as 80:20 or 70:30. With the data ready, the model architecture is defined by specifying the type and sequence of layers, like input, hidden and output layers, that the data will flow through. The TensorFlow library provides the tools for building and running the models. The model is then compiled by configuring key hyperparameters like optimization algorithm, loss function and metrics for training. This sets up the model for the training process. The next step is to fit the compiled model on the training data for multiple epochs, which trains the model by minimiz-

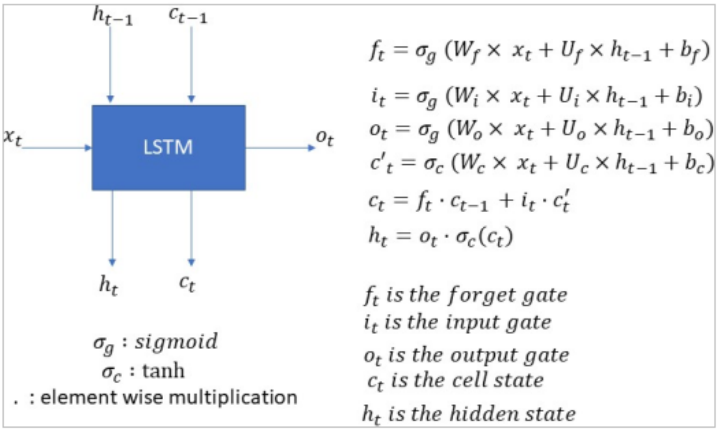

Figure 3: LSTM input outputs and the corresponding equations for a single timestep.

ing the specified loss function. Finally, the trained model can be used to generate predictions on new validation/test data, and its performance evaluated using metrics like accuracy and loss scores. In summary, the methodology involves data preprocessing, model building, training and prediction to develop a machine learning model for water quality forecasting.

The LSTM architecture consists of an input layer that accepts sequences of 10-time -steps, with each step containing 9 input features. This input layer is connected to an encoder layer comprised of 100 LSTM units. The encoder processes the input sequence and outputs an encoded sequence. The encoded sequence is then fed into a decoder layer, also containing 100 LSTM units. Additionally, a repeat vector summarizes the entire input sequence from the encoder's final output and provides this context to the decoder at each time step. Finally, a dense output layer separates the decoder outputs into distinct predictions for each water quality parameter. The model optimizes using the ADAM algorithm combined with a Huber loss function. The Huber loss provides the benefits of lower sensitivity to outliers compared to MSE and lower bias than MAE, as suggested in prior studies. This LSTM model architecture enables effective learning of complex temporal relationships and patterns in the water quality time series data.

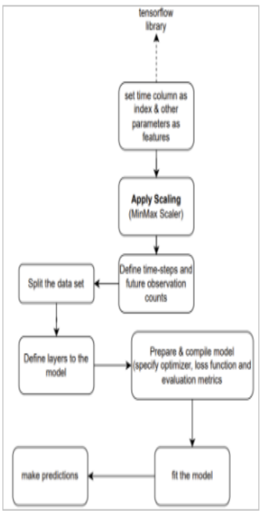

Figure 4: Steps for implementation LSTM Model.

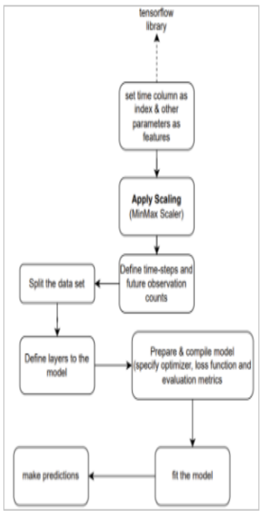

Figure 5: Steps for implementation DNN Model.

3.6 DNN Model Training

Models were trained using time series cross-validation with an 80-20 train-test split. Model parameters were tuned to optimize performance. Techniques like backpropagation and optimization algorithms like ADAM were used to minimize error and improve model accuracy. Model performance was evaluated using metrics like MSE, MAE, AIC and BIC.

4 Results and Discussion

The results of the ARIMA and LSTM models for the prediction of water quality parameters are summarized below:

Table 2: Measured and Predicted Values for Parameters at the Two Nearest Locations

Parameter	Measured Value (Blue dot)	Predicted Value (Red dot)
DO	0	0.259856
pH	7.29	7.202224
Temperature	21.64	21.34662
ORP	-254.6	-259.856

Figure 6: Plot of DO values collected with mobile sensor along the river.

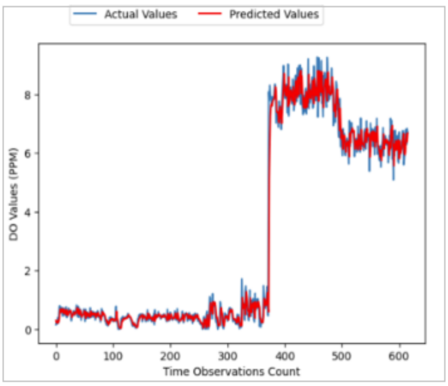

Figure 7: Prediction curve: Actual values vs Predicted values of DO (ARIMA).

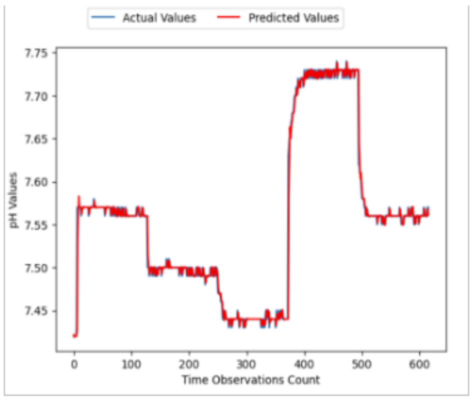

Figure 8: Prediction curve: Actual values vs Predicted values of pH (ARIMA).

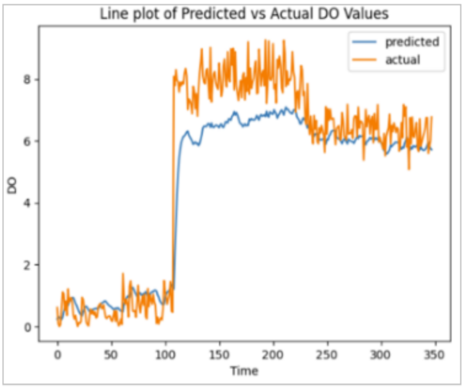

Figure 9: Prediction curve: Actual values vs Predicted values of DO (LSTM).

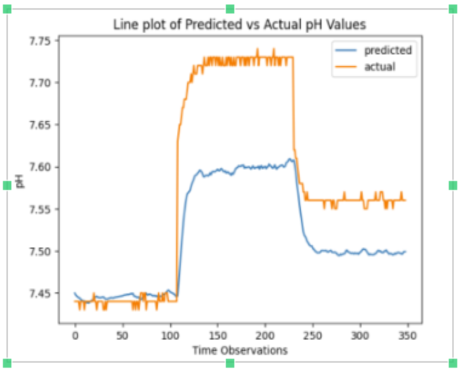

Figure 10: Prediction curve: Actual values vs Predicted values of pH (LSTM).

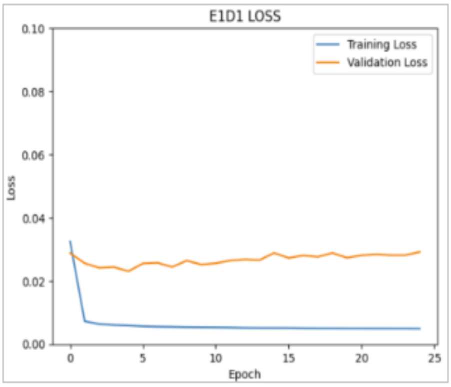

Figure 11: Epoch wise loss plot for E1D1 LSTM.

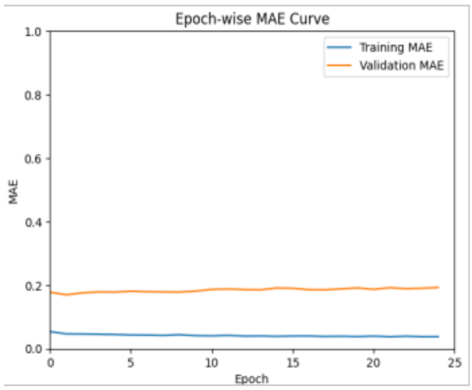

Figure 12: Epoch-wise MAE curve for E1D1 LSTM.

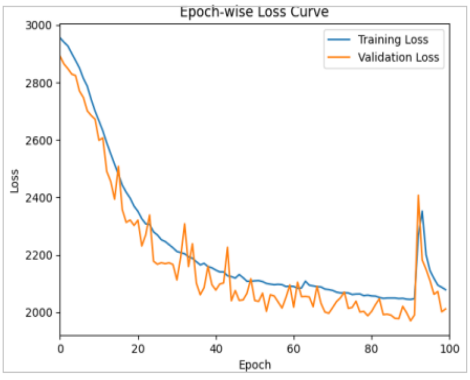

Figure 13: Epoch-wise Loss Curve for DNN Model at 80 : 20 validation split.

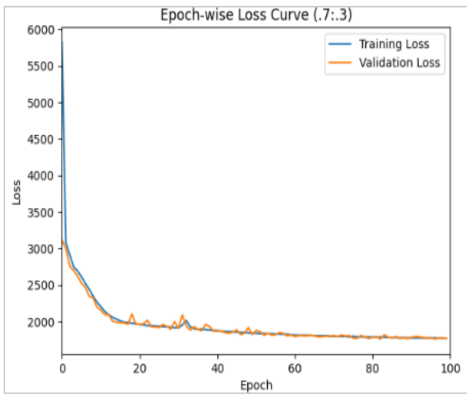

Figure 14: Epoch-wise Loss Curve for DNN Model at 70 : 30 validation split.

4.1 ARIMA Model

The ARIMA time series modeling approach was able to effectively capture the temporal dynamics in the water quality parameters. For dissolved oxygen (DO), the model achieved an RMSE of 0.519, MAE of 0.2901 and R2 of 0.9760 on the test set, indicating a good model fit with minimal errors. The residuals plot showed that most residuals were clustered close to zero, signifying that the actual values were close to the predicted values. The model was able to forecast DO levels reasonably accurately up to 5-time-steps ahead, with errors increasing slightly for longer forecast horizons. Overall, the ARIMA

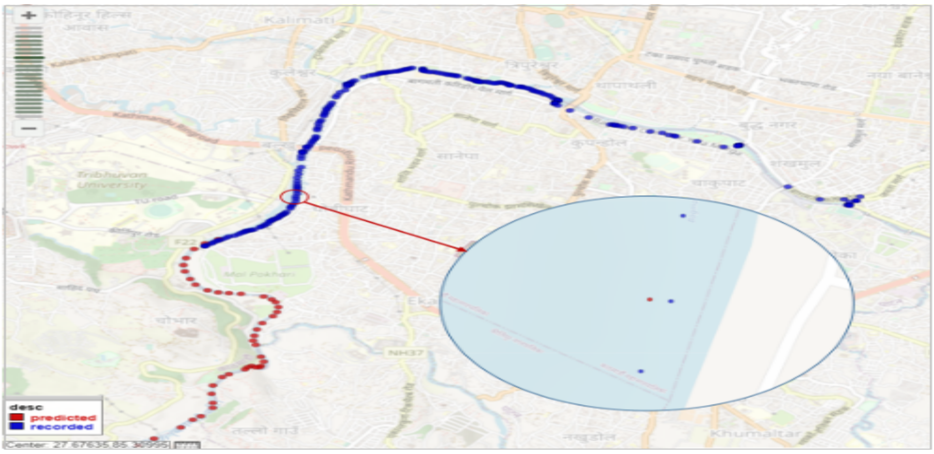

Figure 15: GPS points with values measured (blue) and predicted (red) by DNN.

modelling was successful in modelling the temporal variations in DO and other water quality indicators.

4.2 LSTM Model

The LSTM neural network model demonstrated promising performance for short-term time series forecasting across the different water quality parameters. For temperature predictions, the model achieved the lowest error with MAE scores ranging from 0.2649 to 0.2701 across 5-time-units. The errors were higher but still fairly low for other parameters like pH (MAE 0.0527 to 0.0586) and dissolved oxygen (MAE 1.10 to 1.49). The model was able to learn complex time-dependent patterns in the data. The epoch-wise learning curves showed that the model errors stabilized within 30-40 epochs of training. Overall, the LSTM model provided robust forecasts for the next 5-time-steps based on previous lags and long-term temporal contexts.

4.3 DNN Model

The deep neural network model for spatial prediction yielded optimal results with a 80:20 train-test split, achieving the lowest MAE of around 0.15 after 30 epochs. The model training curves showed the validation loss decreasing and levelling off after 30-40 epochs across different data splits. This indicates that the model was able to learn the underlying spatial relationships between the location coordinates and water quality parameters. The multilayer architecture with increasing number of neurons in the hidden layers likely enabled the model to learn complex nonlinear feature representations. The model demonstrates potential for accurate prediction of water quality indicators like pH and DO levels based on the geographic coordinates.

5 Conclusion

This study demonstrated the feasibility of using machine learning approaches like ARIMA, LSTM, and DNN models for predicting water quality parameters in the Bagmati River in Nepal. The fine-scale, real-time observation data collected provides valuable insights into the spatio-temporal dynamics of key water quality indicators like pH, DO, temperature, and conductivity. The ARIMA model was successfully implemented to capture the temporal patterns in parameters like DO and pH. The LSTM model also showed promising results for short-term time series forecasting of multiple water quality variables. The location-based DNN model achieved reasonable performance in predicting water quality parameters based on geographic coordinates.

However, the limitations of the current dataset underscore the need for expanded data collection across wider spatial and temporal scales, as well as inclusion of more water quality indicators. Addressing these limitations through continued research will further enhance our understanding of the intricate relationships between various natural and anthropogenic factors influencing river water quality. Incorporating turbulence data along with spatio-temporal coordinates could have enabled examining the fluctuations in water quality parameters as water flows through different regions over time. Availability of extensive datasets capturing turbulence, timestamps, and locations would have facilitated developing enhanced forecasting models and uncovering significant dynamic patterns and trends. Specifically, analysis of how turbulence impacts various quality measures could lead to more sophisticated models and insights. However, investigating the role of turbulence poses certain challenges such as complex measurements and simulations. Nevertheless, accounting for turbulence remains an important consideration for future work to advance understanding of aquatic systems and devise optimal water resource management strategies. While turbulence incorporation was beyond the scope of this study, addressing the associated difficulties and limitations in follow-up research could significantly improve water quality assessment capabilities.

Nonetheless, this study represents an important step towards leveraging advanced ML techniques for developing accurate, real-time water quality monitoring systems. The model frameworks presented can inform future efforts to build intelligent decision support systems for water resource management, pollution control and remediation in the Bagmati and similar river ecosystems.

Acknowledgement

This research is supported by University Grants Commission, Nepal through Collaborative Research Grant (CRG-79/80-ST-01).

References

[1] M. P. Adhikari, N. B. Rawal, and N. B. Adhikari. Real-time fine-scale measurement of water quality parameters along the bagmati river in the kathmandu valley. *Nature Environment and Pollution Technology*, 20(3):1047–1057, Sep 2021.

[2] A. Najah Ahmed et al. Machine learning methods for better water quality prediction. *J Hydrol (Amst)*, 578:124084, Nov 2019.

[3] S. C. Azhar, A. Z. Aris, M. K. Yusoff, M. F. Ramli, and H. Juahir. Classification of river water quality using multivariate analysis. *Procedia Environ Sci*, 30:79–84, 2015.

[4] Y. Zhang et al. Accurate prediction of water quality in urban drainage network with integrated emd-lstm model. *J Clean Prod*, 354:131724, Jun 2022.

[5] R. Sharma et al. Analysis of water pollution using different physicochemical parameters: A study of yamuna river. *Front Environ Sci*, 8, Dec 2020.

[6] J. Liu et al. Accurate prediction scheme of water quality in smart mariculture with deep bi-s-sru learning network. *IEEE Access*, 8:24784–24798, 2020.

[7] U. Ahmed et al. Efficient water quality prediction using supervised machine learning. *Water (Basel)*, 11(11):2210, Oct 2019.

[8] Z. S. Khudhair, S. L. Zubaidi, S. Ortega-Martorell, N. Al-Ansari, S. Ethaib, and K. Hashim. A review of hybrid soft computing and data pre-processing techniques to forecast freshwater quality's parameters: Current trends and future directions. *Environments - MDPI*, 9(7):85, Jul 2022.

[9] Yanfei Zhao, Zhihong Zou, and Shenglong Wang. A back propagation neural network model based on kalman filter for water quality prediction. In *2015 11th International Conference on Natural Computation (ICNC)*, pages 149–153, 2015.

[10] T. H. H. Aldhyani, M. Al-Yaari, H. Alkahtani, and M. Maashi. Water quality prediction using artificial intelligence algorithms. *Appl Bionics Biomech*, 2020:1–12, Dec 2020.

[11] Y. Zhou et al. Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques. *J Hydrol (Amst)*, 589:125164, Oct 2020.

[12] A. El Bilali, H. Lamane, A. Taleb, and A. Nafii. A framework based on multivariate distribution-based virtual sample generation and dnn for predicting water quality with small data. *J Clean Prod*, 368:133227, Sep 2022.

[13] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi. Dnn-based prediction model for spatio-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–4, 2016.

[14] S. Liu et al. Anthropogenic disturbances on distribution and sources of pharmaceuticals and personal care products throughout the jinsha river basin, china. *Environ Res*, 198:110449, Jul 2021.

[15] T. Deng, K.-W. Chau, and H.-F. Duan. Machine learning based marine water quality prediction for coastal hydro-environment management. *J Environ Manage*, 284:112051, Apr 2021.

[16] T. Xie, G. Zhang, J. Hou, J. Xie, M. Lv, and F. Liu. Hybrid forecasting model for non-stationary daily runoff series: A case study in the han river basin, china. *J Hydrol (Amst)*, 577:123915, Oct 2019.

[17] K. Y. Jung et al. Evaluation of water quality for the nakdong river watershed using multivariate analysis. *Environ Technol Innov*, 5:67–82, Apr 2016.

[18] V. Senapathi et al. Assessment of river water quality via environmentric multivariate statistical tools and water quality index: A case study of nakdong river basin, korea. 2014.

[19] H. Zhuhua et al. A water quality prediction method based on the deep lstm network considering correlation in smart mariculture. *Sensors (Basel)*, 19(6):1420, 2019.

[20] L. Ping et al. Analysis and prediction of water quality using lstm deep neural networks in iot environment. *Sustainability*, 11(7), 2019.

[21] J. Yigi et al. Accurate prediction of water quality in urban drainage network with integrated emd-lstm model. *Journal of Cleaner Production*, 354:131724, 2022.

[22] B. K. Mishra et al. Assessment of bagmati river pollution in kathmandu valley: Scenario-based modeling and analysis for sustainable urban development. *Sustainability of Water Quality and Ecology*, 9–10:67–77, 2017.