E-ISSN : 2827-8550
P- ISSN : 1978-5569

IMPLEMENTATION OF THE LSTM DEEP LEARNING MODEL FOR WATER QUALITY PARAMETER PREDICTION

Yeza Febriani[1)]*, Dwi Astuti[2)], Yusuf Wahyu Setiya Putra[3)], Riska Dwi Handayani[4)], Fatkhurrohman[5)], Kapti[6)]

[1) 2)] "Informatics Engineering" STMIK BINA PATRIA Magelang
[3) 4)5)] "Sistem Information" STMIK BINA PATRIA Magelang
[6)] "Health Information Management" AKMT Temanggung
Email:yezafebriani@stmikbinapatria.ac.id[1)], dwi.astuti@stmikbinapatria.ac.id[2)], yusuf@stmikbinapatria.ac.id[3)], riska@stmikbinapatria.ac.id[4)], fathur@stmikbinapatria.ac.id[5)], tensmart18@gmail.com[6)]

Abstract

Efficient water quality prediction is crucial amid the threat of increasing pollution. Conventional methods have limitations in terms of cost, time, and coverage, requiring an innovative approach based on artificial intelligence. This study aims to classify drinking water suitability using Long Short-Term Memory (LSTM) architecture, which is known to be effective for sequential data. The research method includes data collection from Kaggle, data pre-processing such as normalization and missing value handling, and data division into 80% for training and 20% for testing. Specifically, the model utilizes key physicochemical parameters including pH, temperature, turbidity, salinity, dissolved oxygen, specific conductance, and chlorophyll.The proposed LSTM model was evaluated using accuracy, precision, recall, and F1-Score metrics. The results show that the LSTM model is capable of achieving an overall accuracy of 95.47%, with a precision of 0.9511 and a recall of 0.9547. Although there are some False Positive classification errors (predicting water as unfit when it is actually fit), the overall performance of the model is excellent and reliable for this classification task. The conclusion of this study is that the LSTM model can be an effective and accurate solution for predicting water quality, supporting early detection and real-time pollution control efforts

Keywords : *prediction, water quality, classification, deep learning, LSTM*

1. Introduction

Water is a vital resource that is essential for human survival and supports various important sectors such as industry and agriculture. However, the pace of human activity and rapid industrial development has significantly increased the threat of pollution to water resources. This has led to a decline in water quality that has serious impacts on human health and ecosystem stability (Lin et al., 2022; Misman et al., 2023; Y. Zhang et al., 2024). Various sources of water pollution, ranging from industrial waste, domestic waste, agricultural activities, to massive urbanization, all contribute to the contamination of both surface water and groundwater (Misman et al., 2023; Syeed et al., 2023; Y. Zhang et al., 2024).

In addressing the challenge of water pollution, conventional methods for testing water quality, such as spectrophotometry, chromatography, and physical-chemical laboratory tests, still have significant limitations. These methods generally require a long time, high operational costs, and require experts and special equipment. In addition, manual approaches are often inefficient for real-time water quality monitoring and are less capable of detecting large amounts of pollutants or complex mixtures. These conditions pose a major obstacle to early detection and effective water pollution control (Yaroshenko et al., 2020) .

Various innovative approaches have now been developed to overcome the limitations of conventional methods. The application of water quality indices (WQI) and pollution indices (PI) is becoming increasingly common to classify and profile water quality in a more integrated and

E-ISSN : 2827-8550
P- ISSN : 1978-5569

efficient manner (Al-Adhaileh & Alsaade, 2021). Furthermore, developments in chemical sensor technology, fluorescence spectroscopy, and microfluidic platforms integrated with artificial intelligence (AI) and machine learning have enabled more accurate, automated, and real-time water quality monitoring . AI-based predictive models also demonstrate high capabilities in processing large volumes of data, accelerating the classification process, and improving the efficiency of water pollution detection (G. Xu et al., 2022).

Various algorithms have been explored, with some showing very promising performance in classifying water quality status based on various physical-chemical parameters. Algorithms such as Support Vector Machine (SVM), Naïve Bayes, and Multi-Layer Perceptron (MLP). MLP, as a type of Artificial Neural Network (ANN), has demonstrated remarkable capabilities in handling complex nonlinear relationships, often providing the highest accuracy compared to other algorithms such as J48 and Naïve Bayes, with an accuracy range of 0.85 to 0.96 (Ahmed et al., 2019; Chen et al., 2023).

With advances in deep learning, these models have begun to be applied to water quality prediction and classification, especially for sequential or time series data. In this context, Long Short-Term Memory (LSTM), a type of recurrent neural network (RNN), has emerged as a highly relevant architecture due to its ability to effectively handle sequential data and model complex temporal dynamics. Recent studies show that LSTM can achieve very high performance, with R2 values of up to 0.9999, even outperforming other models such as XGBoost in predicting water quality indices (G. Xu et al., 2022). However, although LSTM has proven to be superior in predicting time series-based water quality parameters and has shown great potential in environmental monitoring, its specific application for classifying drinking water suitability is still limited. Based on the above background, this study aims to predict water quality classification using the LSTM model. The Kaggle dataset was specifically chosen for this study because it provides a comprehensive and globally relevant collection of critical water parameters, allowing the model to be rigorously trained and validated across diverse environmental conditions for practical application.

2. Research Methods

The research stages were systematically designed to predict water quality using a deep learning approach, as shown in Figure 1. Several main stages, ranging from dataset description, data pre-processing, model architecture design, to model performance evaluation.

E-ISSN : 2827-8550
P- ISSN : 1978-5569

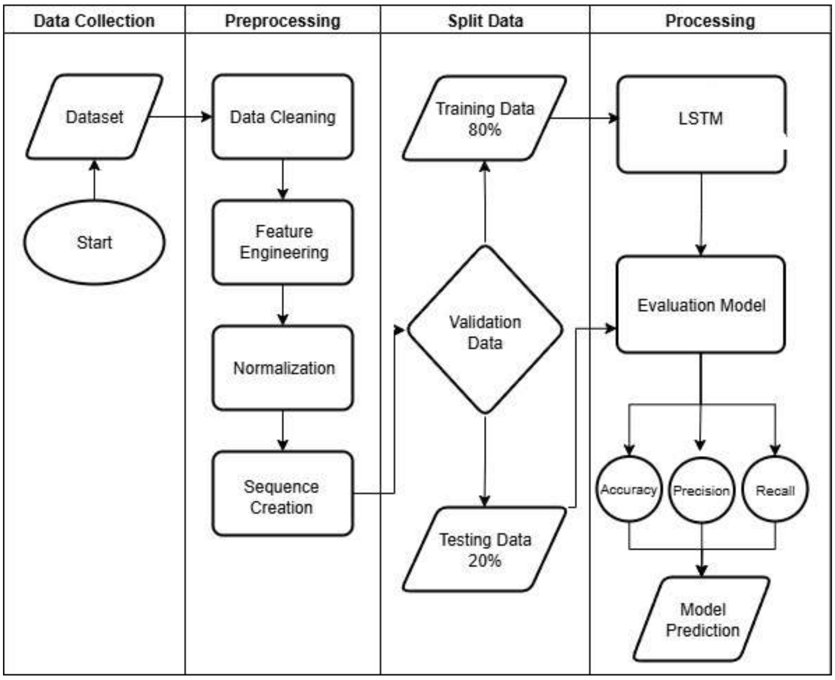

Figure 1. *Research Flowchart*

Dataset Description

The dataset used in this study is a water quality dataset available on the Kaggle platform, which can be accessed at the following link : https://www.kaggle.com/code/briannuevo/water-quality-analysis-and-time-series.

Data Preprocessing

Data preprocessing is a crucial stage in the machine learning pipeline that directly affects the accuracy and reliability of the model. The steps to be taken include handling missing values, data normalization, and feature selection.

1. Handling Missing Values: Missing values in a dataset can reduce the quality of the analysis and cause bias in the model. Strategies used to fill in empty values include imputation using the mean or median of the related feature.
2. Data Normalization: Normalization is the process of scaling feature values into a specific range, often between 0 and 1.
3. Feature Selection: Feature selection aims to reduce data dimensions by identifying and selecting a subset features that are most influential or relevant to the target variable.
4. Splitting data: The processed dataset is divided chronologically into two subsets: 80% for training and 20% for testing. Chronological division prevents data leakage and ensures that the model is evaluated based on its ability to predict future data that it has never seen before.

Long Short-Term Memory (LSTM) Model Architecture

The LSTM architecture proposed in this study will be designed effectively to process water quality data and classify its potability. The architecture design includes:

1. Number of Layers: The LSTM model can be implemented in one or more layers (stacked LSTM). Multi-layer (stacked) architectures are often used to increase the model's capacity to capture complex patterns and feature hierarchies from sequential data (Lin et al., 2022; Lindemann et al., 2021; Q. Zhang et al., 2017). In this study, two LSTM layers were implemented [the number of layers, for example, one or two LSTM layers] will be used to balance the complexity of the model with the expected performance.

E-ISSN : 2827-8550
P- ISSN : 1978-5569

2. Number of LSTM Units per Layer: The number of units (neurons) in each LSTM layer is an important parameter that affects the learning capacity of the model. The number of units varies depending on the complexity of the data and the purpose of the model. General studies often use 32, 64, 128, or more units per layer (J. Xu et al., 2024). In this model, the number of units will be determined, for example, 64 units will be implemented in each LSTM layer.
3. Activation Function: Activation functions play a crucial role in introducing non-linearity into the network. The main activation function used in LSTM cells is sigmoid (σ) for the input, forget, and output gates, which produces values between 0 and 1 to control the flow of information. Meanwhile, the tanh (hyperbolic tangent) function is used for memory cell updates and hidden state activation coming out of the output gate, which produces values between -1 and 1 (Lindemann et al., 2021; Malashin et al., 2024; Q. Zhang et al., 2017).
4. Output Layer: After the LSTM layer processes the input sequence, one or more dense (fully connected) layers are usually added to produce the final output. For binary classification tasks such as determining water potability (drinkable/undrinkable), a single dense layer with one neuron and a sigmoid activation function will be used. The sigmoid function will produce probabilities between 0 and 1 that can be interpreted as the likelihood of water being drinkable (J. Xu et al., 2024; Q. Zhang et al., 2017).

Model Evaluation

1. Accuracy measurement is performed as a basic metric to evaluate the percentage of correct predictions from the total predictions made by the system. The accuracy calculation implementation uses accuracy_score from scikit-learn to ensure consistency in calculations. This metric is applied to binary classification outputs (eligible/ineligible) with a threshold probability of 0.5 for conversion from sigmoid output to binary classification.
2. Precision is calculated using the formula TP / (TP + FP) to measure the accuracy of the model's positive predictions. This metric is very important in the context of water quality monitoring, where false positives (predicting that water is unfit when it is actually fit) can cause unnecessary alarms or treatment.
3. Recall measurement uses the formula TP / (TP + FN) to evaluate the model's ability to detect all existing positive instances. The implementation of recall_score with handling for zero division cases ensures evaluation robustness. Recall has high significance in water quality applications where false negatives (predicting water is safe when it is not) can impact consumer health. Recall evaluation is performed with monitoring to ensure an optimal detection rate.
4. F1-Score is calculated as the harmonic mean of precision and recall using the formula 2 × (precision × recall) / (precision + recall). The implementation of f1_score provides a balanced evaluation that considers both false positives and false negatives.

3. Results and Discussionn

3.1 *Research Result*

Correlation matrix, also known as heatmap, to analyze the relationship between various water quality parameters. Each box in the matrix visualizes the correlation coefficient between two variables, using color gradation to indicate the strength and direction of the relationship.

E-ISSN : 2827-8550
P- ISSN : 1978-5569

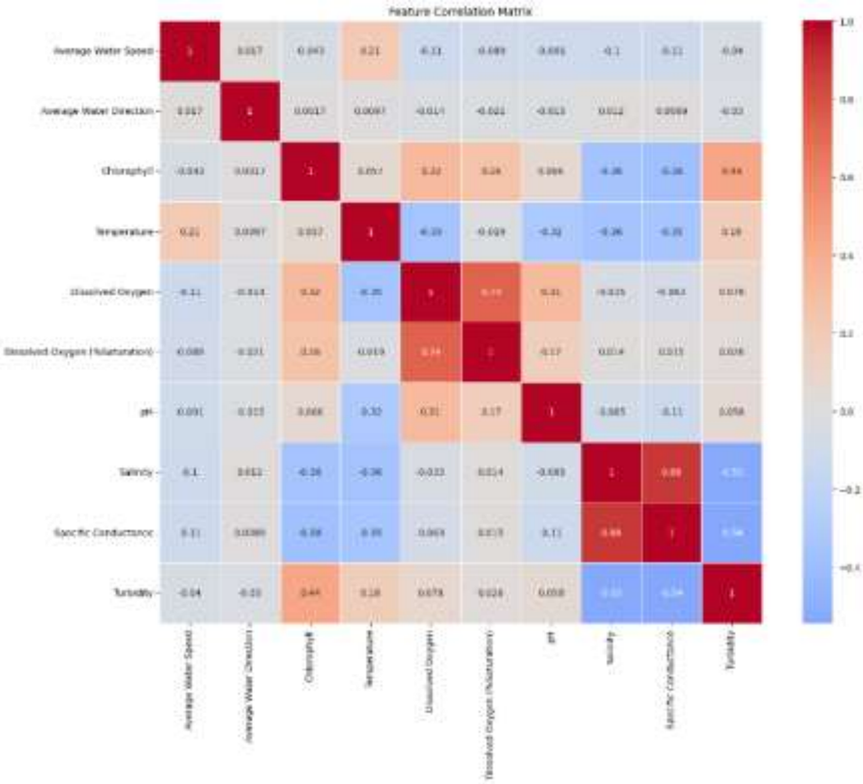

Figure 2. *Water Quality Parameter Correlation Matrix*

Based on Figure 2, it is clear that salinity has a very strong positive correlation with specific conductance, which makes sense because salt content is the main determinant of water conductivity. On the other hand, there is a negative correlation between Temperature and Dissolved Oxygen, which is in line with the law of physics that the solubility of gas in water decreases as the temperature rises. A significant positive correlation can be seen between Dissolved Oxygen and its saturation percentage (0.74), as well as a moderate relationship between Turbidity and Chlorophyll (0.44). Conversely, the matrix shows a moderate negative correlation, particularly between Temperature and Dissolved Oxygen (-0.35), consistent with the principle that warmer water holds less dissolved gas. Additionally, Turbidity negatively correlates well with both Salinity (-0.53) and Specific Conductivity (-0.54), indicating that events such as freshwater inflow can increase turbidity while simultaneously decreasing salinity.

Correlation of Water Quality Parameters: Salinity, Conductivity, Temperature, Dissolved Oxygen, Turbidity, and Chlorophyll

Correlation analysis between water quality parameters such as salinity, specific conductivity, temperature, dissolved oxygen, turbidity, and chlorophyll is very important for understanding the dynamics of aquatic ecosystems. Strong positive correlations between salinity and specific conductivity, as well as negative correlations between temperature and dissolved oxygen, are supported by scientific literature.

Correlation between Salinity and Specific Conductivity

Salinity and specific conductivity have a very strong positive relationship because the main salt (ion) content determines water conductivity. Studies show that this relationship is generally linear in seawater and brackish water, although at very high salinity, the relationship becomes non-linear and is influenced by ionic composition (Rebello et al., 2020; Rusydi, 2018).

E-ISSN : 2827-8550
P- ISSN : 1978-5569

Conductivity is also often used as a proxy for salinity in water quality monitoring (Rebello et al., 2020).

Correlation between Temperature and Dissolved Oxygen

There is a consistent negative correlation between temperature and dissolved oxygen; the higher the temperature, the lower the solubility of oxygen in water. This is in accordance with the laws of physics and is supported by various studies predicting dissolved oxygen, in which temperature is the most influential variable on dissolved oxygen levels (Kisi et al., 2020; Maroufpoor et al., 2022a; J. Xu et al., 2024).

Other Correlations: Dissolved Oxygen, Turbidity, and Chlorophyll

1. Dissolved oxygen and its saturation percentage show a significant positive correlation, as both are influenced by similar physical and biological processes (Maroufpoor et al., 2022b; J. Xu et al., 2024).
2. Turbidity and chlorophyll often correlate positively, as increased chlorophyll can increase turbidity (Giles et al., 2022; Gogoi et al., 2024; Kayalık & Çorumluoğlu, 2022; Nong et al., 2024).
3. Turbidity correlates negatively with salinity and specific conductivity, especially when freshwater inflow increases turbidity and decreases salinity (Gholamalifard et al., 2024; Giles et al., 2022; Jabin et al., 2024).

Figure 3 shows that, overall, of the total samples tested, the vast majority (more than 20,000 samples) were of acceptable quality, while only a small portion (less than 2,500 samples) were considered unacceptable.

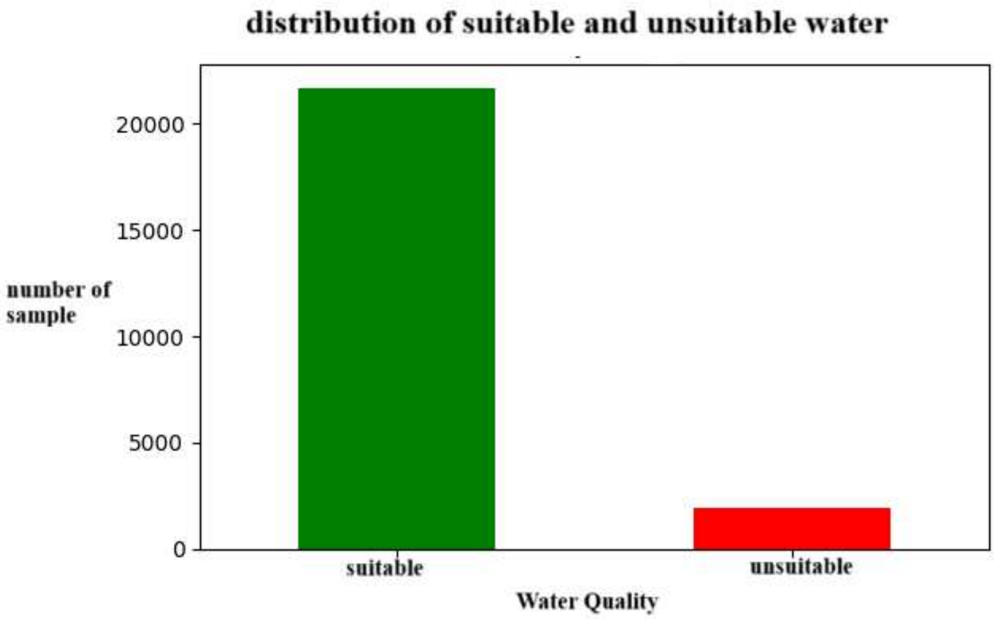

Figure 3. *Water quality distribution*

Figure 4 shows the evaluation matrix of the LSTM deep learning model used to predict water quality classification as suitable or unsuitable.

E-ISSN : 2827-8550
P- ISSN : 1978-5569

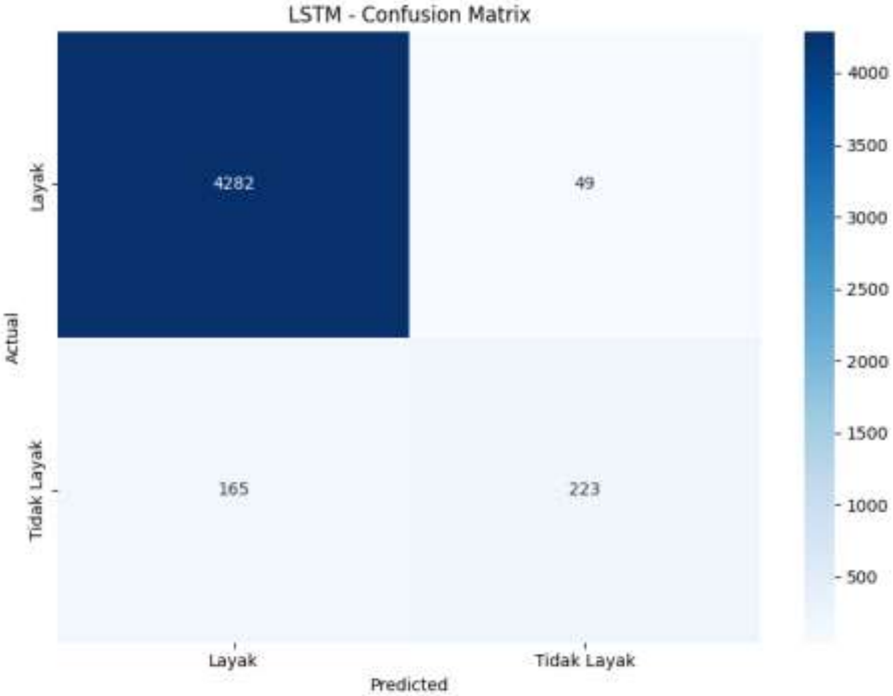

Figure 4. *LSTM Confusion Matrix*

A total of 4,331 water samples were classified as suitable, and the model correctly predicted 4,282 (True Positive), while 49 were incorrectly predicted as unsuitable (False Negative). On the other hand, of the total number of unfit water samples, the model correctly predicted 223 (True Negative), but incorrectly predicted 165 samples as fit (False Positive). Due to the significant data imbalance (over 20,000 suitable samples versus fewer than 2,500 unsuitable samples), a class-specific evaluation reveals a performance discrepancy. While the model is highly robust in detecting suitable water (98.8% class accuracy), its ability to detect the minority "unfit" class requires further optimization. This prediction error (False Positive) indicates that the model still has room for improvement, particularly in identifying unfit water conditions so that they are not classified as fit.

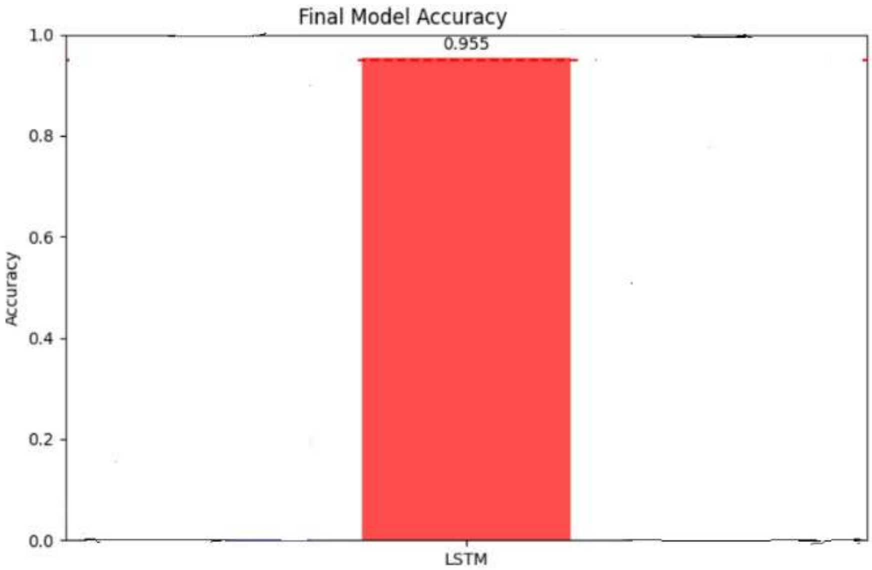

Figure 5. *Accuracy Model*

E-ISSN : 2827-8550
P- ISSN : 1978-5569

Figure 5. This graph shows that the LSTM model used has excellent performance with an accuracy rate of 95.5%, which means that the model is able to predict or classify data correctly in 95.5% of all samples tested. This high accuracy indicates that the model is very reliable for the classification task given.

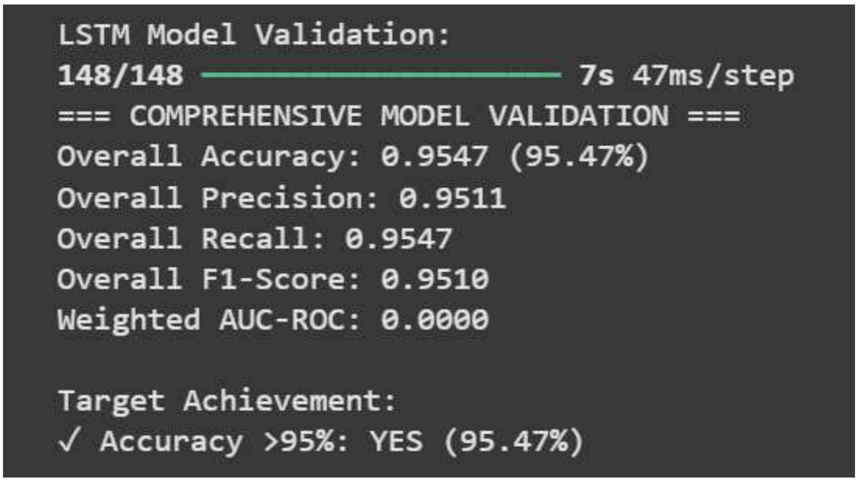

Figure 6. *LSTM model validation*

The comprehensive performance of the deep learning model can be seen in Figure 6, which shows the LSTM model validation. Overall Accuracy shows that the overall accuracy of the model is 0.9547 or 95.47%. This indicates that the model correctly predicts almost 95.5% of all validation data (Acosta et al., 2024). Overall Precision is a precision value of 0.9511. Precision measures how accurately the model predicts positive classes (for example, 'Suitable') from all positive predictions it makes. Overall Recall is a recall value of 0.9547. Recall measures how well the model finds all positive cases that actually exist in the data (Y. Hu et al., 2022; Z. Hu et al., 2024; Xia et al., 2020). The Overall F1-Score is 0.9510. The F1-Score is the harmonic mean of precision and recall, providing a balanced view of the model's performance (Z. Hu et al., 2024; Jailani et al., 2023).

3.2 Discussion

Logical Analysis of the Findings

The deep learning approach using the Long Short-Term Memory (LSTM) model proposed in this study demonstrated excellent performance in predicting and classifying drinking water suitability. Based on the evaluation, the model achieved an overall accuracy of 95.47%, a precision of 0.9511, and a recall of 0.9547. These results logically indicate that the LSTM architecture is highly capable of modeling the complex temporal dynamics and non-linear relationships inherent in sequential water quality data. Although the model showed exceptional performance, it still produced some False Positive classification errors, wherein the model incorrectly predicted unfit water as fit. his specific type of error presents a critical real-world challenge, as classifying contaminated water as safe carries direct and severe health risks for consumers. Future technical strategies to mitigate this error rate and handle the dataset's imbalance should include applying oversampling techniques such as SMOTE (Synthetic Minority Over-sampling Technique), utilizing class-weighted loss functions to heavily penalize False Positives during training, or adjusting the probability decision threshold to prioritize public health safety. In the context of water quality monitoring, paying attention to false positives is crucial to

E-ISSN : 2827-8550
P- ISSN : 1978-5569

prevent unnecessary alarms or treatment interventions. However, the overall results confirm that AI-based predictive models can effectively process large volumes of data, accelerate the classification process, and improve the efficiency of real-time pollution control efforts.

Agreement with Previous Studies and Existing Theories

The findings of this study are in strong agreement with previous research and existing theories regarding the application of Artificial Neural Networks (ANN) in environmental monitoring. Previous studies have reported that Multi-Layer Perceptron (MLP) models can achieve an accuracy range of 0.85 to 0.96 for water quality classification. The LSTM model's accuracy of 95.47% in this study sits perfectly at the upper bound of this benchmark, validating the superiority of deep learning architectures for this specific task.

Furthermore, these results align with the broader consensus in the scientific community that LSTM networks offer enhanced predictive power for water quality parameters due to their unique ability to retain long-term dependencies and automatically learn complex patterns (Pyo et al., 2023). Previous applications of LSTM in this domain have consistently shown high accuracy and stability, establishing it as a more reliable method compared to traditional machine learning algorithms when handling complex environmental data. By achieving these high metrics, this study reinforces the theory that overcoming the limitations of conventional laboratory testing methods which generally require a long time and high operational costs is highly achievable through the integration of artificial intelligence and machine learning.

4. Conclusion

The implementation of the Long Short-Term Memory (LSTM) deep learning model in this study demonstrates exceptional efficacy in predicting and classifying the suitability of drinking water. The proposed architecture achieved a high overall accuracy of 95.47%, complemented by a precision of 0.9511 and a recall of 0.9547. These results logically confirm that the LSTM model is highly capable of capturing the complex temporal dynamics and non-linear relationships inherent in sequential water quality data. Furthermore, the study validates established scientific correlations, such as the strong positive relationship between salinity and specific conductance, as well as the consistent negative correlation between temperature and dissolved oxygen. Ultimately, this AI-based approach provides a reliable and efficient alternative to conventional laboratory testing methods, which are often limited by high operational costs and slow processing times.

For future development, it is highly recommended to integrate this predictive LSTM model into IoT (Internet of Things) sensor devices. Such integration would enable continuous, real-time water quality monitoring directly in the field, providing immediate automated alerts to authorities and further reducing reliance on delayed manual laboratory testing.

Bibliography

Acosta, J. C., Ortiz, J. G. A., P., M. D., Galan, J., Soriano, V. R., & Venegas-Anaya, M. (2024). Variación espacial en la estación seca de las características fisicoquímicas del agua marino-costera en Bahia Damas, Parque Nacional Coiba, Panamá. *2024 9th International Engineering, Sciences and Technology Conference (IESTEC)*, 264–270. https://doi.org/10.1109/IESTEC62784.2024.10820260

Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., & García-Nieto, J. (2019). Efficient Water Quality Prediction Using Supervised Machine Learning. *Water*, 11(11), 2210. https://doi.org/10.3390/w11112210

Al-Adhaileh, M. H., & Alsaade, F. W. (2021). Modelling and Prediction of Water Quality by Using Artificial Intelligence. *Sustainability*, *13*, 4259. https://doi.org/10.3390/SU13084259

E-ISSN : 2827-8550
P- ISSN : 1978-5569

Chen, X., Aljrees, T., Umer, M., Saidani, O., Almuqren, L., Mzoughi, O., Ishaq, A., & Ashraf, I. (2023). Cervical cancer detection using K nearest neighbor imputer and stacked ensemble learning model. *Digital Health*, 9. https://doi.org/10.1177/20552076231203802

Gholamalifard, M., Ahmadi, B., Naghdi, M., & Kostianoy, A. (2024). Monitoring of Optically Active Constituents (OACs) in the Southern Caspian Sea: Concurrent Match-up Between In-Situ and Satellite Observations. *Ecologica Montenegrina*. https://doi.org/10.37828/em.2024.76.10

Giles, A., Correa, R., Santos, I., & Kelaher, B. (2022). Using multispectral drones to predict water quality in a subtropical estuary. *Environmental Technology*, 45, 1300–1312. https://doi.org/10.1080/09593330.2022.2143284

Gogoi, P., Das, S., Jana, C., Das, B., Saha, A., Ramteke, K., Jaiswar, A., Samanta, S., & Roshith, C. (2024). Assessing the trophic status of a tropical microtidal estuary applying TRIX and Random Forest—A combined approach. *Marine Pollution Bulletin*, *200*, 116126. https://doi.org/10.1016/j.marpolbul.2024.116126

Hu, Y., Wei, R., Yang, Y., Li, X., Huang, Z., Liu, Y., He, C., & Lu, H. (2022). Performance Degradation Prediction Using LSTM with Optimized Parameters. *Sensors*, 22. https://doi.org/10.3390/s22062407

Hu, Z., Huang, J., Wang, J., Wang, Z., & Qiao, Y. (2024). Hydrothermal carbonization of sewage sludge for hydrochar valorization: Role of feedwater pH on sulfur removal and transformation. *Fuel*. https://doi.org/10.1016/j.fuel.2024.132813

Jabin, S., Kapoor, J., Chadha, A., Gupta, A., & Jadoun, S. (2024). Assessment of poly(diallyl dimethyl ammonium chloride) and lime for surface water treatment (pond, river, and canal water): Seasonal variations and correlation analyses. *Environmental Monitoring and Assessment*, 196(10), 874. https://doi.org/10.1007/s10661-024-13004-3

Jailani, N. L. M., Dhanasegaran, J. K., Alkawsi, G., Alkahtani, A., Phing, C. C., Baashar, Y., Capretz, L. F., Al-Shetwi, A., & Tiong, S. (2023). Investigating the Power of LSTM-Based Models in Solar Energy Forecasting. *Processes*. https://doi.org/10.3390/pr11051382

Kayalık, M., & Çorumluoğlu, Ö. (2022). SST Correlation Between Chlorophyll and Turbidity by Landsat MS Image Analysis for the Coast of Izmir Province. *International Journal of Environment and Geoinformatics*. https://doi.org/10.30897/ijegeo.1065482

Kisi, O., Alizamir, M., & Gorgij, A. D. (2020). Dissolved oxygen prediction using a new ensemble method. *Environmental Science and Pollution Research*, 27, 9589–9603. https://doi.org/10.1007/s11356-019-07574-w

Lin, L., Yang, H., & Xu, X. (2022). Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. *Frontiers in Environmental Science*. https://doi.org/10.3389/fenvs.2022.880246

Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP*. https://doi.org/10.1016/J.PROCIR.2021.03.088

Malashin, I., Tynchenko, V., Gantimurov, A., Nelyub, V., & Borodulin, A. (2024). Applications of Long Short-Term Memory (LSTM) Networks in Polymeric Sciences: A Review. *Polymers*, 16. https://doi.org/10.3390/polym16182607

Maroufpoor, S., Sammen, S., Alansari, N., Abba, S., Malik, A., Shahid, S., Mokhtar, A., & Maroufpoor, E. (2022). A novel hybridized neuro-fuzzy model with an optimal input combination for dissolved oxygen estimation. *Frontiers in Environmental Science*. https://doi.org/10.3389/fenvs.2022.929707

Misman, N. A., Sharif, M. F., Chowdhury, A. J. K., & Azizan, N. (2023). Water pollution and the assessment of water quality parameters: A review. *Desalination and Water Treatment*. https://doi.org/10.5004/dwt.2023.29433

E-ISSN : 2827-8550
P- ISSN : 1978-5569

Maroufpoor, S., Sammen, S., Alansari, N., Abba, S., Malik, A., Shahid, S., Mokhtar, A., & Maroufpoor, E. (2022b). *A novel hybridized neuro-fuzzy model with an optimal input combination for dissolved oxygen estimation*. *10*. https://doi.org/10.3389/fenvs.2022.929707

Misman, N. A., Sharif, M. F., Chowdhury, A. J. K., & Azizan, N. (2023). Water pollution and the assessment of water quality parameters: A review. *DESALINATION AND WATER TREATMENT*. https://doi.org/10.5004/dwt.2023.29433

Nong, X., Huang, L., Chen, L., Wei, J., & Li, R. (2024). Distribution, relationship, and environmental driving factors of chlorophyll-a and algal cell density: A national view of China. *Global Ecology and Conservation*. https://doi.org/10.1016/j.gecco.2024.e03084

Rebello, L. R. B., Siepman, T., & Drexler, S. (2020). Correlations between TDS and electrical conductivity for high-salinity formation brines characteristic of South Atlantic pre-salt basins. *Water SA*, 46, 602–609. https://doi.org/10.17159/wsa/2020.v46.i4.9073

Rusydi, A. (2018). Correlation between conductivity and total dissolved solid in various type of water: A review. *IOP Conference Series: Earth and Environmental Science*, 118. https://doi.org/10.1088/1755-1315/118/1/012019

Syeed, M., Hossain, M. S., Karim, Md. R., Uddin, M. F., Hasan, M., & Khan, R. (2023). Surface water quality profiling using the water quality index, pollution index and statistical methods: A critical review. *Environmental and Sustainability Indicators*. https://doi.org/10.1016/j.indic.2023.100247

Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access*, 8, 56855–56866. https://doi.org/10.1109/ACCESS.2020.2982225

Xu, G., Wei, H., Wang, J., Chen, X.-B., & Zhu, B. (2022). A Local Weighted Linear Regression (LWLR) Ensemble of Surrogate Models Based on Stacking Strategy: Application to Hydrodynamic Response Prediction for Submerged Floating Tunnel (SFT). *Applied Ocean Research*. https://doi.org/10.1016/j.apor.2022.103228

Xu, J., Wang, Z., Li, X., Li, Z., & Li, Z. (2024). Prediction of Daily Climate Using Long Short-Term Memory (LSTM) Model. *International Journal of Innovative Science and Research Technology (IJISRT)*. https://doi.org/10.38124/ijisrt/ijisrt24jul073

Yaroshenko, I., Kirsanov, D., Marjanovic, M., Lieberzeit, P., Korostynska, O., Mason, A., Frau, I., & Legin, A. (2020). Real-Time Water Quality Monitoring with Chemical Sensors. *Sensors*, 20. https://doi.org/10.3390/s20123432

Zhang, Q., Wang, H., Dong, J., Zhong, G., & Sun, X. (2017). Prediction of Sea Surface Temperature Using Long Short-Term Memory. *IEEE Geoscience and Remote Sensing Letters*, 14, 1745–1749. https://doi.org/10.1109/LGRS.2017.2733548

Zhang, Y., Li, J., Zhou, Y., Zhang, X., & Liu, X. (2024). Artificial Intelligence-Based Microfluidic Platform for Detecting Contaminants in Water: A Review. *Sensors*, 24. https://doi.org/10.3390/s24134350

Zhang, Y., Li, J., Zhou, Y., Zhang, X., & Liu, X. (2024). Artificial Intelligence-Based Microfluidic Platform for Detecting Contaminants in Water: A Review. *Sensors (Basel, Switzerland)*, *24*. https://doi.org/10.3390/s24134350