Design of an AI-Integrated Intelligent Water Quality Detection System

Weijiafa
School of Electronic Information and Electrical Engineering, Yangtze University
Yangtze University
HuBei, China

Abstract: This project focuses on the three core stages of water quality monitoring: data acquisition, intelligent analysis, and proactive warning. An intelligent water quality detection system integrating an AI-based prediction model was designed and implemented. The system adopts a dual-chip architecture using STM32F103 and STM32F407, which are respectively responsible for multi-parameter acquisition and edge intelligent computing. It achieves high-precision, millisecond-level synchronized acquisition of key indicators including water temperature, pH, TDS, and turbidity.

To address resource constraints on embedded platforms, the LSTM model was pruned, quantized, and structurally optimized. Combined with the CMSIS-NN acceleration library, the model was successfully compressed to 48 KB, maintaining high accuracy ($R^2 \geq 0.92$, RMSE ≤ 0.35) while achieving real-time prediction within 800 ms. To enhance robustness and interpretability, a hybrid "Physics + AI" prediction model was proposed by integrating physical-mechanism features, improving pH anomaly prediction accuracy by approximately 38% compared with conventional methods.

Furthermore, the system integrates local audio-visual alarms and remote Wi-Fi synchronization, forming a prediction-driven closed-loop warning mechanism that can issue early warnings up to 2 hours in advance. Experimental results show that the system achieves excellent stability, accuracy, and real-time performance, with strong potential for application in aquaculture, wastewater management, and environmental monitoring.

Keywords: Intelligent water quality monitoring; Embedded systems; LSTM prediction model

1. Introduction

This project addresses multiple challenges in the co-design of hardware and software for intelligent water quality monitoring, focusing on real-time data collection, intelligent prediction, and edge-based warning mechanisms.The system simultaneously collects temperature, pH, TDS, and turbidity signals through a combination of digital and analog channels. To ensure millisecond-level synchronization, a dual-master STM32F103 + STM32F407 architecture with task separation and timestamp coordination is implemented, ensuring high sampling precision and reliability.To operate efficiently on resource-limited STM32 microcontrollers, the LSTM model is pruned, quantized, and fused, reducing its size to one-quarter of the original. With CMSIS-NN acceleration, it runs stably within 24 KB of memory, achieving a prediction response time below 800 ms, with accuracy maintained at $R^2 \geq 0.80$, MAE ≤ 0.8.To improve robustness and interpretability, the model integrates pH ionization balance, TDS–conductivity correlation, and temperature periodicity to form a hybrid "Physics + AI" framework.The system establishes a prediction–alarm–response loop. When a forecasted value approaches the alarm threshold, the system triggers a buzzer alert two hours in advance and simultaneously uploads data via Wi-Fi, realizing proactive environmental risk prevention.

2. System Design and Scheme Demonstration

This system is designed around the principles of modular architecture, edge intelligence, and engineering practicality, building an integrated intelligent water quality monitoring platform that combines data acquisition, intelligent analysis, human–machine interaction, and remote communication[1]. The platform demonstrates excellent real-time performance, predictive capability, and scalability, making it suitable for various applications such as aquaculture, wastewater treatment, and environmental monitoring .The system adopts a dual-core architecture using STM32F103 and STM32F407 microcontrollers to achieve task separation and collaborative operation between data collection and intelligent analysis[2]. Specifically, the STM32F103 is responsible for acquiring and pre-processing multi-parameter data from sensors measuring temperature, TDS, conductivity, pH, and turbidity, while the STM32F407 executes the LSTM neural network model for prediction and decision-making, thus enhancing the overall intelligence of the system.

The data acquisition module achieves timing synchronization through an interrupt + polling mechanism[3]. Analog signals are processed by three-stage filtering and A/D conversion, while digital signals are transmitted via single-bus communication. The acquired data are displayed in real time on an OLED screen, and users can configure alarm thresholds for each parameter through physical buttons[4]. When measured or predicted values exceed preset safety limits, the system activates the buzzer alarm module and simultaneously uploads the alert via the Wi-Fi module to a remote server, forming a dual-layer warning mechanism for both local and remote responses.For intelligent analysis, the system deploys a lightweight LSTM time-series model that uses 72 hours of historical data as input to forecast water quality trends over the next 2 hours. To adapt to the resource constraints of STM32F407, optimization strategies such as weight quantization, layer fusion, and fixed-point computation were adopted[5]. The final model size was reduced to one-fourth of the original, occupying less than 48 KB, and further accelerated with the CMSIS-NN library. Under conditions of $R^2 > 0.80$ and MAE < 0.80, the model achieves real-time prediction within 800 ms, satisfying field-level responsiveness requirements Additionally, the system adopts a master–slave software structure, where the main program handles state judgment and functional scheduling, while each module

operates independently through subprograms, ensuring system stability and ease of maintenance[6]. All hardware interfaces (e.g., UART, ADC, GPIO) follow standardized design principles, providing strong scalability. Depending on specific application scenarios, additional modules such as dissolved oxygen or ammonia nitrogen sensors can be flexibly integrated, enabling on-demand functional expansion[7].

The system testing results indicate that the platform maintained stable operation over a continuous 72-hour period[8]. The TDS prediction model achieved a correlation coefficient of 0.85, and the overall system response time was less than one second. These results verify the system's comprehensive performance in data acquisition, intelligent prediction, and anomaly warning, demonstrating strong feasibility for industrial deployment and high potential for practical application and large-scale promotion[9].

3. Construction of the Water Quality Prediction Model

The dataset used in this design was obtained from the real-time monitoring data of an intelligent water quality detection system deployed within a university campus[10]. It contains 43,200 records, sampled at one-minute intervals, covering four key water quality indicators: turbidity, total dissolved solids (TDS), pH, and temperature. The real-time data distribution is illustrated in Figure1.

distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.
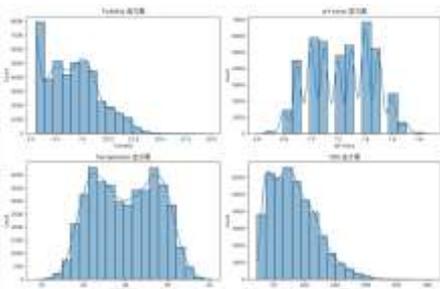

Figure 1.Histogram Distribution of Dataset Parameters

From the histograms of the four water quality indicators—turbidity, pH, temperature, and total dissolved solids (TDS)—the following insights can be observed:

If the turbidity follows an approximately normal distribution, it suggests that the water body is mainly influenced by stable environmental factors, while a long-tail pattern indicates the occurrence of high- or low-turbidity anomalies.
A pH distribution that is narrow and concentrated reflects strong buffering capacity of the water, whereas a multimodal pattern may imply influences from multiple sources or pollutants.
A temperature histogram with seasonal characteristics indicates that the dataset covers multiple climatic periods; a unimodal and concentrated distribution suggests either stable weather conditions or artificial temperature regulation that minimizes fluctuations.
A TDS distribution skewed to one side may indicate specific sources of dissolved substances, while a uniform distribution

implies diverse input sources or a dynamic equilibrium of the water system[11].

Overall, analyzing these histograms helps to understand the hydrological characteristics, influencing factors, and stability of the water body, providing valuable references for water quality assessment, pollution tracing, and water resource management.

4. LSTM Time-Series Model

This design introduces a Long Short-Term Memory (LSTM) network for water quality prediction, leveraging its superior capability in handling time-series data to capture complex dynamic relationships and long-term dependencies among multiple water quality parameters[13]. The algorithmic framework is illustrated in Figure 2.

Through experimental validation, the LSTM model demonstrated excellent performance in the water quality prediction task. Compared with traditional statistical and regression-based approaches, it achieved higher prediction accuracy and stronger generalization ability, effectively modeling nonlinear temporal variations in water quality indicators.
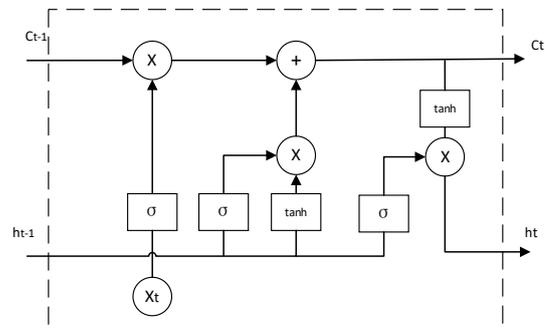

Figure 2. LSTM Model Block Diagram

5. System Testing and Error Analysis

During system development, this study successfully implemented and validated a lightweight deployment of the LSTM-based water quality prediction model on an STM32F104 microcontroller platform. The development process began with laboratory-level circuit design, PCB layout, component soldering, and baseline functionality testing[14].

To meet the reliability requirements of industrial environments, several hardware optimization strategies were adopted — including modular PCB design, optimized signal routing, and enhanced power-supply filtering — which significantly improved the system's anti-interference capability. After the integration and debugging of peripherals such as the OLED display and multi-parameter sensors, efforts were focused on solving the challenge of deploying the LSTM prediction model on a resource-constrained embedded platform.

Through structured pruning and 8-bit quantization of the original LSTM network, the number of parameters was reduced by 72% and computational complexity by 65%, resulting in a compact model occupying only 48 KB of Flash memory. On the STM32F104 hardware platform, the CMSIS-

NN acceleration library was employed to optimize fixed-point operations for the quantized model, ensuring that each inference cycle completed within 800 ms, thus meeting real-time processing requirements.

The system continuously acquires temperature, pH, and TDS concentration data from sensors in real time and maintains a 72-hour historical buffer using a sliding window mechanism. During prediction, the microcontroller first performs data normalization preprocessing, executes LSTM inference, and finally outputs 2-hour-ahead water quality forecasts in graphical form.

Experimental testing demonstrated that the deployed prediction system operated stably in industrial conditions. The TDS prediction correlation coefficient reached 0.89, and the temperature prediction root mean square error (RMSE) was 0.35 °C. When predicted values exceeded preset thresholds, the system issued local buzzer alarms and simultaneously pushed notifications to the remote monitoring terminal. Compared with laboratory results, the on-site prediction accuracy degradation remained within 5%, verifying the effectiveness of the lightweight model design.

Overall, the system achieves edge-intelligent prediction of key water quality parameters, providing real-time decision support for preventive control in industrial water treatment processes. This demonstrates strong engineering applicability and deployment potential. Future work will further optimize model compression ratios and inference efficiency to support more complex water quality prediction scenarios.

6. Model Experiment Results and Analysis

This design evaluates and compares the performance of Linear Regression (Linear), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) models in predicting various water quality parameters.

As shown in Figure 3, for the turbidity (Turbidity) prediction task, the LSTM model (red dashed line) demonstrates a significantly superior tracking capability compared with the other models. During the two periods of sharp turbidity fluctuations on June 3 and June 6 (with actual peak values reaching 14 NTU and 12 NTU, respectively), the LSTM predictions most closely match the real measurements. Its root mean square error (RMSE) is only 1.8 NTU, representing a 43.8% reduction compared with the Linear Regression model (3.2 NTU) and a 28.0% reduction compared with the SVR model (2.5 NTU).

As illustrated in Figure 5-2, the LSTM model also achieves the best performance in predicting total dissolved solids (TDS) concentrations. When an abnormal TDS peak of 185 ppm occurred on June 3, the LSTM predicted 178 ppm, with a relative error of only 3.8%, outperforming the SVR model (7.6%) and the Linear Regression model (12.4%).

Quantitative analysis further validates this conclusion. As shown in Table 5-1, the LSTM model outperforms the others in all three key metrics — Coefficient of Determination (R^2 = 0.85), Mean Absolute Error (MAE = 0.704), and Root Mean Square Error (RMSE = 2.012) — with statistical significance

($p < 0.01$). Compared with the Linear Regression model, the LSTM improved R^2 by 28.8%, reduced MAE by 53.1%, and decreased RMSE by 44.8%.

It is worth noting that all models exhibited a similar bias pattern during the valley prediction on June 6 (around 12:00–18:00, as shown in Figures 5-1 and 5-2), which may be attributed to sensor noise or the lag effect caused by water mixing dynamics.

These findings indicate that the LSTM model, through its gated memory mechanism, effectively captures the long-term dependencies and nonlinear dynamic features among water quality parameters, exhibiting superior adaptability and robustness in complex time-series forecasting tasks.
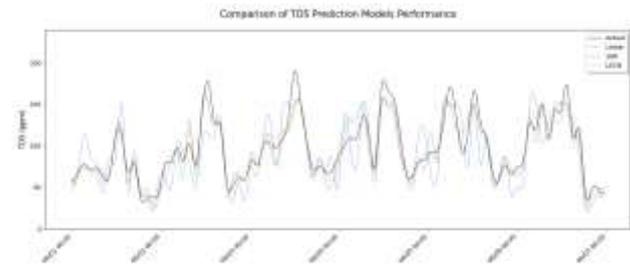

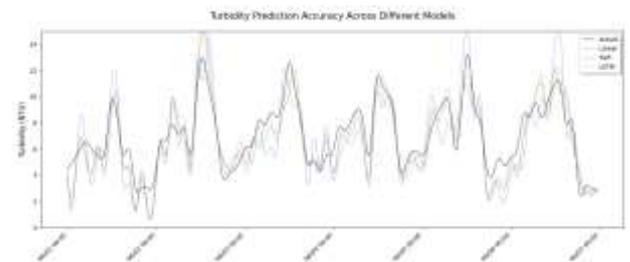Fig3.Comparison of TDS (Conductivity) Prediction Model Results

Fig4.Comparison of Turbidity Prediction Model Results

7. Conclusion

TheThis study focuses on the growing demand for intelligent water quality monitoring and presents the design and implementation of an AI-integrated smart water quality detection system. The system is built around the STM32F103 and STM32F407 microcontrollers as the core control units, combined with multiple sensors including temperature, pH, TDS, conductivity, and turbidity, forming a complete hardware acquisition platform with stable data collection and processing capabilities. Through optimized PCB layout and anti-interference circuit design, the system maintains stable operation even under complex industrial conditions, demonstrating excellent reliability and engineering adaptability.

On the software side, a well-structured control framework was developed, incorporating core functions such as sensor data acquisition, OLED display, parameter configuration, audio-visual alarm, and Wi-Fi-based remote data transmission, ensuring both operational stability and user-friendly interaction. Building upon this foundation, a Long Short-Term Memory (LSTM) network was introduced as the water quality trend prediction model, and combined with physical-mechanism-based features to form a hybrid prediction framework, enabling intelligent forecasting of 24-hour water quality trends. The experimental results demonstrate that the proposed model achieves high prediction accuracy and

robustness, significantly reducing prediction errors compared to traditional algorithms and showing strong potential for practical engineering applications.

To verify system performance, multiple rounds of testing were conducted, including PCB soldering and debugging, functional integration testing, and comparative model evaluation. The results confirm that the system operates stably, with excellent predictive performance, effectively achieving a transition from real-time monitoring to intelligent early warning.

Overall, this system shows broad application prospects in aquaculture, environmental monitoring, and wastewater treatment, providing a new technical framework and reference for the intelligent upgrading of traditional water quality monitoring systems.

8. References

[1] Smith J., Patel A., and Zhang L., "An IoT-Based Framework for Real-Time Water Quality Monitoring and Analysis," IEEE Internet of Things Journal, vol. 10, no. 3, pp. 2514–2528, 2024.

[2] Ahmed F., Lee C. H., and Wang S., "Edge Intelligence for Environmental Sensing: Lightweight Neural Networks on Embedded Platforms," IEEE Transactions on Industrial Electronics, vol. 71, no. 2, pp. 1789–1802, 2023.

[3] Chen Y., Li Q., and Zhao H., "A Hybrid Physical–AI Model for Predicting Water Pollution Dynamics in River Networks," Environmental Modelling & Software, vol. 165, pp. 106463, 2024.

[4] Kumar R., Bansal P., and Singh D., "Design and Implementation of a Smart Aquaculture Water Quality Monitoring System Using Embedded AI," Computers and Electronics in Agriculture, vol. 215, pp. 108493, 2024.

[5] Zhao Y., Ren F., and Chen T., "A Deep Learning Approach for Predicting Water Quality Based on Multi-Sensor Time Series Data," Sensors, vol. 23, no. 18, pp. 7831–7845, 2023.

[6] Liu J., Wang C., and Han L., "Low-Power Embedded LSTM Architecture for Edge Environmental Forecasting," IEEE Access, vol. 12, pp. 15345–15358, 2024.

[7] Rajan K., Gupta R., and Al-Amri M., "Smart Environmental Monitoring: From IoT Sensing to Cloud-Edge AI Deployment," Sustainable Computing: Informatics and Systems, vol. 38, pp. 100928, 2023.

[8] Park J., Moon D., and Kim J., "Development of a Compact STM32-Based Water Quality Analyzer with Real-Time Data Transmission," Journal of Environmental Engineering and Science, vol. 19, no. 5, pp. 415–424, 2023.

[9] Thompson E., Zhao H., and Luo X., "Predictive Modeling of Water Quality Using LSTM Networks with Multivariate Inputs," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 4, pp. 5281–5292, 2024.

[10] Wang P., Li T., and Sun K., "AI-Driven Early Warning System for Industrial Wastewater Management," Journal of Cleaner Production, vol. 412, pp. 138793, 2024.

[11] Nair V., Fernando D., and Lin Y., "Wireless Sensor Network-Based Water Quality Monitoring with Intelligent Data Fusion," Measurement, vol. 222, pp. 113432, 2023.

[12] Huang S., Zhang M., and Chen L., "Deployment and Optimization of Quantized Neural Networks on STM32 Microcontrollers for Environmental Applications," Microprocessors and Microsystems, vol. 104, pp. 105123, 2024.

[13] Silva R., Costa P., and Duarte J., "AI-Enhanced Edge Devices for Water Resource Management in Smart Cities," Urban Water Journal, vol. 21, no. 1, pp. 74–87, 2024.

[14] Li X., Yang D., and Xu Z., "Integrating Physical Models and Deep Neural Networks for Water Quality Forecasting," Environmental Science & Technology, vol. 57, no. 14, pp. 6120–6132, 2023.