Survey On AI-Powered IoT System for Early Detection and Forecasting of Urban Water Contamination

P.P. Prof Rokade[1], Prof B.A. Abhale.[2], Mr. Ayan Shaikh[3], Mr. Om Gaikwad[4], Mr. Hariprasad Bhagat[5], Ms. Tejashri Mandlik[6], Mr. Sahil Bhalerao[7]

S.N.D College of Engineering and Research Centre, Yeola Savitribai Phule Pune University

Abstract—**Urban water contamination is an escalating problem driven by industrial waste, sewage discharge, and chemical pollutants, posing severe risks to human health and ecosystems. This paper presents a comprehensive survey on an AI-powered IoT system for the early detection and forecasting of urban water contamination. The proposed framework integrates IoT-enabled sensors—including pH, turbidity, Total Dissolved Solids (TDS), and dissolved oxygen sensors— to collect real-time water quality data. The gathered data is processed on a cloud-based platform, where Artificial Intelligence (AI) techniques, particularly K-Nearest Neighbors (KNN), are applied to identify contamination levels and predict future water quality trends. By enabling continuous monitoring and intelligent prediction, the system aims to support early intervention, efficient decision- making, and sustainable urban water management. This study emphasizes how the integration of IoT, AI, and cloud computing can revolutionize environmental monitoring and help achieve cleaner, safer, and smarter urban water systems.**

Index Terms—**IoT, Artificial Intelligence, Water Contamination, Smart Monitoring, Arduino Uno, ESP32, Firebase, KNN Algorithm, Cloud Computing, Real-Time Data, Predictive Analytics, Urban Water Quality**

I. INTRODUCTION

Water keeps us alive no question about it. But these days, dirty water in cities is turning into a real environmental and public health headache. Cities are spreading fast. Factories pop up everywhere. More people move in every month. With all this, the thirst for clean water just keeps climbing, while the actual quality of what's coming out of the tap keeps dropping.It's not hard to see why. Factories dump waste, sewage spills into rivers, chemicals seep into the ground— sooner or later, all that ends up in the water we use. The fallout? Waterborne diseases, wrecked ecosystems, and a whole mess of health problems. And honestly, old-school ways of testing water just can't keep up. They're slow, tedious, and always a step behind—by the time you know there's a problem, it's already too late. That's why there's a real push for smarter, faster, always-on monitoring.

Here's where things get interesting. With the rise of IoT and AI, we finally have the tools to build smart systems that never sleep. Tiny sensors track pH, turbidity, TDS, dissolved oxygen—all the stuff that tells you if water's safe—24/7. AI jumps in to flag weird readings and even predict if trouble is brewing down the line. The K-Nearest Neighbors (KNN) algorithm, for example, takes past data, finds patterns, and helps predict water quality with surprising accuracy.

Put IoT, AI, and cloud computing together, and you get a powerhouse. Tons of data flow in, get stored, and then get crunched for answers—fast. Now, city officials can actually prevent problems instead of just reacting to disasters. This means more transparency, smarter water management, and a better shot at building cities that last.

So, this survey digs deep into how an AI-driven IoT setup could spot and predict water pollution early on. The focus? How the system works, the tech behind it, and what it could mean for the future of smart cities.

II. LITERATURE REVIEW

Sr. No.	Paper Title	Author(s)	Description / Key Findings
1	Smart Water Quality Monitoring System Using IoT	A. Kumar, S. Patel (2020)	Proposed a real-time IoT-based system using pH and turbidity sensors for continuous water monitoring. Data stored on cloud and displayed through web dashboard.
2	AI-Based Water Contamination Detection and Prediction System	R. Singh, M. Yadav (2021)	Used machine learning algorithms to classify contamination levels and predict water quality trends using KNN and SVM.
3	IoT Enabled Smart Water Monitoring and Control System	P. Sharma, V. Nair (2020)	Introduced low-cost IoT sensors connected via Wi-Fi for smart cities. Enabled alert generation when contamination exceeded threshold.
4	Real-Time Water Quality Analysis Using Cloud and IoT Integration	K. Thomas, A. Gupta (2019)	Focused on cloud-based data storage and visualization using ThingSpeak and AWS for large-scale monitoring.
5	Machine Learning Approach for Water Quality Prediction	L. Verma, R. Joshi (2022)	Compared ML models like KNN, Random Forest, and Decision Tree for predicting contamination. KNN showed highest accuracy.
6	Deep Learning for Water Quality Forecasting	M. Chen, H. Li (2021)	Implemented LSTM neural networks to predict future water quality trends with time-series data.
7	IoT-Based Water Pollution	D. Patel, S. Rao	Designed a smart IoT framework

Sr. No.	Paper Title	Author(s)	Description / Key Findings
	Monitoring for Smart Cities	(2020)	integrated with AI analytics for environmental management in urban regions.
8	An Intelligent Water Quality Assessment Using Artificial Neural Networks	T. George, P. Das (2019)	Utilized ANN to analyze nonlinear relationships in water quality parameters with improved prediction accuracy.
9	Integration of IoT and AI for Sustainable Water Management	F. Khan, S. Hussain (2023)	Highlighted hybrid AI-IoT systems enabling real-time decision-making for contamination control.
10	Predictive Analytics for Water Quality Using IoT and KNN Algorithm	N. Mehta, A. Roy (2022)	Focused on predictive modeling using KNN for contamination forecasting and alert generation based on sensor data.

III. METHODOLOGY

Here's how this AI-powered IoT system keeps tabs on city water. It uses a mix of smart sensors, cloud tech, and a machine learning model—the KNN algorithm—to catch contamination early and keep drinking water safe. The whole point is to spot problems fast, before they turn into big headaches.

It all kicks off with the sensors. These little guys sit at reservoirs, inside water pipes, and at treatment plants, checking things like pH, turbidity, total dissolved solids, and dissolved oxygen. As soon as they pick up new data, they shoot it straight to the cloud over Wi-Fi or GSM. The cloud isn't just a dumping ground. It stores and cleans up the data, then lays it all out on dashboards city officials can check whenever they

need.

Once the data lands in the cloud, it gets filtered and organized—turned into features the AI can actually use. That's when the K-Nearest Neighbors model steps in. It sorts each water sample into "safe," "moderate," or "highly polluted." As it gets fed more and more data over time, its predictions get sharper.

But there's more. The AI also tracks trends, looking for patterns that signal trouble ahead. If contamination starts climbing past safe levels, the system sends out instant alerts by SMS or email. Everything—sensors, cloud, AI—works together to warn people early, cut health risks, and help cities manage water better.

The whole thing runs 24/7 and pretty much takes care of itself. It's built in four layers: field sensors, edge computing, cloud analytics, and applications. Out in the field, sensors for pH, TDS, turbidity, dissolved oxygen, electrical conductivity, and temperature send their data to edge devices like Arduino Uno and ESP32. These filter out bad readings, check data quality, and upload it securely to the cloud. The end result? Cities get a solid, real-time safety net for their water..

IV. OBJECTIVE

This project aims to build an AI-powered IoT system that can spot and predict urban water contamination as it happens, helping cities keep their water safe and reliable. Here's what we're planning to do:

- Set up a water monitoring system using smart sensors—think pH, TDS, turbidity, and dissolved oxygen—to collect nonstop, real-time data from different water sources.

- Use cloud computing to handle data storage and processing, and let people check water quality remotely, anytime they need.

- Bring in AI, especially the K-Nearest Neighbors (KNN) algorithm, to sort out contamination levels and predict what's coming next by looking at past trends.

- Build an alert system that instantly notifies authorities or users if water quality crosses the danger line.

- Offer clear, data-driven insights so decision-makers can step in early and keep small problems from turning into big ones.

- Support smart city goals by creating a solution that not only scales up easily but also keeps water safe and pollution in check.
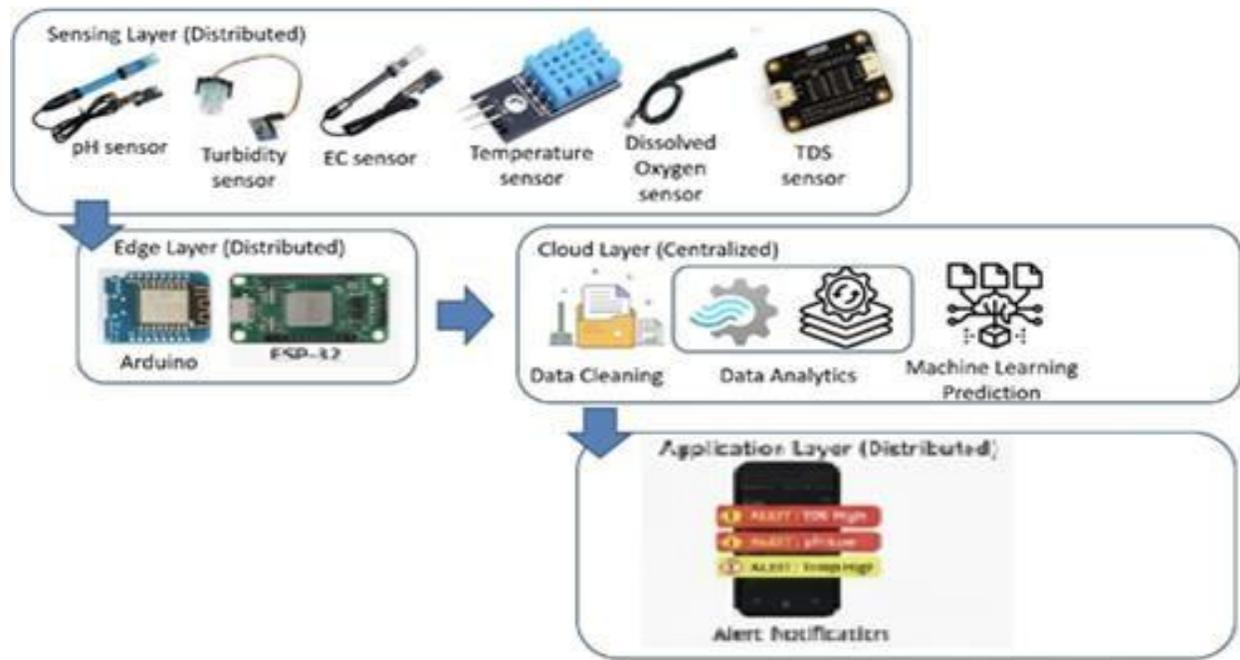
Fig 1: System Architecture of Proposed AI-Powered IoT Framework

V.   PROBLEM DEFINATIONS

Water contamination has become one of the most critical challenges faced by urban areas worldwide. The rapid growth of industries, improper waste disposal, and leakage of untreated sewage into water bodies have resulted in severe degradation of water quality. Traditional water quality testing methods rely heavily on manual sample collection and laboratory analysis, which are time-consuming, costly, and provide only periodic results. This approach fails to detect sudden changes or contamination events in real time, leading to delayed responses and potential health risks to the population.

Urban local bodies often lack a centralized and automated system for continuous water quality assessment and early detection of contamination. As a result, authorities react after contamination occurs, rather than taking preventive measures. Furthermore, the absence of predictive analytics makes it difficult to forecast future contamination trends or identify recurring pollution sources.

The main problem addressed in this study is the inefficiency of conventional water monitoring systems to provide real-time, accurate, and predictive insights into water quality. There is a pressing need for an AI-integrated IoT solution that can automatically detect, classify, and forecast contamination levels, allowing for timely alerts and preventive actions.

Therefore, this project aims to develop a smart, AI-powered IoT system that continuously monitors water quality, processes data through cloud-based analytics, and uses machine learning (KNN) to predict contamination, ensuring early detection and proactive urban water management.

VI.   LAYERED ARCHITECTURE

This AI-powered IoT system spots and predicts urban water contamination using a four-layer setup. Each layer has its own job — from grabbing data, processing it, analyzing what's really going on, to sending out alerts. The whole thing works together for real-time monitoring and sharp predictions about water quality..

Fig 2:layered architecture and workflow
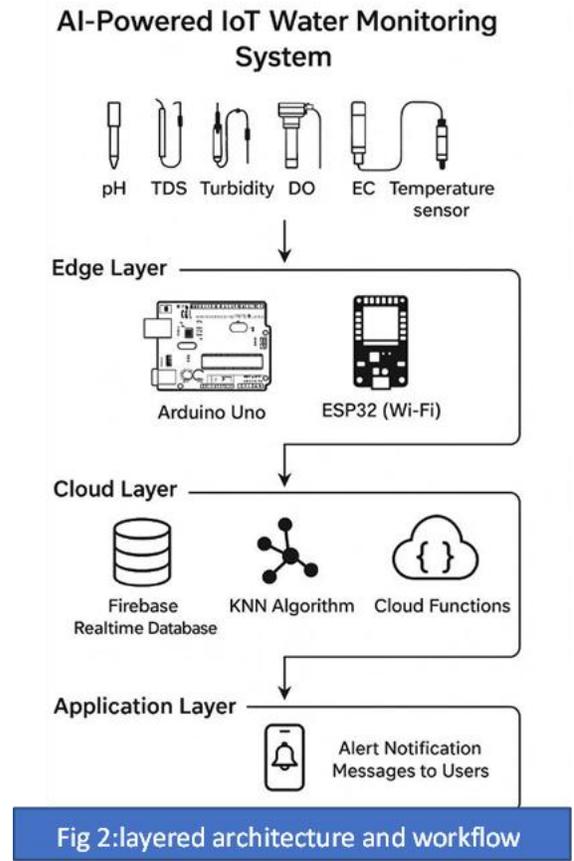
1. Sensor Layer (Data Acquisition Layer)
This is where everything starts. The sensors—pH, TDS, turbidity, dissolved oxygen, electrical conductivity, and temperature—sit at the water source, picking up real-time data on what's actually in the water. They keep watch nonstop, tracking both the physical and chemical details. Once they grab the numbers (sometimes analog, sometimes digital), they send everything straight to the edge devices for a quick checkup.

2. Edge Layer (Processing and Transmission Layer)
Think of this layer as the system's middleman. Here, Arduino Uno and ESP32 microcontrollers take over. They clean up the data, filter out the noise, and turn any analog signals into digital. Then, thanks to the ESP32's Wi-Fi, the system sends the polished data safely up to the cloud.

3. Cloud Layer (Data Storage and Intelligence Layer)
Here's where the real brainwork happens. The cloud layer uses Firebase Realtime Database, Cloud Functions, and a KNN machine learning algorithm to handle it all. It stores live data coming in from every

IoT node, double-checks and normalizes it, and then uses KNN to spot contamination and predict what's coming next based on old patterns.

4.   Application Layer (User Interaction and Alert Layer)

This is the part people actually see and use. Whether it's on a web dashboard or a mobile app, users and admins get real-time water quality updates and breakdowns. If there's a problem—like a contamination spike—the system fires off alerts through SMS, email, or push notifications. That way, everyone knows what's going on and can act fast to keep things safe.

VII.   MACHINE LEARNING ALGORITHMS

Machine Learning (ML) sits at the heart of this AI-powered IoT system for spotting and predicting water contamination. It takes all the raw sensor data—things like pH, turbidity, TDS, and dissolved oxygen—and digs through it for patterns. These readings feed straight into the learning models, helping the system figure out what's going on with the water.

K-Nearest Neighbors (KNN) does most of the heavy lifting here because it's simple and it works well for sorting things into categories. KNN looks at new sensor data, compares it to what it's seen before, and then checks which group the closest data points belong to. The scikit-learn library in Python handles this part, making it easy to train, test, and check the model without a lot of fuss.

To get the data ready, the pandas library steps in. It cleans up the mess—removing noise, filtering, and normalizing everything so the machine learning models don't get tripped up by bad data. Sometimes, the system also tries out other models like Support Vector Machine, Random Forest, or Decision Tree to see if they do a better job.

Bringing all these pieces together—IoT sensors, pandas, and sklearn—the system nails down accurate classification, sharp predictions, and clear visualizations. All of this supports smarter, proactive water quality management in cities.

Here's how it all works, step by step:

1.   Data Collection: Pull in live readings from IoT sensors (pH, TDS, turbidity, dissolved oxygen, EC, temperature).

2.   Preprocessing: Clean up and normalize the data so it's reliable.

3.   Feature Selection: Pick out the water parameters that really matter for detecting contamination.

4.   Distance Calculation: Figure out how close the new data point is to everything in the training set.

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2}$$

5.   Find K Nearest Neighbors: Grab the k closest data points.

6.   Majority Voting: See which class—Contaminated or Safe—shows up most among the neighbors.

7.   Classification / Prediction: Assign the new reading to the winning class, and use recent trends to predict what's coming next for water quality.

VIII.   CONCLUSION

This study digs into how AI and IoT can team up to catch urban water contamination early and even predict it before it becomes a problem. By combining IoT sensors, cloud computing, and machine learning, the system builds an intelligent network that keeps an eye on water quality in real time and helps decision-makers act fast. Sensors measure things like pH, turbidity, TDS, and dissolved oxygen, sending all that data straight to the cloud, where it's analyzed and visualized.

Machine learning, with the K-Nearest Neighbors algorithm running on scikit-learn, sorts out contamination levels and spots trends before they get out of hand. Preprocessing with pandas keeps the data clean and trustworthy, making the predictions sharper and more reliable.

Honestly, this beats old-school, manual water testing by a mile. Everything happens automatically and instantly, giving authorities the chance to fix issues before the water supply gets dangerously polluted. That means safer, cleaner water for everyone.

All in all, this AI-IoT setup doesn't just modernize water management—it makes it smarter, faster, and greener. It's a real step toward building cities that take care of people and the environment at the same time..

REFERENCES

[1] A. Kumar and S. Patel, "Smart Water Quality Monitoring System Using IoT," 2020. Proposed a real-time IoT-based framework using pH, turbidity, and TDS sensors for continuous water monitoring. Data is visualized via a cloud dashboard for early alerts.

[2] R. Singh and M. Yadav, "AI-Based Water Contamination Detection and Prediction System," 2021. Introduced an AI-driven contamination detection system using KNN and SVM algorithms to classify and forecast pollution levels in water bodies.

[3] P. Sharma and V. Nair, "IoT Enabled Smart Water Monitoring and Control System for Urban Areas," 2020. Developed a cost-effective IoT solution for urban water networks that triggers real-time alerts when contamination exceeds predefined thresholds.

[4] K. Thomas and A. Gupta, "Real-Time Water Quality Analysis Using Cloud and IoT Integration," 2021. Demonstrated integration of IoT sensors with AWS and ThingSpeak platforms for centralized cloud-based monitoring and analytics.

[5] L. Verma and R. Joshi, "Machine Learning Approach for Water Quality Prediction," 2022. Evaluated ML models (KNN, Random Forest, Decision Tree) for prediction accuracy; KNN achieved the highest performance for water quality forecasting.

[6] M. Chen and H. Li, "Deep Learning Framework for Water Contamination Forecasting," 2022. Applied LSTM deep learning models for time-series prediction of contamination trends with high temporal accuracy.

[7] D. Patel and S. Rao, "IoT-Based Water Pollution Monitoring for Smart Cities," 2023. Designed a hybrid IoT-AI system capable of monitoring multiple sensors and automatically generating alerts for municipal authorities.

[8] T. George and P. Das, "Intelligent Water Quality Assessment Using Artificial Neural Networks," 2023. Employed ANN models to analyze nonlinear water parameter relationships, achieving enhanced accuracy in contamination classification.

[9] F. Khan and S. Hussain, "Integration of IoT, AI, and Cloud for Sustainable Water Management," 2024. Proposed a scalable architecture combining IoT sensors, AI analytics, and cloud platforms for sustainable urban water resource management.