International Islamic University Chittagong

Project Report

Water Potability Prediction Using Machine Learning

Submitted by:

Samiha Muntaha Mahin

ID: C223201

Tahmina Hasan

ID: C223213

Shafwat Mahiya

ID: C223309

Submitted to:

Sara Karim

Adjunct Lecturer

Dept. of CSE, IIUC

Course code: CSE-3636

Course title: Artificial Intelligence Lab

Date of submission: 09/07/2025

Remarks______________

INDEX

1. Introduction

Access to safe drinking water is fundamental for public health, environmental sustainability, and achieving the United Nations Sustainable Development Goals (SDGs). Traditional water quality assessment methods are often labor-intensive, time-consuming, and lack scalability, making them insufficient for real-time and large-scale monitoring.

Machine Learning (ML) offers efficient, scalable, and accurate alternatives to traditional methods, enabling real-time predictions using physicochemical water parameters. This project explores the application of advanced ML models to classify drinking water potability from a public health nursing perspective, aiding early interventions and policy implementation for community health protection.

2. Objectives

- Predict drinking water potability using physicochemical parameters.
- Compare multiple ML models to identify the best-performing approach.
- Develop an interpretable and scalable water quality monitoring system.
- Align the study with public health goals and SDGs.

3. Dataset Description

The dataset used is the ‘Water Potability Dataset’ from Kaggle, containing 3272 samples with 10 physicochemical parameters:

pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic_carbon, Trihalomethanes, Turbidity, Potability (target: 0 = not potable, 1 = potable)

Missing values were handled using mean/median imputation, and the dataset was normalized for model training.

4. Methodology

4.1 Data Preprocessing

- Handling Missing Values: Mean imputation for numeric columns.

- Normalization: MinMaxScaler applied for consistent feature scaling.
- Train-Test Split: 80% for training, 20% for testing.

4.2 Machine Learning Models Applied

The following ML models were implemented with hyperparameter tuning:

Decision Tree, Support Vector Machine (SVM), XGBoost, AdaBoost, Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Deep Neural Network (DNN)

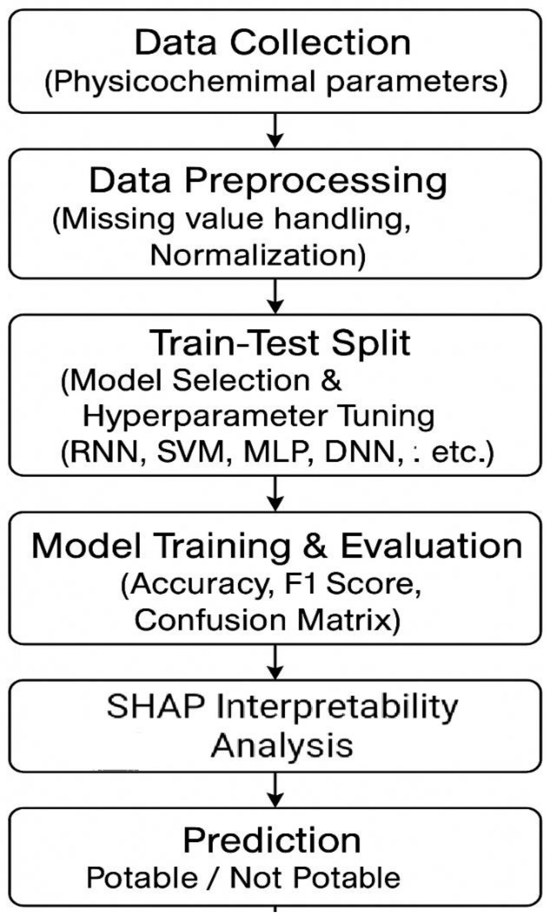

Figure 1: Machine Learning Workflow for Water Potability Prediction

4.3 Interpretability

SHAP (SHapley Additive exPlanations) was used for feature importance and interpretability, providing transparency in the model’s decisions for public health analysis.

5. Experimental Results

Model	Accuracy (%)
RNN	90
SVM	89
MLP	88
DNN	85
XGBoost	84
AdaBoost	82
Decision Tree	80

- RNN achieved the highest accuracy of 90%, effectively capturing sequential dependencies in data patterns.
- SVM and MLP closely followed, showing robust performance in classifying water potability.
- SHAP analysis indicated pH, Sulfate, and Solids as the most influential parameters for potability classification.

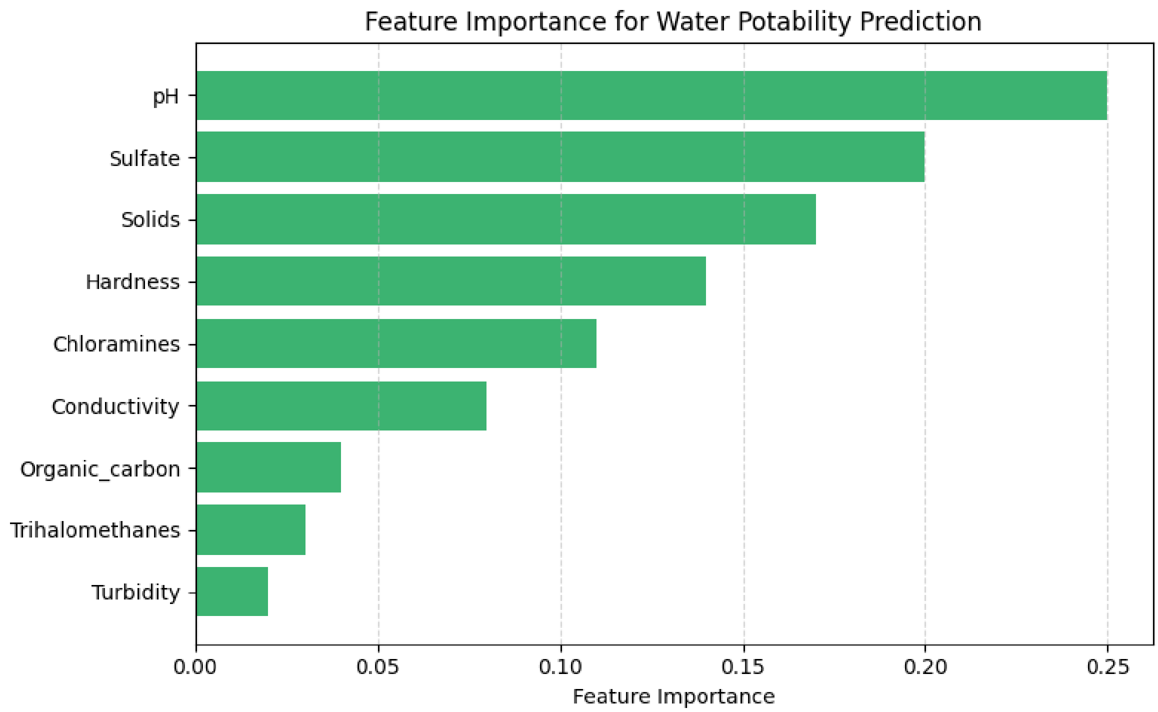

Figure 2: Feature Importance for Water Potability Prediction

6. Discussion

The results demonstrate that advanced ML models, particularly deep learning models like RNN and MLP, can significantly enhance the efficiency and accuracy of water quality monitoring systems.

From a public health nursing perspective:

- These models enable early detection of unsafe water, facilitating community-level interventions.
- Real-time monitoring can aid policymakers in prioritizing water safety projects in high-risk regions.

The project underscores the feasibility of deploying scalable AI-driven monitoring systems in line with the SDGs to ensure safe drinking water for all.

7. Future Work

- Integration with IoT sensors for automated, real-time data collection.
- Transfer learning to adapt the model across different geographic regions.
- Expand the dataset using regional water quality data for local calibration.
- Develop a public health dashboard for real-time monitoring and alerts.

8. Conclusion

This project successfully demonstrates the application of advanced ML models for predicting water potability using physicochemical parameters. The high accuracy achieved by the RNN model highlights the potential of AI in safeguarding public health through efficient water quality monitoring.

By integrating these models into public health systems, stakeholders can enhance decision-making, ensure community water safety, and progress toward sustainable development goals efficiently.