Ensemble Learning for Early Flood Warning System in Wastewater Networks

1st Sampreet Ajjanagouda Patil
Computer Science Dept.
Texas A&M University-Corpus Christi
Corpus Christi, USA
spatil1@islander.tamucc.edu

2nd Gowtham Kilaru
Computer Science Dept.
Texas A&M University-Corpus Christi
Corpus Christi, USA
gkilaru@islander.tamucc.edu

3rd Khaja Faizan Mohammad
Computer Science Dept.
Texas A&M University-Corpus Christi
Corpus Christi, USA
kmohammad@islander.tamucc.edu

4th Mohammed Taha Ayub Khan
Computer Science Dept.
Texas A&M University-Corpus Christi
Corpus Christi, USA
mkhan30@islander.tamucc.edu

Overleaf link: https://www.overleaf.com/read/hpzmdnqgtvcy

I. Introduction

Urban flooding due to overwhelmed wastewater systems during heavy rainfall events is a serious concern for city infrastructure and public health.(1) Wastewater overflow can lead to contamination, infrastructure damage, and hazardous conditions. Traditionally, hydraulic models such as SWMM have been used for flood simulation, but their computational intensity makes them less suitable for real-time alerts. This project introduces a machine learning-based ensemble approach for predicting water depths at critical junctions in wastewater networks.(2) Using time-series data on water depth and inflow rate, along with node-specific metadata, the model aims to trigger alerts when potential flood conditions are detected.(1)

II. Related Work

Recent years have seen an increasing application of AI to water management. Table I summarizes recent work:

TABLE I
Summary of Related Work

Study	Method	Focus
Rathore et al.	LSTM	Time-series water level prediction
Rao et al.	IoT + ML	Real-time flood alert and monitoring
Wang et al.	Ensemble ML	Combination of models for hydrology

These studies demonstrate the potential of data-driven models and smart infrastructure. Our method extends these concepts to the wastewater domain using an ensemble system combining different learners tailored for both structural and temporal attributes.

III. Proposed Method

We designed an ensemble architecture that integrates five machine learning models. Each model captures different relationships in the data.(2) Table II describes their type and contribution.

TABLE II
Model Roles in the Ensemble

Model	Type	Strength
MLP	Neural Net	Captures nonlinear patterns
Random Forest	Tree-Based	Handles categorical and numeric data well
XGBoost	Boosting Tree	Good for structured data and outlier resistance
Linear Regression	Statistical	Baseline interpretability
SVM	Kernel Method	Handles high-dimensional data

The final prediction is computed via soft voting. Classification into 'flood' and 'non-flood' is done using a water depth threshold.(3) All components were developed in Python3 using PyTorch and scikit-learn.

IV. Experiments

A. A. Datasets

We utilized:

- **Flow Depth:** Time-stamped water level at each node.
- **Flow Rate:** Inflow rates per node.
- **Node Metadata:** Slope, pipe length, invert elevation, etc.

After loading, all data were merged by time index and cleaned. Missing values were imputed with column-wise mean.

B. B. Baselines

We compared the ensemble with each model individually. The models were trained with standard hyperparameters:

- MLP: Two hidden layers (128, 64), ReLU, Adam optimizer
- Random Forest/XGBoost: 100 trees
- SVM: RBF kernel
- Linear Regression: Ordinary least squares

C. C. Evaluation Results

70% of the data was used for training and 30% for testing. Evaluation was based on MSE and F1 Score.

TABLE III
PERFORMANCE COMPARISON OF MODELS

Model	**MSE**	**F1 Score**
Neural Network	0.0112	0.89
Random Forest	0.0104	0.91
Gradient Boost	0.0098	0.93
Linear Regression	0.0153	0.84
SVM	0.0127	0.88
Ensemble (Voting)	**0.0087**	**0.94**

The ensemble model consistently outperformed individual models in both regression and classification metrics.

REFERENCES

[1] R. Rao and Others, "Iot and machine learning for smart flood monitoring," *Sensors*, 2020.
[2] A. Rathore and Others, "Flood prediction using lstm in water systems," *Journal of Hydrology*, 2021.
[3] Y. Wang and Others, "Ensemble learning in hydrological forecasting," *Environmental Modelling*, 2022.