Pipe Roughness Estimation in Water Distribution Networks Using EPANET Net3 Data

Linyue Luo

CIE 500 Project

05/08/2025

1. Introduction

Pipe roughness is a key parameter affecting hydraulic performance in water distribution systems. As infrastructure ages, pipe conditions degrade, altering hydraulic resistance and flow behavior. Accurate prediction of pipe roughness can support proactive maintenance and improve simulation reliability.

However, direct measurement of roughness is rare, and many systems lack sufficient historical or condition-based data. This project explores data-driven approaches to infer unknown pipe roughness values using features available from hydraulic simulations and network topology. The goal is to test both Graph Neural Networks (GNNs) and Multilayer Perceptrons (MLPs) using data from the Net3 benchmark model.

2. Problem Statement

The problem addressed is the prediction of pipe roughness coefficients based on available hydraulic, topological, and node-level information in the Net3 EPANET model. This is a supervised regression task where the model learns to map input features (pipe and node attributes) to roughness labels.

Challenges include:

- Limited or noisy label data (roughness values are partially synthetic)
- Highly imbalanced feature-label relationships
- Small dataset size
- No historical failure or aging records

The primary goal is to evaluate the performance of GNNs and MLPs in predicting roughness under these constraints.

3. Data and Preprocessing

Net3 is a well-known benchmark water distribution network used in many studies to test water system models and simulations. It has 2 reservoirs, 97 junctions, and 117 pipes. It's a real-world-inspired urban water distribution system with a set number of junctions, pipes, and reservoirs, often used for testing hydraulic and contaminant transport models.

The Net3 EPANET model was used as the base hydraulic network. The following preprocessing steps were performed:

- A hydraulic simulation was run using WNTR for 3 days at 2-hour intervals.

- The average flow rate over time was computed for each pipe.
- Pipe-level features were collected: length, diameter, average flow rate, and headloss.
- Node-level features (elevation and base demand) were extracted for both start and end nodes of each pipe.
- Additional features: elevation difference and demand sum.

Roughness Labels:

The original roughness values in Net3 were synthetically modified:

- 90% of pipe roughness values were randomly perturbed into 3 ranges (low, medium, high) to simulate aging and heterogeneity.

4. Methodology

Method 1: Graph Neural Network (GNN) with Edge-Node Features

A GNN model was developed using the PyTorch Geometric library.

- Graph Construction:
  - Nodes represent junctions, tanks, and reservoirs.
  - Edges represent pipes.
  - Edge attributes included pipe-level features.
  - Node features included elevation and demand.
- Flipping Strategy:
  - To match edge attributes with source/target node features, a flipped edge index was constructed.
  - Node attributes from start and end nodes were concatenated per edge.
- Target: Roughness coefficient for each pipe (edge).
- Challenge:
  - High dimensionality mismatch when pairing edge attributes with node features.
  - Limited improvement in prediction accuracy. Model clustered predictions near the mean.

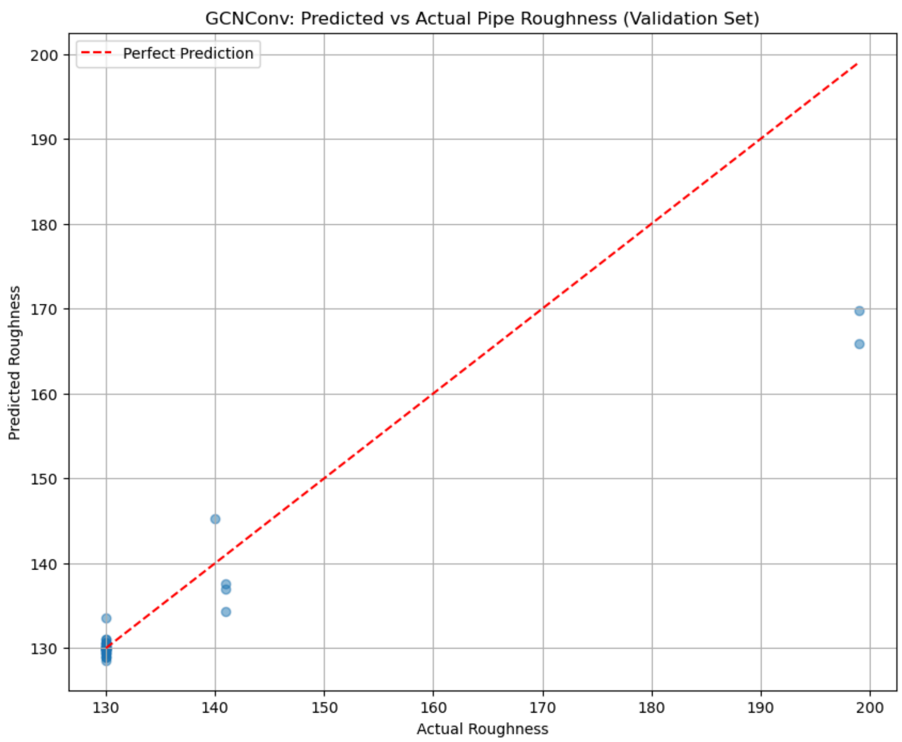

Fig. 1- GNN Results with a Non-modified Roughness Label

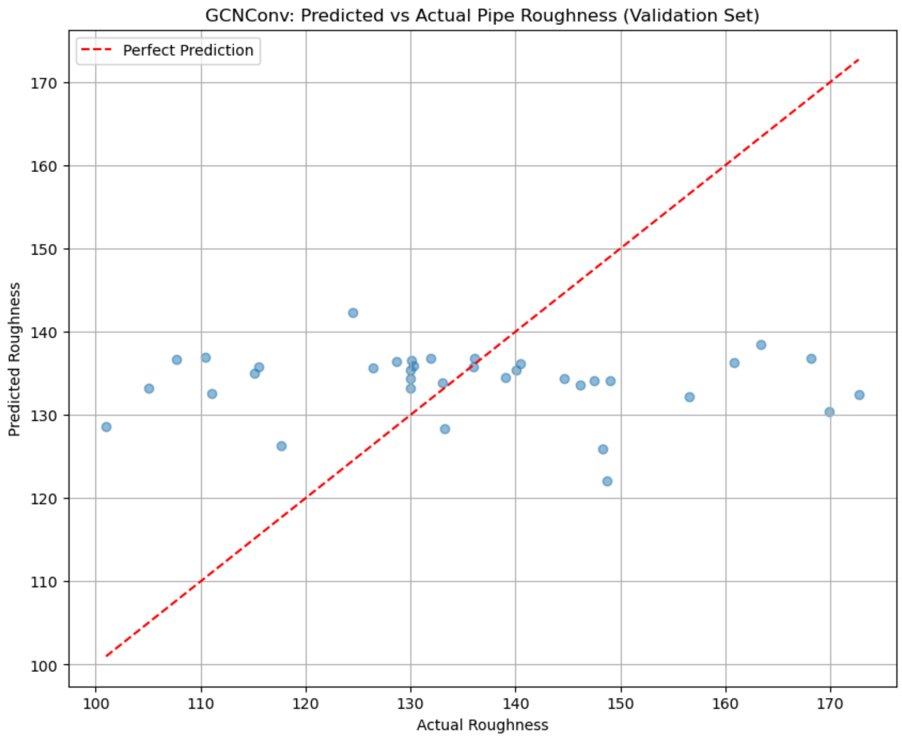

Fig. 2 - GNN Results with a Modified Roughness Label

Evaluation Metrics for Validation Set:

R^2 Score: -0.0736

MAE: 15.2530

RMSE: 18.9067

Method 2: Basic MLP with Pipe Features Only

A baseline MLP model was developed using only pipe-level attributes:

- Input Features:
  - Length
  - Diameter
  - Flowrate
  - Headloss
- Architecture:
  - Two hidden layers (64 and 32 units)
  - ReLU activations
  - Final output node for regression
- Result:
  - The model was unable to capture the variation in roughness values
  - Predicted values clustered near the global mean
  - High validation loss due to a lack of descriptive features

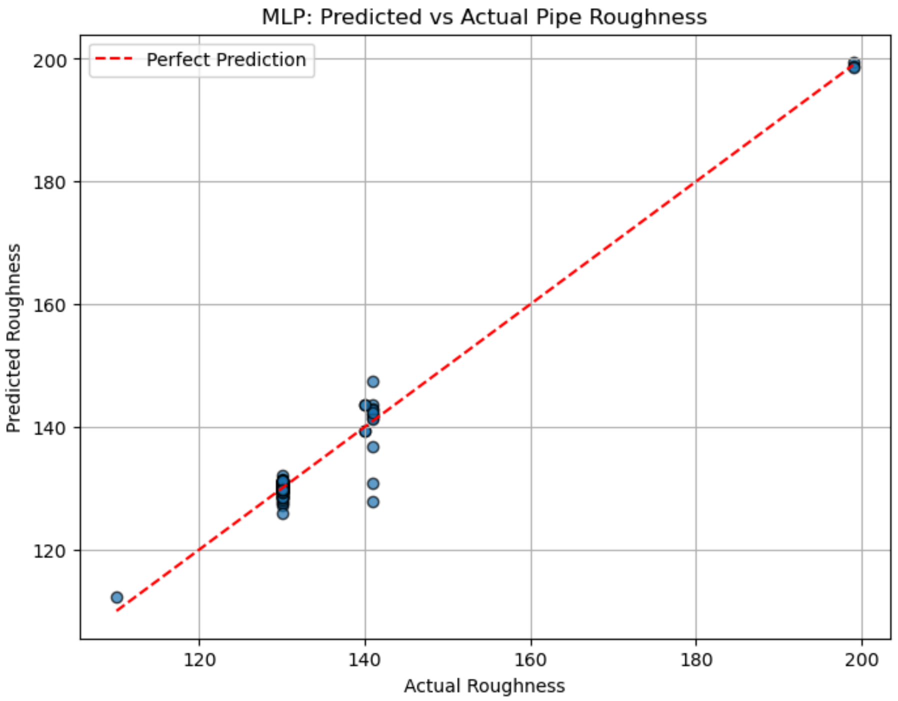

Fig. 3 - MLP Results (Pipe Features) with a Non-modified Roughness Label

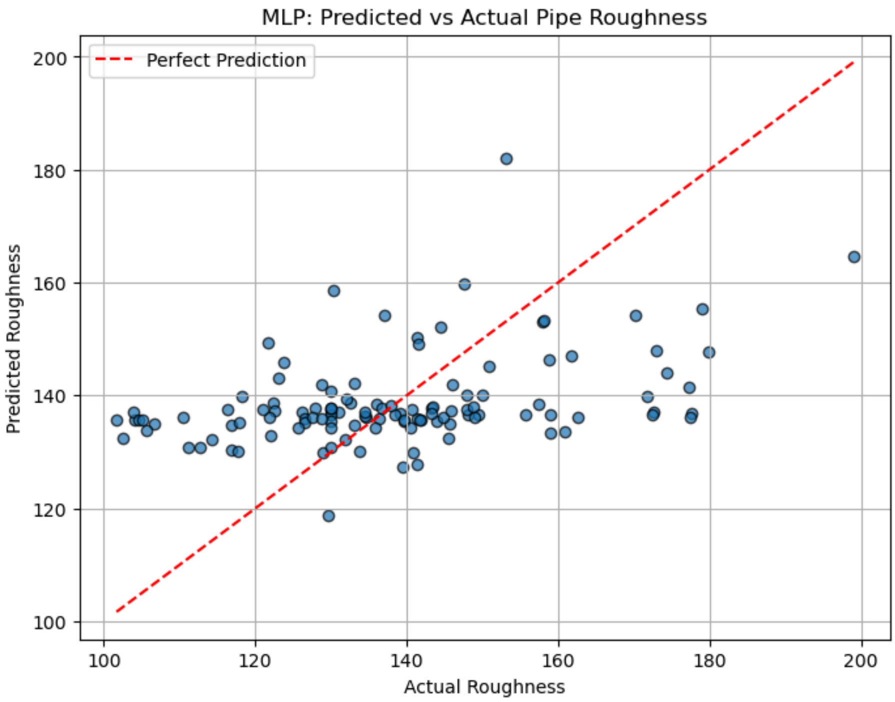

Fig. 4 - MLP Results (Pipe Features) with a Modified Roughness Label

- Evaluation Metrics for Validation Set:
- R^2 Score: 0.0634
- MAE: 15.2901
- RMSE: 19.3281

<u>Method 3: Basic MLP with Pipe and Node Features</u>

To improve upon the basic MLP, node attributes were integrated into the feature set.

- Additional Features:
  - Elevation at start and end nodes
  - Elevation difference
  - Demand at start and end nodes
  - Demand sum
- Total Input Dimensions: 10 features per pipe
- Model and Training:
  - Same MLP architecture as before
  - Train-validation split (80/20)
  - Features normalized using StandardScaler
- Result:
  - Slight improvement in prediction distribution
  - Still significant underfitting: model unable to learn the non-linear structure of roughness variation
  - Validation predictions remain clustered in a narrow range

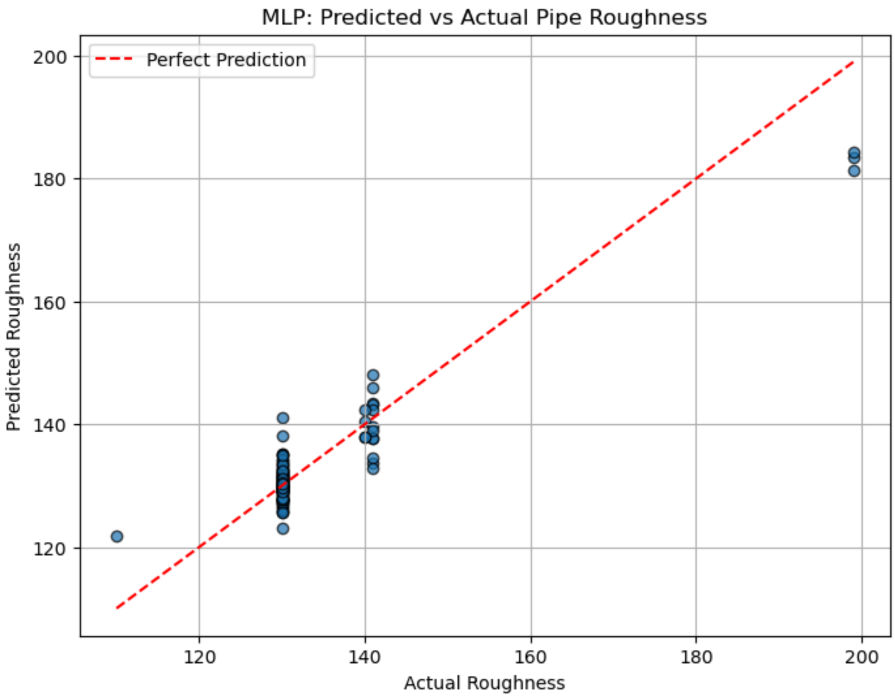

Fig. 5 - MLP Results (Pipe and Node Features) with a Non-modified Roughness Label

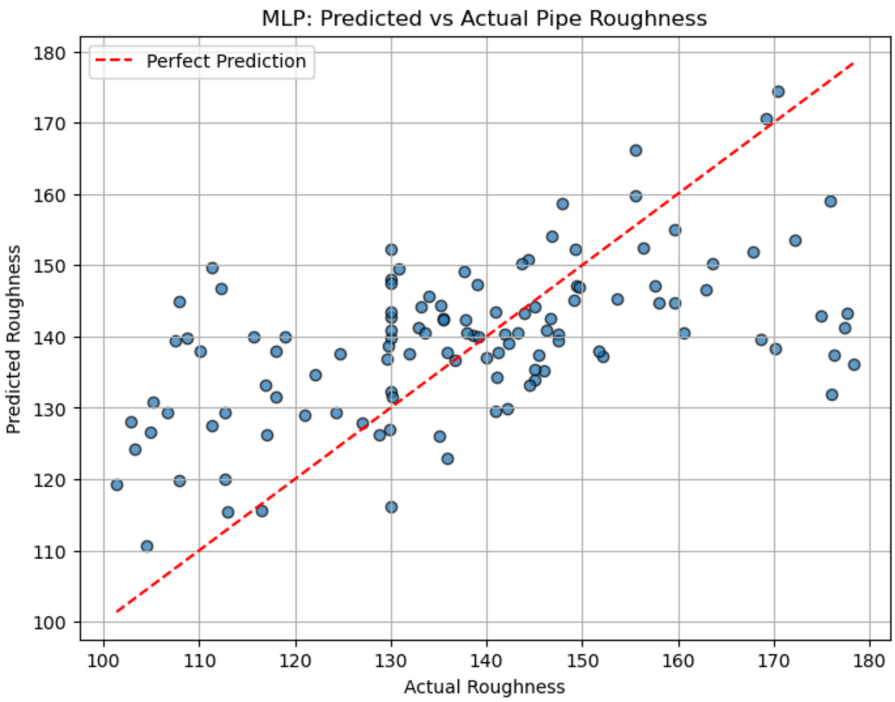

Fig. 6 - MLP Results (Pipe and Node Features) with a Modified Roughness Label

Evaluation Metrics for Validation Set:

R^2 Score: 0.1341

MAE: 13.2036

RMSE: 16.9162

5. Summary

Three approaches were implemented to predict pipe roughness:

1. GNN with flipped edge and node features
2. Basic MLP with pipe-level attributes only
3. MLP with expanded feature set including node-level attributes

Despite engineering richer features and experimenting with neural architectures, performance remained limited. Key challenges include:

- Lack of real variation in synthetic roughness labels
- Highly nonlinear or latent dependencies not captured by available features
- Small dataset size leading to underfitting

6. Next Steps

- Integrate historical maintenance records or known roughness measurements if available
- Explore graph-based message-passing with edge-conditioned networks (e.g., NNConv)
- Use ensemble models as a benchmark
- Apply physics-informed loss functions if appropriate