UMBC

Modeling-II

Team Members:
Chandra Sekhar Katipalli
Sindura Reddy Challa
Sanjana Reddy Soma

Research Questions

1. How does the fusion of HPMS and FAF datasets (2013–2022) enhance the predictive performance of highway deterioration models in estimating IRI, compared to traditional statistical and machine learning approaches using single-source data?

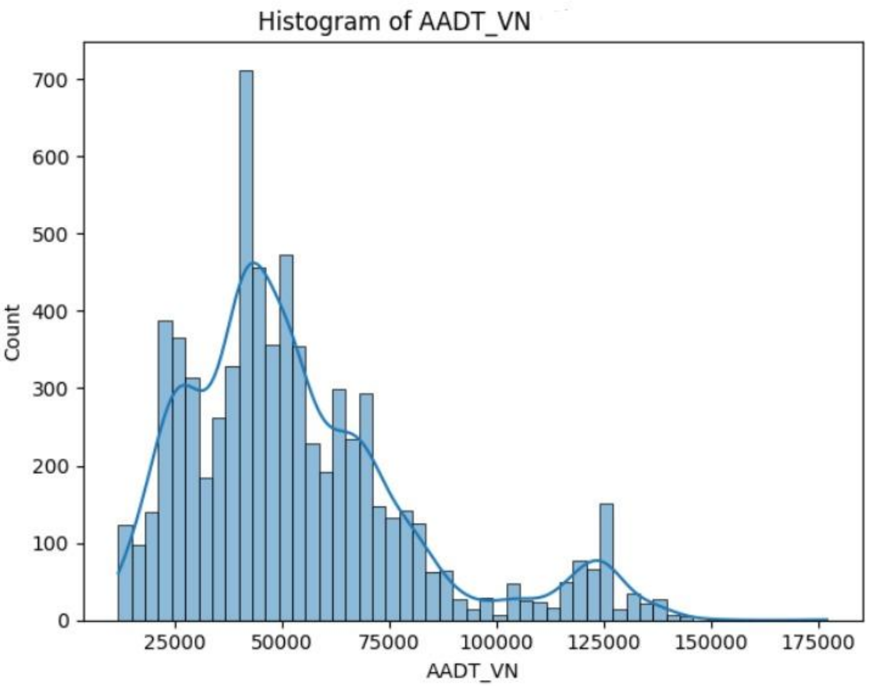
2. What are the most influential predictive features—such as traffic volume, freight load, and pavement condition indices—derived from the integrated datasets, and how do their contributions vary across different machine learning models?

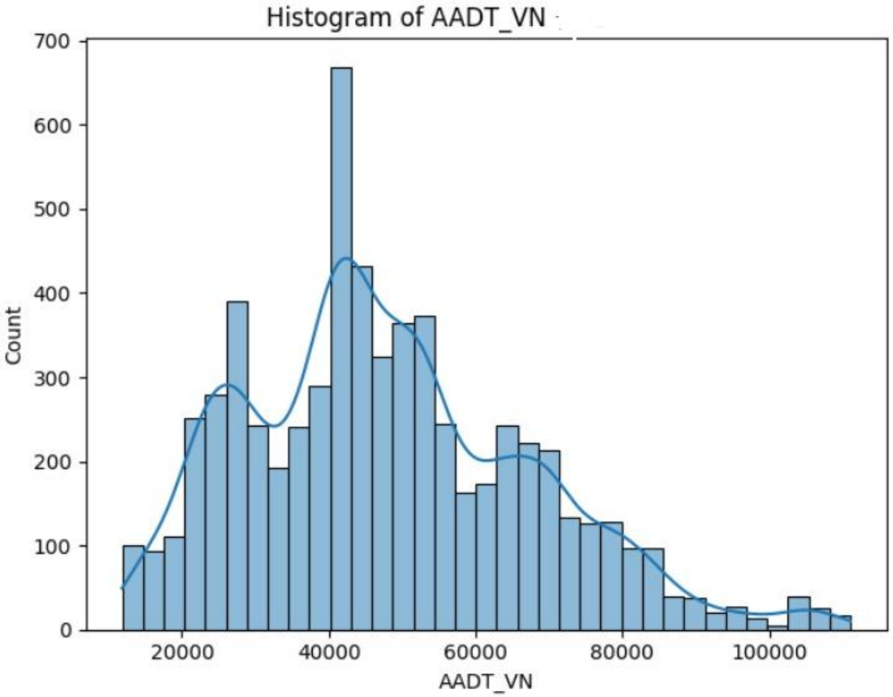
3. How effectively can the proposed predictive model, leveraging data fusion and advanced machine learning techniques, minimize forecasting errors and improve the optimization of maintenance scheduling to reduce unplanned highway repairs?

4. What is the optimal approach to forecasting IRI at different levels of granularity—both for entire highway routes (RouteID level) and for specific highway sections (0.1-mile segments)—to support more precise maintenance planning?

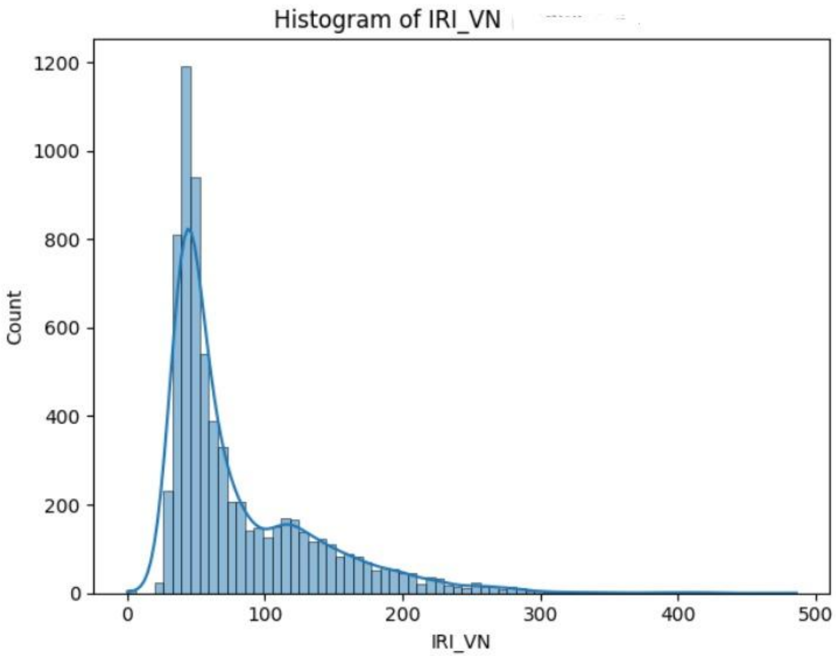
5. What might be the most effective method for visualizing and presenting findings to highway maintenance teams, like using geospatial mapping to find the roughest sections along a highway and their projected deterioration over time?

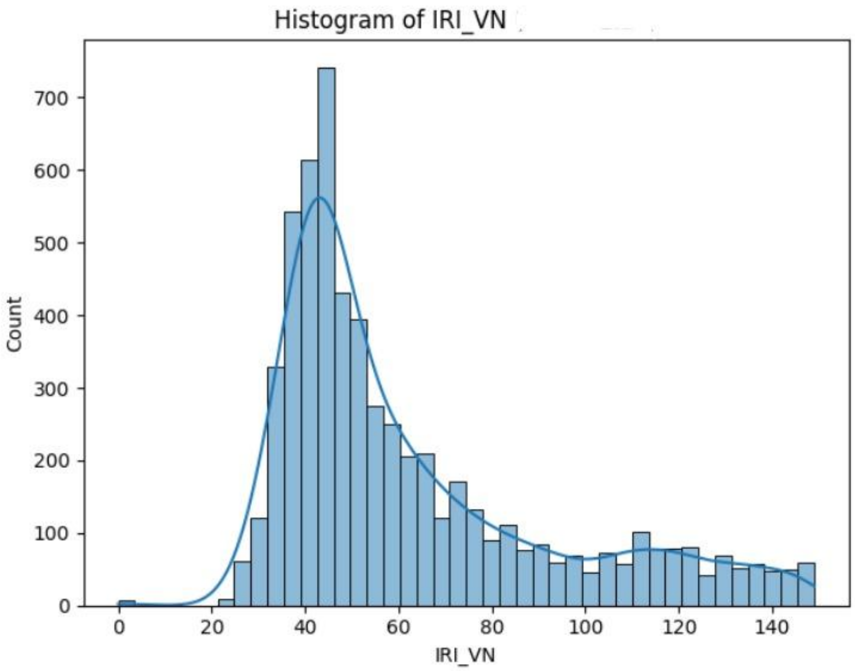
Skewness and Distribution





Skewness Graphs





Ensemble Models Built

Random Forest Regressor (Bagging-type method)

Gradient Boosting Regressor (Boosting method)

XGBoost Regressor (Boosting method)

Voting Regressor combining both RF, GB (Stacking-like ensemble)

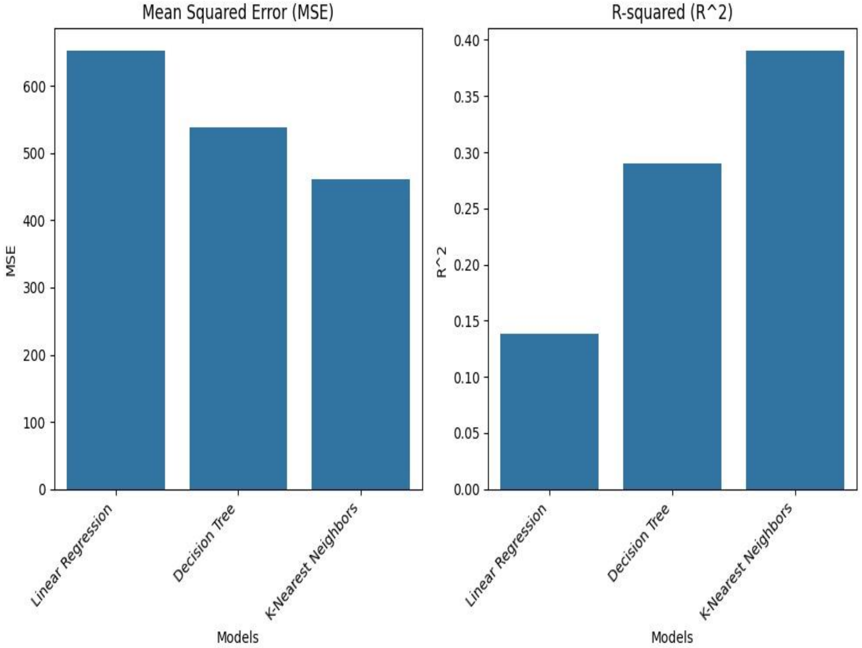
UMBC

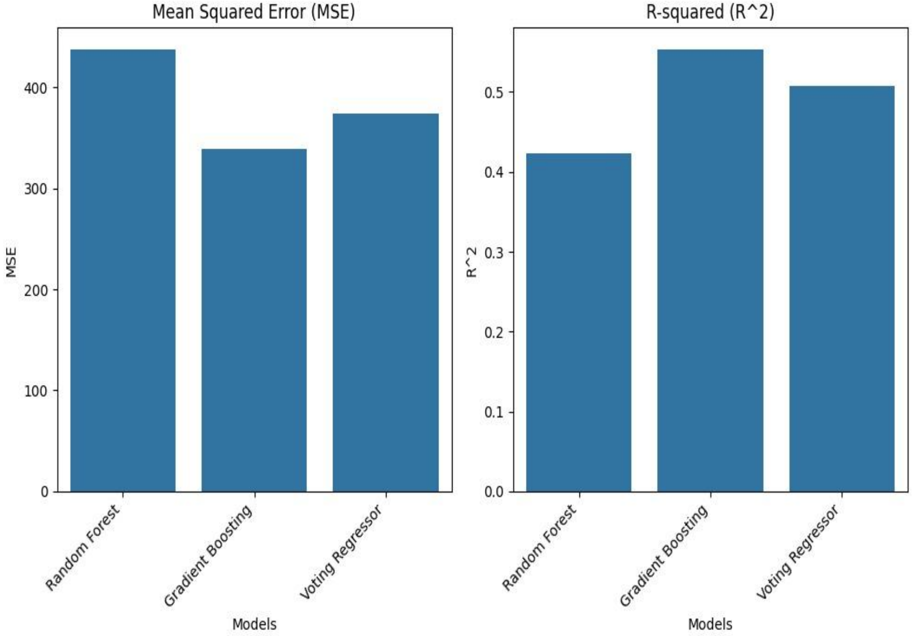
Model Evaluation (Performance Metrics)

Model	MSE	R^2
Linear Regression	652.63	0.138
Decision Tree	537.88	0.289
KNN	461.50	0.390
Random Forest	437.03	0.423
Gradient Boosting	338.69	0.553
Voting Regressor	373.44	0.507
XGBoost	330.27	0.562

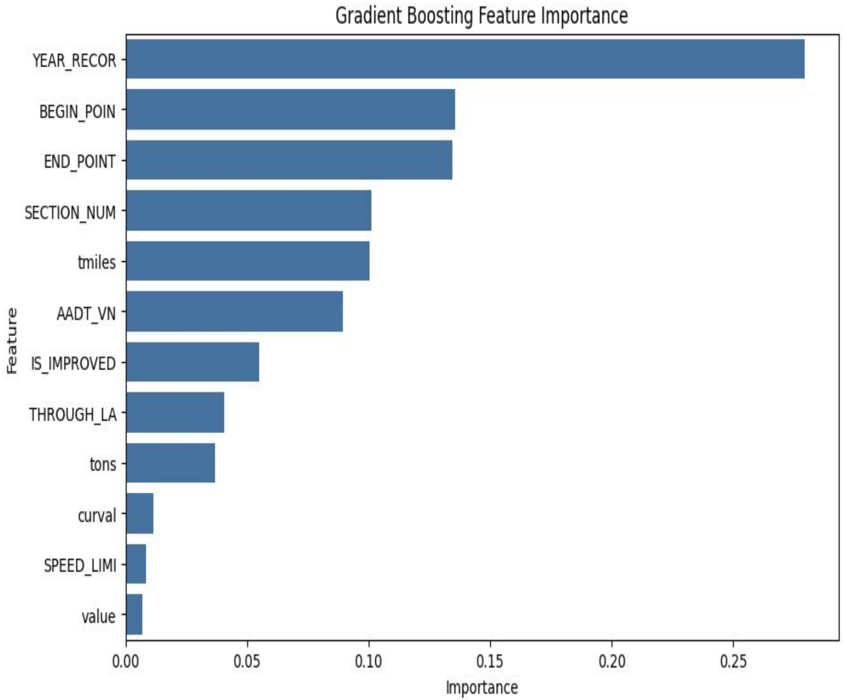
- XGBoost achieved the lowest MSE and highest R^2.
- Ensemble methods performed better than individual models.

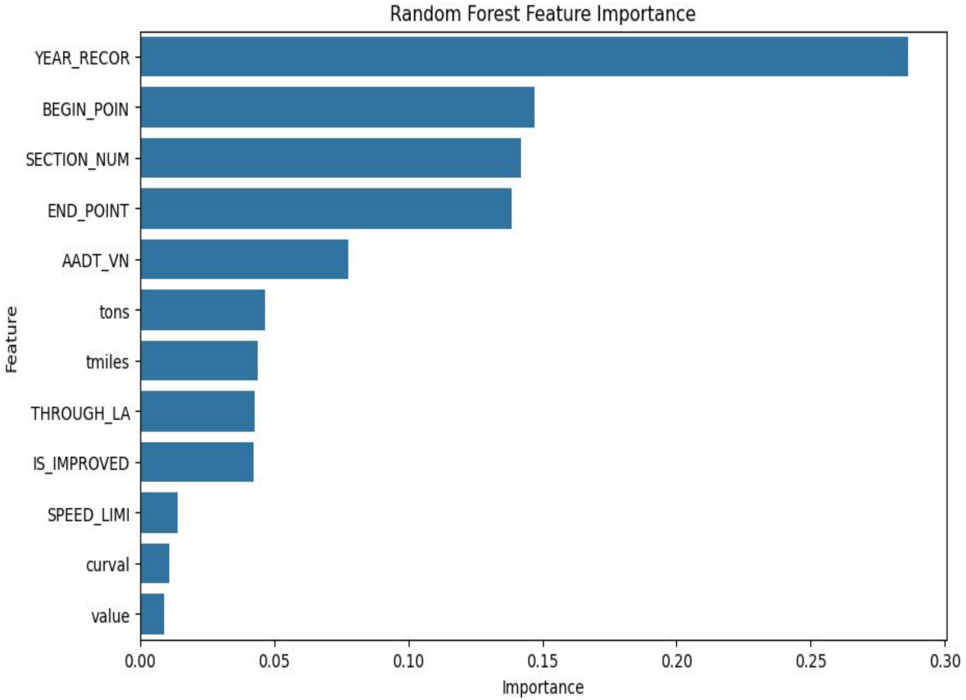
Traditional VS Ensemble





Feature Importance





Bias Variance Tradeoff

Models	Train MSE	Test MSE	Train R^2	Test R^2
Random Forest	52.96	466.34	0.904	0.3842
Gradient Boosting	119.20	383.09	0.785	0.494
Voting Regressor	67.85	357.32	0.877	0.528
XGBoost	58.08	331.35	0.895	0.562

Models perform **too well** on training data but fail on unseen test data.
Current models might not capture sequential dependencies in data
Road conditions (IRI) **change gradually** over the years.
Sequential models will consider how IRI evolved over time.

If AADT_VN (traffic volume) gradually increases over time, a sequential model can track its long-term impact on IRI_VN

Next steps

- Since the models are not performing well with more errors and less variance
- we will convert the data into sequential data and train Recurrent neural networks (RNN) for better predictions with less errors.

THANK YOU