George Mason University

AIT 614 Big Data Essentials

Selected Student Projects

Spring 2025

Prof. Dr. Duoduo (Lindi) Liao

AIT614-002 Instructor: Prof. Dr. Lindi Liao

AIT614-002 Team 8:

Traffic Incident Prediction Using Machine Learning and Deep Learning Models

Aishwarya Jonnagiri, Venkat Chunduru, Rohit Venkata Krishna Javvadi, Sri Sai Suhaas Ponnada

[Presentation video]

Abstract:

Traffic accidents are the leading reason for deaths, property destruction and system disruptions in worldwide transportation networks. Road safety improvement alongside infrastructure guidance now depends heavily on effective traffic incident prediction because vehicle numbers are growing alongside transforming seasonal road conditions. This research develops a complete data-based system for crash prediction using modern machine learning and deep learning techniques. The research obtains various datasets from FHWA InfoPave and InfoBridge that combine information about road pavement indicators and truck traffic flows accompanied by environmental factors including freezing conditions and humidity variables. The modeling solution uses AutoRegressive Integrated Moving Average (ARIMA) models for statistical baseline prediction and Long Short-Term Memory (LSTM) networks as sophisticated time-series analysis methods. The application of KMeans clustering groups high-risk regions through identification based on linked environmental and structural factors. All data-related processes run within PySpark pipelines on Databricks which process more than 200,000 records in parallel through their expandable framework. The experimental analysis demonstrates that LSTM outperforms ARIMA in terms of forecasting metrics RMSE, MAE, and R^2 as well as cluster analysis identifies spatial patterns for targeted intervention development. Big data analytics and the findings from this study created essential bases for advanced preventive crash systems while showing the usefulness of data analytics for planning smart transportation infrastructure. The paper ends by presenting insights regarding potential benefits and existing obstacles alongside future possibilities to apply this system for real-time severe crash prediction.

AIT614-002 Team 2:

Predicting Asphalt Pavement Deterioration under Climate Variability Using Big Data, Machine Learning, and Large Language models

Nidhi Janivara Somashekar, Manisha Suguru, Sreyas Kolli

Presentation video

Abstract:

The endurance of asphalt pavement is severely hampered by climate variability, which includes variations in temperature, precipitation, and freeze-thaw cycles. The intricacy of large-scale, multi-variable datasets necessary for precise deterioration forecasts is frequently not captured by traditional mechanistic models. In order to forecast pavement deterioration trends under various environmental situations, this study uses machine learning and big data techniques. To create a final dataset of more than 22,000 entries from five western U.S. states—California, Arizona, Nevada, Oregon, and Washington—we combined datasets from NASA's MERRA-2 climate archives and the FHWA InfoPave portal. To deal with outliers, skewed distributions, and missing values, extensive data preprocessing was done. By integrating traffic and climate data, feature engineering improved the forecasting ability. Distress Score, International Roughness Index (IRI), and Rut Depth are three important pavement performance measures that were accurately predicted by the final modeling technique, Gradient Boosted Trees (GBT). Despite the difficulties with the dataset, the models showed good predictive accuracy when assessed using RMSE, MAE, and R2 metrics. According to experimental studies, pavement degradation is greatly influenced by the combined effects of traffic and climate conditions. This project serves as an example of the effectiveness of integrating AI-driven reporting, machine learning, and big data analytics to support proactive, data-driven infrastructure management techniques. Future research will concentrate on adding real-time climate forecasts and broadening the geographic coverage.

AIT614-002 Team 1:

Enhancing Pavement Structural Health Assessment Using FWD-Based NDE and Machine Learning

Varsha Ponnaganti, Rohit Annepu, Harsha Raja Neathala, Anirudh Pudipeddi, Erica Mathias

Presentation video

Abstract:

This study presents a machine learning driven approach to enhance pavement structural health assessment using Falling Weight Deflectometer (FWD) based non-destructive evaluation (NDE) data. The research integrates deflection measurements, climate variables, and structural attributes. To address three core objectives: (a) proactively identifying pavement segments that are likely to fail within two to three years using supervised classification models and SHAP-based interpretation; (b) evaluating the impact of climate extremes such as freeze-thaw cycles and precipitation on pavement deterioration using regression analysis; and (c) forecasting remaining structural service life through multivariate time-series modeling with long short-term memory (LSTM) networks. Experimental results from a large multi-state dataset demonstrate that ensemble methods (e.g., Random Forest, XGBoost) and deep learning models yield accurate and interpretable predictions of pavement distress trends. SHAP analysis revealed that key contributors to deterioration include asphalt thickness, freeze index, and climatic severity. This framework enables data-informed maintenance planning and contributes to early failure detection, particularly under variable environmental conditions. The

findings offer practical tools for transportation agencies to support proactive and cost-effective infrastructure management.

AIT614-002 Team 6:

Enhancing Pavement Maintenance through Machine Learning: Predicting Freeze-Thaw Cycle Impacts

Sonal Ashok Kumar Masid, Srinija Tummalapally, Uttam Yeramalla, TarunDeep Krishnakumar

Presentation video

Abstract:

The project "Enhancing Pavement Maintenance Through Machine Learning: Predicting Impacts of Freeze-Thaw Cycles" aims particularly at countering the adverse effects of freeze-thaw cycles on pavement, a major cause of pavement degradation in the cold states of the U.S.A. By involving big data technology and other machine learning techniques, the study proposes to develop predictive models that will be able to predict the deterioration of pavement due to climatic effects, especially freeze-thaw effects. The ultimate intention of this project is to provide a data-based approach toward better and proactive maintenance for our infrastructure systems through integration of past climatic data with pavement conditions. Machine learning methods are used with Spark MLlib to analyze the factors that influence the conditions of the roads. Additionally, relevant visualizations were created to gain actionable and meaningful insights into patterns and trends. The project commenced at a very important phase, involving the collection of huge volumes of data, some cleaning and preparation of the dataset, and performing first exploratory analyses to identify trends that are presently being worked upon. The next steps involve extracting relevant features from the data, and evaluating several machine learning algorithms, in order to determine the most accurate and reliable predictive models that can be implemented later for setting requirements of planning and maintenance of the infrastructure, thereby reducing long-term costs and improving roadway safety in locations susceptible to extreme weather conditions.

AIT614-002 Team 5:

Predictive Analysis of Pavement Friction and Road Deterioration Using Big Data

Venkata Lakshmi Parimala Pasupuleti, Kavya Annam, Kiran Dabbiril, Haripriya Varanasi, Sumanth Naidu

Presentation video

Abstract:

Pavement deterioration in snow-prone areas is a major concern for road safety and maintenance in these communities. Snow, ice and freeze-thaw cycles dramatically amplify distress to road pavement, and lead to decreases in friction and skid resistance, all of which play a significant role in vehicle control in winter weather. Deterioration and decreased friction allow for an increase in accident potential and road closures. This project has the sole focus of developing a predictive model that can ultimately predict pavement conditions by leveraging big-data analytic and machinelearning techniques. A massive compilation of data is collected,

stored, manipulated, and analyzed including historical pavement data, the continuous collection of skid resistance data, aggregated traffic data, and environmental weather and climate condition data (for example, snowfall, temperatures and snow/ice accumulation, plus average daily temperature change). The accepted analyses for this study; use Apache Spark & PySpark. To make accurate predictions, we use machine learning models like XGBoost and Random Forest. These models learn from the data to predict when and where pavement deterioration likely occurs. Our system is hosted on Databricks, and as a result, it is scalable and can monitor roads in various geographic locations simultaneously and in real time. The early warning and forecasting capabilities of this model can enable transportation departments to make more informed decisions when planning repairs, lowering risk for accidents, and prolonging the life of road infrastructure, particularly in areas with harsh winter climates.

AIT614-002 Team 7:

The Interplay of Traffic Loads and Environmental Factors in Asphalt Pavement Degradation

Bala Ashrith Reddy Kommareddy, Mohan Silambarasu Elangkumaran, Surthesh Velu Samy

Presentation video

Abstract:

Asphalt pavements are of national transportation significance but are subject to progressive stress from multiple axle loading and environmental degradation patterns. This study investigates the interaction between mechanical stress due to traffic and climatic degradation patterns in different U.S. states. We utilize a large data, multi-source database with over 15 million entries, including annual traffic volume data, Equivalent Single Axle Load (ESAL) patterns, longitudinal cracking data, Mean Roughness Index (MRI), precipitation totals, and freeze-thaw events. This research aims to determine important predictive variables for pavement distress and provide data-driven results in an effort to enable infrastructure maintenance and design. In order to process and analyze such high-speed, large-scale data, we used a scalable data pipeline with PySpark and ran all the processing and modeling operations on the Databricks cloud platform. Data preprocessing included removal of duplicates, type casting, missing value handling, and noise removal by attribute elimination of unwanted attributes. Descriptive analysis and correlation heatmaps were used to explore statistical correlations, finding regular negative correlations between truck load measures (ESAL, AADTT) and surface quality measures (MRI), in support of our main hypothesis that increased loading is associated with more rapid deterioration. Environmental analysis showed that precipitation and freeze-thaw had regionally distinct but very large effects, with unexpected outcomes such as Florida and Maryland experiencing extreme cracking with low freeze-thaw activity. This is suggestive of interactions between material performance or maintenance practice and climate. Scatter plots, boxplots, and pair plots further reflected non-linear trends, skewed distributions, and clustering behaviors that reduced the predictive ability of linear models. Our findings highlight the requirement for more robust analysis methods in the form of non-linear machine learning models like Random Forests for more precise prediction of pavement deterioration. These models can effectively capture the complex, multi-dimensioned associations that are inherent in civil infrastructure data. This study contributes both a methodological framework and empirical insights for transportation engineers, policy makers, and researchers concerned with optimizing pavement management systems in light of evolving climate and traffic patterns.

AIT614-002 Team 4:

Evaluating the Influence of Cracking Metrics and Pavement Types on Performance Degradation

Deekshitha Reddy Kalluru, Medha Chada, Bharath Vardhan Reddy Ravula, Rishika Reddy Baddam, Aditya Indurkhya

Presentation video

Abstract*:*

The present work investigates applications of machine learning methods in pavement management decisions. Based on sidewalk condition and climate data from LTPP InfoPave, present work addresses three core tasks: classification of pavements, predictive modeling of cracking and rutting failures, and clustering of segments of pavements into strategies of maintenance. Pavement classification was achieved using Naive Bayes and Logistic regression, of which Logistic Regression performed better. Prediction was performed using, among other methods, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression, of which Logistic Regression performed best in terms of ROC-AUC. Lastly, unsupervised clustering from KMeans and DBSCAN was applied in cluster detection of maintenance strategies, where KMeans produced understandable clusters. The results indicate potential of machine learning in generating actionable knowledge owing to its vulnerability to pitfalls such as data complexity and class imbalance.

AIT614-002 Team 3:

Leveraging Big Data and Machine Learning for Predictive Pavement Condition Analysis

Rushika Arvapalli, Sushanth Buddhala, Sumanth Veluvolu, Sainath Kammiti

Presentation video

Abstract:

Transportation engineering needs proper pavement management as an essential element of its core practices. Insurance road conditions experience difficulties because of extending transportation systems and shifting environmental conditions along with traditional visual check protocols. The study creates a predictive data system through the combination of Big Data analysis mechanisms with Machine Learning methods aimed at predicting pavement deterioration states. Various predictive models consisting of clustering models together with ensemble regressors and classification-based life prediction models were developed. GBT proved to be the most effective model for predictive purposes. Research results indicate that units of predictive modeling with clustering allow predictive maintenance planning which decreases expenses while enhancing road network protection.

AIT614-003 Instructor: Prof. Dr. Eddy Zhang

AIT614-003 Team 6:

Impact of Bridge Design Materials and Climate Exposure on Deterioration and Condition

Yaseen Trombati, Neha Rao, Bhargavi Jammi, Kashyap Kandibanda, Pragnya Reddy Vanga

Presentation video

Abstract:

Bridges are a critical component of transportation infrastructure, and their long-term performance is shaped by both the materials used in construction and the environmental conditions they endure. This study investigates how bridge materials—such as concrete, steel, and wood—perform across varying climate zones, including cold, hot, rainy, and windy regions. Using data from the FHWA Long-Term Bridge Performance (LTBP) program, we examined key deterioration factors such as snowfall, freeze-thaw cycles, and bridge age. The analysis involved extensive data cleaning, feature engineering, and exploratory assessment to identify patterns in bridge degradation. Predictive models were developed to estimate bridge condition and classify structural health based on environmental and structural features. Visualizations, including geospatial maps and comparative charts, were used to illustrate how material performance varies across regions. Results indicate that concrete and steel tend to offer greater durability in colder and moderate climates, whereas wood deteriorates more rapidly, particularly under high-moisture or freeze-thaw conditions. These findings provide actionable insights for improving material selection, maintenance strategies, and design practices, ultimately contributing to the development of safer and more resilient bridge infrastructure.

AIT614-003 Team 8:

Decoding Multi-Domain Data: An Analysis of Weather, Experiment, and Site

Raja Ruthvik Shetty, Phani Satya Sai Pamarthi, Dhanushi Panga, Suraj Bharadwaj Mandava Venkata, Yatish Chandra Nalla

Presentation video

Abstract:

The project entails multi-domain data integration and analysis including weather, experimental pavement data, and site conditions for pavement performance prediction and analysis under various environmental and operational conditions. We built scalable models for pavement condition classification, deterioration forecasting, and anomaly detection using Apache Spark and PySpark for distributed data processing and MLlib for machine learning. The pipeline employed rigorous cleaning, normalization, feature engineering, and techniques like clustering and outlier detection. Employee burnout reviews were also fed through an experimental NLP module in order to combine human-centric information with infrastructure data. Streamlit interactive Streamlit dashboards and static Matplotlib and Seaborn plots are the final deliverables to enable data-driven decision-making for infrastructure management.

AIT614-003 Team 3:

Enhancing Bridge Safety in Georgia: A Predictive Approach Using Machine Learning and Big Data

Shashank Reddy Mallepally, Pravalika Godala, Pavan Chandu Annepu, Shanthi Srujan Pratti, Mahender Tanniru

Presentation video

Abstract:

Bridge structural integrity is vital to economic mobility and public safety. With the reality of aging infrastructure and increased climate volatility, traditional inspection methods no longer give assurances of timely maintenance. This project introduces a data-driven predictive maintenance and risk management system for the bridges of Georgia using big data and machine learning technologies. Drawing on the Georgia InfoBridge dataset, we conducted extensive exploratory data analysis and applied clustering and classification models to uncover hidden patterns, evaluate condition trends, and forecast possible degradation. Our approach enables proactive infrastructure management through enabling authorities to determine priority bridges at risk and allocate resources optimally. The combination of data engineering and machine learning techniques in this study demonstrates the transformative potential of intelligent systems in tracking civil infrastructure.

AIT614-003 Team 7:

Impact of Daily Traffic Variability on Long-Term Structural Deterioration of Bridges in High-Congestion Areas

Devi Naga Akshitha Kondru, Harshitha Singannagari, Sai Sahitya Chebrolu, Sandhya Kanduri, Swija Reddy Gaddam

Presentation video

Abstract:

Bridges serve as critical elements of modern transportation networks, facilitating economic and social connectivity across regions. However, the increasing variability in daily traffic patterns, especially in urban areas with chronic congestion, presents a significant threat to the structural health of these infrastructures. Continuous stress from fluctuating traffic volumes accelerates deterioration, leading to increased maintenance demands, service disruptions, and safety risks. This study investigates how daily traffic load variations influence the long-term structural degradation of bridges in high-congestion zones. Leveraging the Selected Bridges dataset, which encompasses over 100,000 bridge records with key indicators such as average daily traffic, material types, and condition ratings, the project applies machine learning and statistical methods to identify critical deterioration patterns. Predictive models, including Random Forest Regression and XGBoost, are used to estimate structural decline, with the goal of enabling proactive infrastructure maintenance. The findings are visualized through dashboards using Power BI and Tableau, while geospatial patterns are mapped using QGIS and GeoPandas to highlight high-risk congestion clusters. These insights aim to guide transportation agencies in optimizing maintenance schedules and resource allocation. Ultimately, the study contributes to the development of a scalable, data-driven framework for monitoring and managing the health of critical bridge infrastructure.