Road Condition Prediction Using Machine Learning

Submitted to: Prof. Ana Marbut

Course: Data Analytics and Internet of Things (AAI-530-02)

Date: Feb 24, 2025

Abstract

This project explores the integration of machine learning and IoT in road condition prediction. Using real-world vehicular sensor data, we trained multiple ML models (Random Forest, LSTM, and GRU) to classify road types based on vibration and sensor readings. Our IoT system design incorporates edge and cloud computing for real-time processing. The results demonstrate the effectiveness of deep learning in road classification and suggest potential applications for smart infrastructure.

Introduction

This research explores the implementation of machine learning and IoT to enhance vehicular safety by classifying road conditions based on sensor data. The study aims to provide a predictive framework that can assist in smart city infrastructure, vehicle optimizations, and autonomous driving systems.

Uneven road conditions significantly impact vehicle safety and efficiency. Traditional approaches to road monitoring rely on manual surveys, which are costly and inefficient. This project proposes an automated sensor-based machine learning approach to classify road conditions and enhance road safety measures.

IoT System Design

Our IoT system is built on edge computing principles, leveraging onboard vehicle sensors to classify road conditions in real time. The dataset used for model training primarily consists of sensor readings recorded directly from the vehicle, emphasizing local processing over cloud dependency.

Current IoT System Setup (Based on Dataset)

The existing system focuses on sensor data collection and local storage, with minimal external communication. The dataset includes:

- **Sensors: GPS:** accelerometer, gyroscope, magnetometer, temperature sensor, and HD camera.
- **Edge Computing:** Local data processing and storage on an SD card, minimizing reliance on cloud computing.

Suggested IoT Enhancements

While the dataset primarily reflects edge computing, our proposed system design expands the architecture to integrate cloud and communication capabilities, making it more practical for real-world deployment. The enhancements include:

- **Cloud Integration:** AWS IoT Core (or similar platforms) for centralized analytics, remote monitoring, and predictive insights.
- **Vehicle-to-Vehicle (V2V) Communication**: Enabling direct communication between vehicles using MQTT messaging to share hazard alerts (e.g., icy roads, potholes).
- **Hazard Detection & Predictive Maintenance:** AI-based road condition analysis in the cloud to assist smart city infrastructure and transportation systems.

Key Takeaways

- **Our dataset is edge-based,** but our **proposed system suggests cloud integration for scalability**.
- **The additional components in our IoT diagram** represent f**uture enhancements** rather than existing functionalities.

- **This system can significantly impact smart transportation**, providing **real-time insights for both drivers and city planners**.

System Diagram

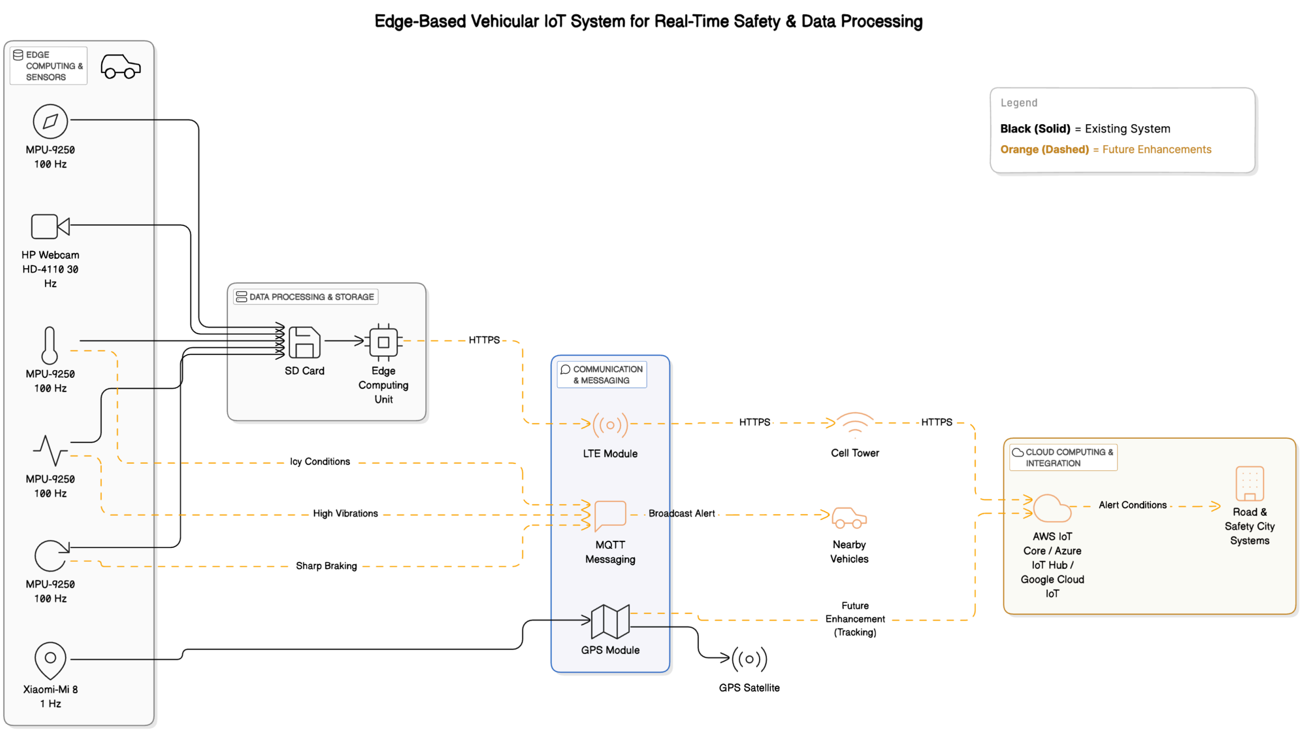


Figure 1

IoT System Architecture for Road Condition Prediction

This diagram illustrates the IoT system's architecture, integrating onboard sensors, edge computing, and cloud analytics. The system captures real-time vehicular data using accelerometers, gyroscopes, GPS, and cameras, processes it locally for rapid insights, and transmits aggregated data to cloud services for further analysis and predictive modeling. Future enhancements include vehicle-to-vehicle (V2V) communication and automated hazard detection.

Machine Learning Models

Model Selection

- **Random Forest (RF):** Traditional classification model used for baseline comparisons.
- **LSTM:** A recurrent neural network capable of handling sequential sensor data.
- **GRU:** A simplified RNN variant with improved computational efficiency.

Data Preprocessing

- **Data Size:** 1,080,905 rows, 81 features**.**
- **Feature Engineering**: Speed, acceleration, vibration, terrain type.
- **Handling Missing Values and Scaling.**

Model Training & Evaluation

Model Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
RF	74%	73%	74%	73%
LSTM	90.5%	90.7%	90.3%	90%
GRU	96%	96.8%	96.1%	96.5%

Model Performance Summary (LSTM vs. GRU vs. RF)

After training and evaluating three different models: LSTM, GRU, and Random Forest; our findings indicate GRU outperforms LSTM in classification accuracy, achieving 96% vs. LSTM's 90%. While LSTM performed well, it exhibited higher validation loss and required more epochs for convergence. The Random Forest model provided a baseline but was less effective than deep learning methods, achieving 74% accuracy. Our Tableau dashboard visually compares these results, highlighting GRU's advantage in faster convergence and generalization across road types.

Temporal Fusion Transformer (TFT) Model Attempt

The Temporal Fusion Transformer (TFT) is a deep learning model specifically designed for multi-horizon time-series forecasting. Given that our dataset contains time-dependent sensor readings, TFT was explored as a potential model to improve predictive accuracy beyond standard LSTM and GRU architectures.

Why TFT?

TFT leverages self-attention mechanisms and gated residual networks (GRN) to enhance interpretability and adapt to various temporal patterns. Unlike LSTMs or GRUs, which process sequences sequentially, TFT captures long-range dependencies more efficiently.

Challenges & Observations

Despite TFT's advantages, training the model for our dataset presented several challenges:

- **High Computational Demand** – TFT required significantly more resources compared to LSTM and GRU.

- **Overfitting Risks** – Due to our dataset size and structure, TFT showed signs of overfitting despite regularization.
- **Complex Hyperparameter Tuning** – Finding optimal learning rates and dropout values was difficult given time constraints.
- **Sensitive to Data Imbalance** – The model struggled with underrepresented classes (e.g., dirt road conditions).

Key Takeaways

While TFT remains a promising candidate for future work, our analysis showed that LSTM and GRU were more practical given our dataset's characteristics and available computational resources. Future iterations could refine data preprocessing, hyperparameter tuning, and feature selection to improve TFT's performance.

Conclusion

TFT was attempted but not included in the final model comparison due to training complexity and computational constraints. However, it remains a potential avenue for future research in real-time vehicular IoT applications.

Results & Analysis

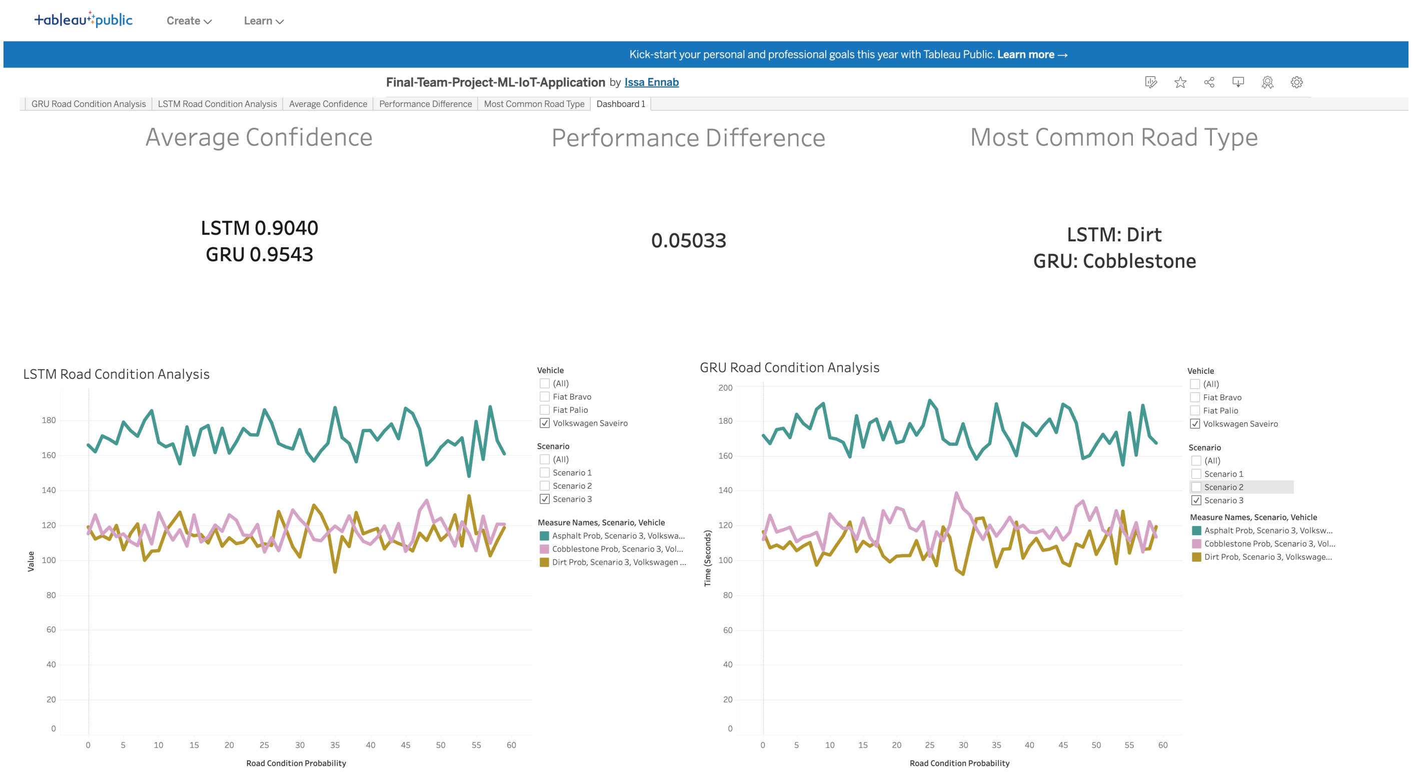
Confusion Matrices

- LSTM and GRU models achieved high classification accuracy across road types.
- The confusion matrix for GRU showed minimal misclassifications, confirming its superior performance.

Tableau Dashboard Insights

- Road condition trends were visualized using **time-series analysis**.

- Comparative graphs highlight the differences in classification accuracy among models.



Discussion

Implications

- **Smart Infrastructure:** The proposed system can aid in proactive road maintenance.

- **Autonomous Vehicles:** Real-time road classification enhances vehicle safety and decision-making.

Figure 2

Tableau Dashboard: Road Condition Predictions and Model Comparisons

This Tableau dashboard presents the visualization of road condition classification using machine learning models (LSTM, GRU, and Random Forest). It includes real-time probability distributions, model performance metrics, and interactive filters for vehicle scenarios.

Challenges & Limitations

- Data Imbalance: Unequal distribution of road types in the dataset.
- Computational Cost: Training deep learning models requires significant resources.

Future Work

Potential Enhancements

- **Improved Feature Engineering:** Incorporating additional vehicular parameters such as brake pressure and suspension dynamics.
- **Model Optimization:** Exploring hybrid models combining CNNs with RNN architectures.
- **Expanded Dataset:** Collecting data across diverse geographical locations for improved generalization.
- Future enhancements include integrating more advanced AI-driven road classification models for larger datasets.

Real-World Applications

- **Smart City Integration:** Real-time road analytics for city planners.
- **Automotive Industry:** Data-driven tire and suspension recommendations based on road conditions.
- **V2V Communication:** Enhancing real-time road safety alerts between connected vehicles.

Conclusion

This study successfully demonstrated the potential of machine learning in road condition classification. The GRU model exhibited superior performance, making it suitable for real-time applications. Future iterations should focus on expanding dataset diversity and refining feature selection for enhanced accuracy and deployment feasibility.

References

- Anaconda Inc. (2024). *Jupyter Notebook* [Computer software]. https://jupyter.org
- Chollet, F. (2015). *Keras* [Computer software]. https://keras.io
- Google Research. (2024). *TensorFlow* [Computer software]. https://www.tensorflow.org
- Menegazzo, J. (2024). *PVS: Passive Vehicular Sensors Datasets* [Data set]. Kaggle. https://www.kaggle.com/datasets/jefmenegazzo/pvs-passive-vehicular-sensors-datasets
- Menegazzo, J. (2024). *PVS Data Exploration* [Data set analysis]. Kaggle. https://www.kaggle.com/code/jefmenegazzo/pvs-data-exploration
- Microsoft Corporation. (2024). *Visual Studio Code* [Computer software]. https://code.visualstudio.com
- OpenAI. (2024). *ChatGPT* (Feb 2025 version) [Large language model]. https://openai.com
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32. https://pytorch.org

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830. https://scikit-learn.org

- Tableau Software. (2024). *Tableau Desktop* [Computer software]. https://www.tableau.com

- The Google Colaboratory Team. (2024). *Google Colab* [Cloud-based notebook environment]. https://colab.research.google.com