Evaluation of Hybrid Machine Learning Models for Risky Driving Behavior Classification: A Comparative Study Using RNN-AdaBoost, GANs and XGBoost

Elena Theodoraki

Transportation Engineer, Research Associate

Together with:
Thodoris Garefalakis, Paraskevi Koliou, George Yannis

Department of Transportation Planning and Engineering
National Technical University of Athens

ICTR 2025

12th International Congress on Transportation Research

16-18 October 2025, Thessaloniki, Greece

HIT- HELLENIC INSTITUTE OF TRANSPORT

HELLENIC INSTITUTE OF TRANSPORTATION ENGINEERS

Introduction

- Road Safety is a global challenge, with 1.19 million people dying every year according to WHO.

- Most crashes, up to 95%, are caused by driver behavior, like speeding, distraction, fatigue, or impairment.

- Advancements in Intelligent Transportation Systems (ITS) and Machine Learning are powerful tools for real-time driver behavior analysis and accident risk prediction.

- Most models lack interpretability and fail on imbalanced data, making them unreliable for safety.

Research Gaps & Objectives

- **Research Gaps**:
  - Most Machine Learning models act as black boxes
  - Driving datasets are highly imbalanced, with risky driving events being rare
  - Limited validation across different countries using real-world data

- **Objectives**:
  - Evaluate 3 hybrid ML models **RNN-AdaBoost, XGBoost,GANs**
  - Apply **Conditional GANs** (cGANs) for data augmentation
  - Use **SHAP** for interpretability and feature analysis
  - Test generalizability on **Belgium & UK datasets**

Experimental Design

Naturalistic driving experiment:

- Belgium: 43 drivers → over 7,000 trips
- UK: 26 drivers → over 8,000 trips
- Method: On road naturalistic driving study
- Duration: 4 months

The naturalistic experimental design has been subdivided into four consecutive phases:

- **Phase 1:** Basiline phase
- **Phase 2:** Real-time warning using ADAS
- **Phase 3:** Driver feedback through mobile app
- **Phase 4:** Gamification features to promote safer driving

Methodology

- Data was segmented into 30-second intervals and categorized into three safety levels using the Safety Tolerance Zone concept, based on headway distance and speed thresholds:
  - normal (low risk)
  - dangerous (moderate risk)
  - avoidable accident (high risk)

- Hybrid Machine Learning Models:
  1. RNN-AdaBoost: combines temporal sequence modeling with boosting
  2. XGBoost-RF: powerful ensemble approach
  3. GANs: used for data augmentation through synthetic samples of rare risky events

Multi-class Classification

- A major challenge was the imbalance of real-world driving data, to address this the SMOTE technique was applied to generate synthetic samples for rare risky behaviors

- Feature selection process:
  - Variance Threshold → removed low-information variables
  - Mutual Information → kept most relevant predictors
  - Normalization (Min-Max scaling) for consistent input ranges

- Selected Features:

Belgium	UK
GPS_distances_sum– Total distance traveled by the vehicle	GPS_distances_sum– Total distance traveled by the vehicle
GPS_spd_mean– Average speed of the vehicle during the trip	GPS_spd_mean– Average speed of the vehicle during the trip
ME_Car_speed_mean– Average speed of the vehicle	ME_Car_speed_mean– Average speed of the vehicle
DEM_evt_hb_lvl_H_mean– Mean level of harsh braking events recorded during the trip	IBI_value_mean– Mean time interval between successive heart beats

Evaluation & Interpretability

- **Evaluation Metrics**:
  1. Accuracy → overall % of correct classifications
  2. Precision → % of predicted risky behaviors that were actually risky
  3. Recall → % of actual risky behaviors correctly detected
  4. F1-score → balance between Precision and Recall
  5. False Positive Rate (FPR) → % of safe driving wrongly classified as risky

- **Interpretability**:
  Used SHAP to explain feature contributions, increasing transparesy and trust in predictions.

Results (1/2)

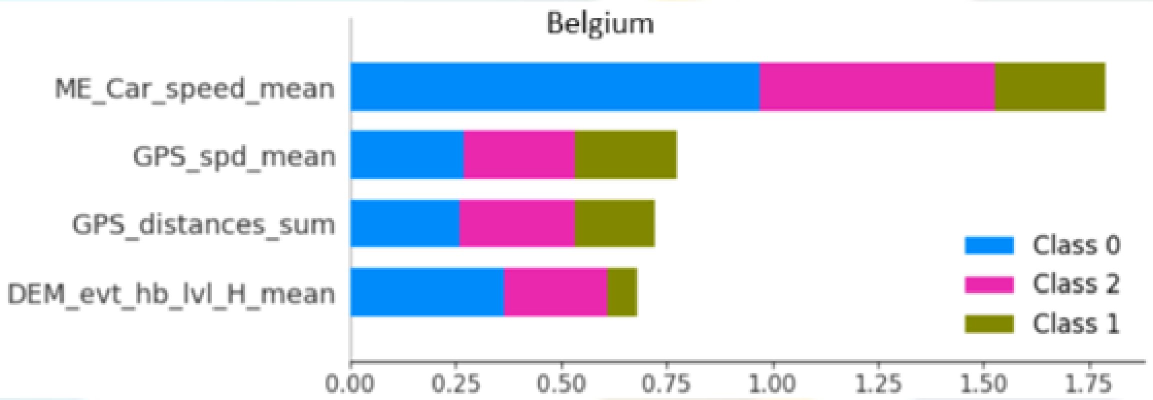
- Comparison of classification model evaluation metrics for Belgium and UK

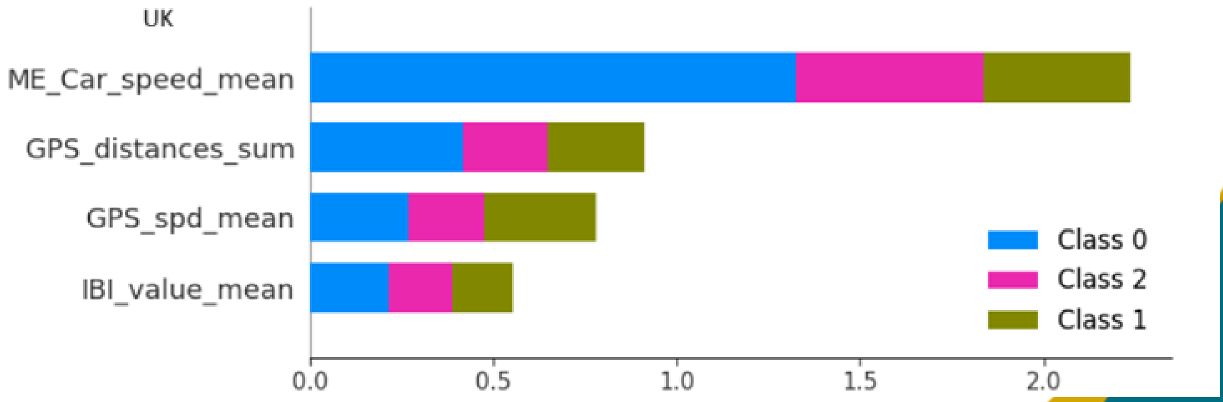
Dataset	Model	Accuracy	Precision	Recall	FPR	F1-score
	XGBOOST & RF	93%	93%	93%	7.4%	93%
Belgium	RNN & AdaBoost	83%	82%	83%	14.7%	82%
	GANS	76%	64%	76%	23.2%	66%
	XGBOOST & RF	92%	92%	92%	9.8%	91%
UK	RNN & AdaBoost	85%	84%	85%	10.3%	84%
	GANS	79%	80%	79%	21%	74%

  - XGBoost-Rf: Delivered the best results in both datasets, with highest accuracy and lowest FPR
  - RNN-AdaBoost: Lower performance with higher FPR, more prone to misclassification
  - GANs: exhibits the lowest performance, lowest accuracy and F1-score and highest FPR, indicating a weaker ability to correctly classify instances
- Cross-Country Insights:
- Belgium: Risk mostly explained by speed and harsh braking
- UK: Speed remains dominant, heartbeat intervals (IBI) play stronger role
- In both datasets Trip Distance has a moderate but consistent effect

Results (2/2)

- **Top Influential Features:**
  1. Speed (ME_Car_speed_mean, GPS_spd_mean): Strongest driver of predictions.
  2. Distance (GPS_distances_sum): Distinguishes normal vs risky trips.
  3. Harsh braking (DEM_evt_hb_lvl_H_mean): Key risk factor in Belgium.
  4. Heartbeat intervals (IBI): Stress-related influence, stronger in UK.





- **Key Takeaways:**
  - Speed Metrics→ strongest and most consistent predictors of risky driving.
  - Physiological signals (e.g., heart rate/IBI) → add interpretability and enable personalized risk assessment.

Conclusions & Future Work

- **Conclusions**:
  - XGBoost-RF proved the most accurate and robust model across both datasets.
  - SHAP analysis increased interpretability, showing speed as the dominant risk factor, with braking and physiological signals providing added insights.
  - GANs showed limited performance as classifiers but remain promising for data augmentation to improve class balance.

- **Future Work**:
  - Expand datasets with more drivers and diverse conditions to improve reliability and generalization.
  - Optimize models for real-time deployment in ADAS and mobile systems.
  - Integrate additional data (e.g., demographics, psychological traits, environmental conditions) for more personalized assessments.
  - Develop hybrid approaches combining GANs with tree-based or sequential models to enhance both accuracy and robustness.

Evaluation of Hybrid Machine Learning Models for Risky Driving Behavior Classification: A Comparative Study Using RNN-AdaBoost, GANs and XGBoost

Elena Theodoraki

Transportation Engineer, Research Associate

Together with:

Thodoris Garefalakis, Paraskevi Koliou, George Yannis

Department of Transportation Planning and Engineering

National Technical University of Athens

ICTR 2025

12th International Congress on Transportation Research

16-18 October 2025, Thessaloniki, Greece

HIT- HELLENIC INSTITUTE OF TRANSPORT

HELLENIC INSTITUTE OF TRANSPORTATION ENGINEERS