Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2026) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2026 Toulouse 6-10 July 2026

Enhancing Risky Driving Behavior Classification Using Conditional GANs (cGANs): A Data Augmentation Approach

Eleni Maria Theodoraki[a]*, Paraskevi Koliou[a], George Yannis[a]

[a]*Department of Transportation Planning and Engineering, National Technical University of Athens, 5 Heroon Polytechniou Str., GR-15773, Athens, Greece*

Abstract

Accurately classifying risky driving behaviors is essential for road safety but is hindered by class imbalance, with dangerous behaviors underrepresented in most datasets. This study explores the use of Conditional Generative Adversarial Networks (cGANs) to generate synthetic data for underrepresented risk categories and improve classification performance. Unlike traditional GANs, cGANs enable controlled data generation based on predefined risk levels. Driving behaviors were initially categorized as Normal, Dangerous, or Avoidable Accident. Following augmentation, a binary classification scheme (Normal vs. Avoidable Accident) was adopted. A cGAN was trained on Normal data to generate synthetic high-risk scenarios, which were used to augment the dataset. Classifiers, GAN-based, XGBoost-RF, and RNN-AdaBoost, were then evaluated on both original and augmented datasets.
Results showed that cGAN-generated data significantly improved GAN model accuracy (from 76% to 90% in Belgium; 79% to 91% in the UK). However, hybrid models achieved near-perfect accuracy on augmented data, indicating overfitting and limited generalizability. This study demonstrates the benefits and challenges of using cGANs for synthetic data augmentation in driving behavior classification and underscores the need for careful validation to ensure real-world applicability.

© 2026 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the World Conference on Transport Research – WCTR 2026.

*Keywords:*Risky Driving Behavior; Classification Models; Deep Learning; Machine Learning; GBoost-RF; RNN-AdaBoost; Synthetic Data Augmentation; Generative Adversarial Networks (GANs); Conditional GANs (cGANs)

* Corresponding author. Tel.: +30 2107721575.
E-mail address: e_theodoraki@mail.ntua.gr

2352-1465 © 2026 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the World Conference on Transport Research – WCTR 2026.

1. Introduction

The rapid increase in vehicular traffic has raised serious concerns regarding road safety, as traffic accidents remain a leading cause of injuries and fatalities worldwide. Human driving behaviors, including **speeding, aggressive acceleration, and harsh braking**, significantly contribute to crash risks. Consequently, accurately classifying and predicting risky driving behaviors is **critical** for developing proactive safety measures aimed at accident prevention.

Machine learning models have been widely used for **driving behavior classification**, leveraging real-world driving data to identify and predict hazardous patterns. **Hybrid classification methods**, such as **XGBoost-RF and RNN-AdaBoost**, have demonstrated strong predictive capabilities. However, a major challenge in these approaches is the **imbalance in driving behavior datasets**. Risky driving behaviors, which are crucial for model learning, often constitute only a **small fraction** of the available data, leading to **biased models that favor normal driving patterns**.

Generative Adversarial Networks (**GANs**) have recently emerged as a powerful tool for **data augmentation**, offering a means to synthesize **realistic** samples for underrepresented categories. However, standard GANs **lack explicit control** over category labels, limiting their effectiveness in structured classification tasks. **Conditional GANs (cGANs)** address this limitation by conditioning data generation on specific labels, allowing for **targeted augmentation** of dangerous driving scenarios.

This study explores the effectiveness of **cGANs in improving the classification of risky driving behaviors**. Initially, the dataset categorizes driving behaviors into three risk levels: **Normal, Dangerous, and Avoidable Accident**. However, following data augmentation, the dataset structure is modified, introducing a **binary risk classification** (**Normal and Avoidable Accident**) to enhance model performance. A **cGAN is trained exclusively on normal driving instances**, generating synthetic data for the underrepresented categories. The augmented dataset is subsequently used to

evaluate the performance of **GAN-based classification models**, along with **XGBoost-RF and RNN-AdaBoost**, to determine the impact of synthetic data augmentation on classification accuracy.

The results indicate that while **cGAN-generated data significantly improves the performance of the GAN model**, increasing its classification accuracy from **76% to 90% (Belgium) and 79% to 91% (UK)**, it also **introduces overfitting** in the hybrid models. XGBoost-RF and RNN-AdaBoost achieve near-perfect accuracy on the augmented dataset, suggesting that these models may have **over-learned synthetic patterns that do not generalize well** to real-world scenarios.

2. Literature Review

The classification of driving behaviors has gained significant attention in recent years due to its implications for **road safety, insurance risk assessment, and autonomous vehicle development**. Researchers have applied various **machine learning and deep learning** techniques to identify hazardous driving patterns, predict accident risks, and develop driver-assistance systems.

Early studies focused on **traditional machine learning algorithms**, such as **Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), and Gradient Boosting Machines (GBM)**, to classify driving behaviors based on real-world datasets. These models utilized **speed, acceleration, braking, and steering patterns** to determine risk levels. However, a major drawback of these approaches was their **susceptibility to data imbalance**, where normal driving instances significantly outnumbered risky driving behaviors. This imbalance often resulted in **poor recognition of hazardous events**, leading to suboptimal classification performance.

To overcome these limitations, researchers introduced **deep learning models**, such as **Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks**, which effectively captured **sequential dependencies** in driving behavior data. These models significantly improved the ability to detect temporal patterns associated with unsafe driving actions. Additionally, **hybrid learning approaches**, such as **combining RNNs with boosting techniques (e.g., RNN-AdaBoost)**, were explored to enhance classification accuracy and reduce bias toward majority classes. While these methods improved overall performance, they still struggled with imbalanced datasets, which affected their generalization capabilities.

To overcome the issue of data imbalance, recent work has introduced Generative Adversarial Networks (GANs) as a means of generating synthetic driving behavior data. Traditional GANs demonstrated potential by enriching datasets with plausible artificial samples. However, their lack of label control limited their application in classification tasks where class-specific synthesis is critical. Conditional GANs (cGANs) addressed this gap by conditioning data generation on predefined risk labels, thus enabling targeted augmentation for rare classes such as Dangerous and Avoidable Accident behaviors.

Several studies support the application of cGANs in driving behavior analysis. For instance, (Park et al., 2025) developed a real-time crash risk prediction model using cGANs, identifying key crash predictors such as temporal speed differences and weather conditions. Similarly, (Zhang et al., 2018) proposed DeepRoad, a GAN-based metamorphic testing system that enhanced the reliability of autonomous driving models under varied conditions. Other approaches, such as Dual Pyramid GAN (Giannoulopoulos, 2024) and hybrid CycleGAN pipelines (Rajagopal et al., 2023), further expanded GAN utility into urban mobility and photorealistic scene generation.

Recent advances in conditional generative modeling have also emphasized the importance of diversity in synthetic data generation. (Yang et al., 2019) introduced Diversity-Sensitive GANs (DSGANs), a regularization-based enhancement to standard cGANs that explicitly addresses the problem of mode collapse, a common issue where generators produce overly similar outputs. Their work demonstrated that incorporating a diversity-promoting regularization term significantly improves the variety and quality of generated samples, leading to more generalizable models across image translation, inpainting, and video prediction tasks.

Building on these theoretical insights, (Giannoulopoulos Antonios, 2024) applied cGANs in a real-world driving context to improve the classification of risky behaviors using naturalistic driving data. By generating synthetic instances of Avoidable Accidents, the study significantly improved recall and overall accuracy of GAN-based classifiers. However, it also revealed overfitting in hybrid models such as XGBoost-RF and RNN-AdaBoost, primarily due to limited diversity in synthetic data. These findings reaffirm the relevance of Yang et al.'s approach, highlighting the need for diversity-sensitive mechanisms in conditional data augmentation pipelines to ensure real-world robustness and model reliability.

Despite their benefits, cGAN-based augmentation methods also present notable challenges. Overfitting remains a primary concern, especially when classifiers learn synthetic patterns too well, limiting their performance on unseen real-world data. Additionally, altering the original data distribution through synthetic augmentation may inadvertently introduce biases or shift feature importance, as revealed by SHAP analysis. To address these issues, future research must explore advanced augmentation techniques, such as noise regularization, adversarial training, or diversity-sensitive frameworks, to improve generalization and reduce over-reliance on artificial data. Given these advancements, the next section outlines the methodology adopted in this study, including data sources, preprocessing steps, cGAN architecture, and model evaluation.

3. Methodology

3.1. Data collection

As part of the i-DREAMS research initiative, a naturalistic driving study was conducted with participants from Belgium and the UK. The Belgian cohort consisted of 43 drivers, generating a dataset of 7,163 trips and 147,337 minutes of driving data. In the UK, 26 drivers participated, contributing 8,226 trips and 118,175 minutes of recorded driving time. The study aimed to collect extensive data on driving behavior and road conditions to enable a detailed analysis of risky driving behaviors.

The experiment was conducted over a four-month period and was divided into four distinct phases:

1. Baseline Phase (4 weeks): No interventions were applied, allowing for the collection of natural driving behavior data.
2. ADAS Intervention Phase (4 weeks): Participants received real-time vehicle warnings through an Advanced Driver Assistance System (ADAS).
3. Performance Feedback Phase (4 weeks): Participants received post-drive feedback via a mobile app.
4. Gamification Phase (4 weeks): Feedback mechanisms were expanded with gamification elements to promote safer driving habits.

Throughout all phases, real-time driving behavior was continuously monitored. The study assessed the effectiveness of ADAS warnings and post-drive feedback mechanisms in influencing driver behavior. Figure 1 provides an overview of the experimental design.

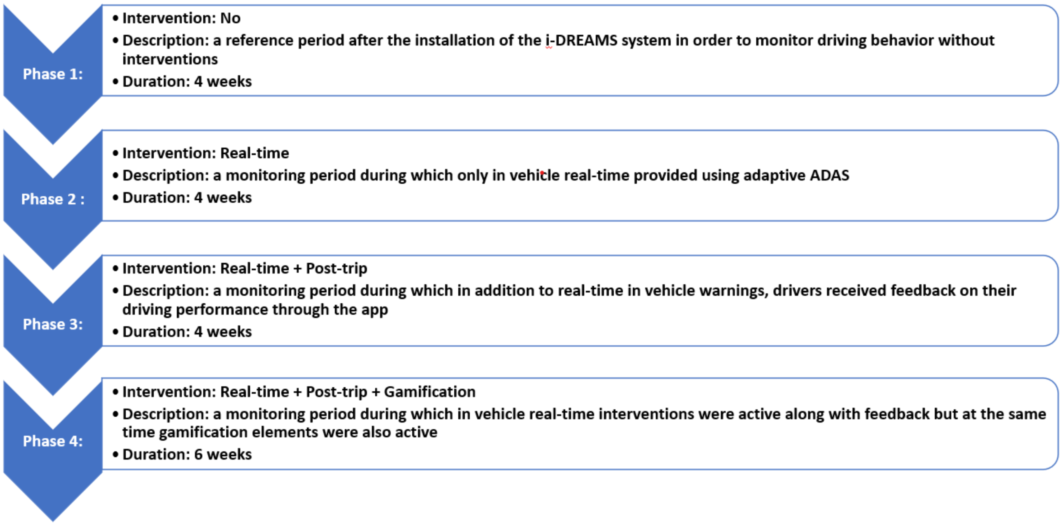

Fig. 1: Overview of the different phases of experimental design

Data collection was performed using an OBD-II device installed in each vehicle, which captured hundreds of driving parameters. Additionally, a mobile data acquisition system enabled seamless data transmission without requiring user input. To classify driving behavior, each 30-second segment of a trip was categorized into one of three safety levels: Normal, Dangerous, or Avoidable Accident. These classifications were based on predefined intervention thresholds and key driving parameters such as speed and headway distances.

3.2. Definition of 'Safety Tolerance Zone'

Before implementing classification algorithms, the dataset was structured into distinct levels of the Safety Tolerance Zone, essential for accurate analysis and modeling. The Safety Tolerance Zone was divided into three categories based on established intervention thresholds:

- **Normal:** Low-risk driving behavior.
- **Dangerous:** Medium-risk behavior, requiring caution.
- **Avoidable Accident:** High-risk behavior, requiring immediate intervention.

This classification was based on two primary performance indicators:

- **Headway distance** (distance between the driver's vehicle and the preceding vehicle).
- **Speeding behavior** (excessive deviation from speed limits).

Given that safety-critical driving events are less frequent, it was expected that Normal driving instances would dominate the dataset, while Dangerous and Avoidable Accident cases would be underrepresented. The dataset contained intervention variables labeled as iDreams_Headway_Map_level_i and iDreams_Speeding_Map_level_i, where i represented intervention levels from -1 to 3:
0: The intervention level does not correspond to i.
1: The intervention level corresponds to i.

To determine the Safety Tolerance Zone for each timeframe, intervention variables were analyzed to identify the most severe safety level encountered. The final classification framework was:

- Normal: Highest intervention level recorded as -1, 0, or 1.
- Dangerous: Highest intervention level recorded as 2.
- Avoidable Accident: Highest intervention level recorded as 3.

This hierarchical classification ensured an accurate representation of real-world driving risks, providing a foundation for subsequent machine learning models aimed at predicting and mitigating unsafe driving behaviors.

3.3. Data Collection and Preprocessing

The dataset utilized in this study consists of real-world driving data collected from naturalistic driving experiments. The raw dataset contains various driving parameters, including speed, acceleration, braking intensity, and trip duration. To ensure consistency and enhance model training, several preprocessing steps were performed:

1. **Data Cleaning:** Missing values (-9999) were identified and removed to prevent inconsistencies.
2. **Feature Normalization:** All numerical features, including *GPS_distances_sum* and *GPS_spd_mean*, were normalized using Min-Max scaling to standardize feature distributions.
3. **Risk Level Categorization:** Each driving instance was classified into one of three risk categories based on excessive speeding and harsh braking behaviors (see Table 1).
4. **Class Balancing:** Since the dataset exhibited significant class imbalance, with "Normal" behaviors dominating, data augmentation techniques were applied to ensure a more equitable class distribution before training the models.

Table 1: Risk Level Classification Criteria

Risk Level	**Condition**
Normal	iDreams_Speeding_Map_level_3_mean ≤ 0.2 AND DEM_evt_hb_lvl_H_mean = 0
Dangerous	0.2 < iDreams_Speeding_Map_level_3_mean ≤ 0.6 OR 0 < DEM_evt_hb_lvl_H_mean ≤ 0.3
Avoidable Accident	iDreams_Speeding_Map_level_3_mean > 0.6 OR DEM_evt_hb_lvl_H_mean > 0.3

These thresholds were established based on prior research in road safety analytics, ensuring a structured and data-driven classification of driving trajectories.

3.4. Conditional GAN (cGAN) Architecture

To create fictitious examples of dangerous driving behavior, a Conditional GAN (cGAN) was used. The model was trained mainly on "Normal" driving trajectories, which were the majority class. Its goal was to generate balanced representations of the underrepresented "Dangerous" and "Avoidable Accident" categories.
The architecture consists of:

- **Generator:** Using fully connected layers with batch normalization and Leaky ReLU activation, the generator creates realistic, label-specific driving data after receiving a noise vector and risk label.
- **Discriminator:** Uses convolutional layers trained with gradient penalty and Wasserstein loss for stable convergence and reduced mode collapse to assess sample authenticity and class consistency.

This configuration improved classifier robustness and increased data diversity by enabling the controlled synthesis of high-risk driving patterns.

Training Procedure

The cGAN was trained iteratively using the following approach:

1. Discriminator Pretraining: Created a solid classification baseline by learning to differentiate between real and fake driving data.

2. Generator Training: Label-conditioned synthetic samples that were optimized to trick the discriminator were created using noise and risk labels.
3. Adversarial Training: To guarantee stable convergence, the generator and discriminator were trained in turn using the Adam optimizer, which included early stopping and adaptive learning rates.
4. Validation: Discriminator performance was tracked to avoid model collapse, and generated samples were compared to actual data to guarantee distribution alignment.

To address class imbalance, this procedure allowed the cGAN to produce high-quality, risky driving samples and model realistic driving behavior.

3.5. Classification Model Training and Evaluation

Three models were used to evaluate the efficacy of the cGAN-augmented dataset:

- **Extreme Gradient Boosting and Random Forests** are combined in **XGBoost-RF**, a tree-based ensemble that was chosen for its interpretability and good performance on structured data.
- **RNN-AdaBoost:** Addresses class imbalance by combining adaptive boosting with recurrent neural networks to capture temporal patterns.
- **cGAN Classifier:** Prior to and following augmentation, the cGAN's capacity to distinguish between Normal and Avoidable Accident cases was assessed through testing as a classifier.**Evaluation Metrics**

Classification performance was assessed using standard metrics:

- **Accuracy:** Overall correctness of predictions.
- **Precision:** Proportion of correctly identified risky driving instances.
- **Recall:** Ability to detect all actual risky cases.
- **F1-score:** Harmonic mean of precision and recall for balanced evaluation.

To ensure robustness, **k-fold cross-validation** was applied. Additionally, **SHAP analysis** was used to interpret feature importance and explain model decision-making.

3.6. Assessing the impact of Data Augmentation

To evaluate the effectiveness of **cGAN-generated synthetic data**, model performance was compared across two datasets:

1. **Original Dataset:** Training and testing models solely on real-world driving data.
2. **Augmented Dataset:** Incorporating **cGAN-generated synthetic data** into training to balance class distributions.

While augmentation improved classification metrics, it also introduced **overfitting risks**. Specifically:

- **XGBoost-RF exhibited high accuracy and precision** but had artificially low misclassification rates due to the homogeneity of synthetic samples.
- **RNN-AdaBoost benefited from increased recall**, yet its performance was **inflated** and might not generalize well in real-world driving conditions.
- **The cGAN model itself showed enhanced classification ability**, suggesting that the generated data was effective at improving representation of high-risk behaviors.

The impact of augmentation is further analyzed in the **Results** section, where quantitative findings are presented and overfitting challenges are discussed in depth.

4. Results

This section presents the results of the classification models trained on both the original and augmented datasets. The performance of the models was evaluated based on key classification metrics, including accuracy, precision, recall, and F1-score. Additionally, the impact of synthetic data augmentation on model generalization and overfitting was examined.3.1 Performance of Classification Models. Table 2 provides a comparative analysis of classification performance across the XGBoost-RF, RNN-AdaBoost, and GAN-based models for both the original and augmented datasets.

Table 2: Performance Comparison of Classification Models

Dataset	Model	Accuracy	Precision	Recall	F1-score
Belgium (original dataset)	XGBOOST & RF	93%	93%	93%	93%
	RNN & AdaBoost	83%	82%	83%	82%
	GANS	76%	64%	76%	66%
UK (original dataset)	XGBOOST & RF	92%	92%	92%	91%
	RNN & AdaBoost	85%	84%	85%	84%
	GANS	79%	80%	79%	74%
Belgium (Augmented dataset)	XGBOOST & RF	100%	100%	100%	100%
	RNN & AdaBoost	100%	100%	100%	100%
	cGANS	90%	90%	90%	90%
UK (Augmented dataset)	XGBOOST & RF	100%	100%	100%	100%
	RNN & AdaBoost	100%	100%	100%	100%
	cGANS	91%	90%	91%	91%

The results indicate a substantial improvement in classification performance when synthetic data was introduced. The cGAN model's performance improved significantly, particularly in recall and overall accuracy. However, XGBoost-RF and RNN-AdaBoost exhibited near-perfect scores on the augmented dataset, raising concerns about overfitting and real-world generalization.

4.1. Overfitting Concerns and Generalization

The perfect scores for **XGBoost-RF** and **RNN-AdaBoost** suggest possible overfitting to synthetic data, which would limit generalizability even though the augmented dataset increased accuracy. While class balance was achieved, limited variability in generated samples likely caused models to memorize rather than truly learn risk patterns.
Models may have become overconfident in detecting risky behaviors, leading to a higher false positive rate, as evidenced by K-fold cross-validation, which demonstrated a significant recall improvement but only modest precision gains.

4.2. Feature Importance and Model Interpretability

Key feature contributions to risk classification were identified by SHAP analysis:

- The most significant variables across datasets were those related to speed (GPS_spd_mean and iDreams_Speeding_Map_level_3_mean).
- In order to differentiate between "Dangerous" and "Avoidable Accident" situations, harsh braking events (DEM_evt_hb_lvl_H_mean) were essential.
- Concerns about overfitting were raised by the fact that models trained on the augmented dataset depended more on synthetic features.

Future research should use techniques like dynamic noise injections or alternative generative approaches to increase the variability in synthetic samples in order to increase robustness.

The key insights from the results are as follows:

- Synthetic data augmentation significantly improved cGAN model performance, particularly in recall and overall accuracy.
- XGBoost-RF and RNN-AdaBoost exhibited overfitting when trained on the augmented dataset, achieving near-perfect scores that may not generalize to real-world conditions.
- Speed and braking intensity remained the most influential features in predicting risky driving behaviors, consistent across both datasets.
- SHAP analysis revealed over-reliance on synthetic features, suggesting the need for improved augmentation diversity.
- While synthetic data improved minority class representation, it also introduced potential biases, emphasizing the need for careful validation and real-world testing.

These findings underscore the dual impact of synthetic data augmentation—while it enhances classification performance and mitigates class imbalance, it also raises concerns about model robustness and generalization. Future work should explore alternative augmentation strategies and adversarial training techniques to mitigate overfitting and improve real-world applicability. These findings are explored in more depth in the following discussion, which considers practical implications, limitations, and future directions.

5. Discussion

The findings of this study highlight the significant impact of using cGAN-generated synthetic data to address class imbalance in risky driving behavior classification. While model performance improved across key evaluation metrics, several critical challenges and limitations must also be considered.

5.1. Effectiveness of Data Augmentation

The results show that adding synthetic data notably improved recall and overall accuracy across all models, indicating that cGAN-generated samples effectively supplemented high-risk instances and helped address class imbalance. Improved recall suggests better identification of risky driving behaviors, which is vital for accident prevention. However, the gain in precision was limited. Although models detected more risky events, they also generated more false positives, likely due to overfitting on synthetic patterns that may not generalize well to real-world data. This trade-off highlights that while synthetic data enhances risk detection, it can also increase misclassification of normal driving as high-risk.

5.2. Overfitting and Generalization Challenges

Although accuracy improved, analysis of precision and false positive rates reveals concerns about generalization. The growing gap between precision and recall in the augmented dataset suggests models may have overfit to synthetic patterns instead of learning true high-risk behaviors. This was especially clear in the XGBoost-RF and RNN-AdaBoost models, which achieved perfect accuracy on augmented data but showed inconsistent performance on real-world samples. Such inflated metrics indicate the models likely memorized synthetic data rather than forming robust, generalized decision boundaries.

The SHAP analysis further confirmed overfitting tendencies, revealing an over-reliance on synthetic features. While the synthetic dataset improved class balance, the uniformity of generated samples may have introduced biases that the models exploited. This highlights the need for future research to explore more diverse data augmentation techniques, such as:

- Noise regularization methods to introduce variability in synthetic samples.
- Adaptive synthetic sampling, ensuring that generated data reflects a broader range of real-world driving behaviors.
- Adversarial training techniques to improve model robustness against overfitting.

5.3. Practical Implications for Road Safety

From a real-world perspective, enhancing the identification of high-risk driving behaviors has substantial implications for road safety applications. Improved classification models can be leveraged in real-time risk assessment systems, enabling:

- Adaptive driver assistance warnings, reducing the likelihood of collisions.
- Autonomous vehicle safety measures, improving self-driving systems' ability to react to hazardous conditions.
- Insurance telematics applications, allowing for more accurate driver risk profiling.

However, ensuring that machine learning models generalize effectively across diverse driving conditions and driver behaviors is essential for practical deployment. Future research should focus on refining the training process by integrating:

- Controlled driving simulations to validate model performance in dynamic environments.
- Real-time telematics data to enhance real-world applicability.
- Model interpretability frameworks, improving trust in AI-driven risk assessment systems for safer deployment in real-world driving applications.

The key findings of the discussion are as follows:

- cGAN-generated synthetic data significantly improved recall and overall accuracy, effectively addressing class imbalance.
- Overfitting was observed, particularly in hybrid models (XGBoost-RF and RNN-AdaBoost), which achieved artificially high accuracy on augmented data.
- SHAP analysis revealed an over-reliance on synthetic features, emphasizing the need for more diverse data augmentation techniques.
- Practical applications in road safety highlight the potential of improved classification models, but generalization remains a key challenge.

These findings underscore the dual impact of synthetic data augmentation—while it enhances model performance, it also raises concerns about overfitting and real-world applicability. Future research should prioritize diverse augmentation strategies and advanced validation techniques to develop more generalizable risk classification models.

References

Giannoulopoulos Antonios. (2024). ARISTOTLE UNIVERSITY OF THESSALONIKI FACULTY OF SCIENCES SCHOOL OF INFORMATICS Utilising GANs for Training Intelligent Agents for Autonomous Vehicles in a Simulator with Deep Learning.

Park, N., Park, J., & Lee, C. (2025). Conditional Generative Adversarial Network-Based roadway crash risk prediction considering heterogeneity with dynamic data. Journal of Safety Research, 92, 217–229. https://doi.org/10.1016/j.jsr.2024.12.001

Rajagopal, B. G., Kumar, M., Alshehri, A. H., Alanazi, F., Deifalla, A. farouk, Yosri, A. M., & Azam, A. (2023). A hybrid Cycle GAN-based lightweight road perception pipeline for road dataset generation for Urban mobility. PLoS ONE, 18(11 November). https://doi.org/10.1371/journal.pone.0293978

Yang, D., Hong, S., Jang, Y., Zhao, T., & Lee, H. (2019). Diversity-Sensitive Conditional Generative Adversarial Networks. http://arxiv.org/abs/1901.09024

Zhang, M., Zhang, Y., Zhang, L., Liu, C., & Khurshid, S. (2018). Deeproad: GaN-based metamorphic testing and input validation framework for autonomous driving systems. ASE 2018 - Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, 132–142. https://doi.org/10.1145/3238147.3238187