IRF INTERNATIONAL ROAD FEDERATION | WORLD CONGRESS 2024

ISTANBUL 15-18 October Türkiye

15 October 2024

Enhancing cyclist safety: Predictive analysis of injury severity and advocacy for evidence-based interventions

Virginia Petraki, Stella Roussou, Apostolos Ziakopoulos, George Yannis

National Technical University of Athens

Virginia Petraki

National Technical University of Athens

Research Associate, PhD Candidate

Introduction

- Cycling is **growing in popularity** promoting health and sustainability.
- Urban cycling increases as a response to **traffic congestion** and **air pollution**.
- Cyclists **face high risks** from crashes, especially in urban areas.
- They are **vulnerable**, lacking physical protection compared to vehicle occupants.
- It is important to **understand injury factors** to improve road safety and promote sustainable transportation.

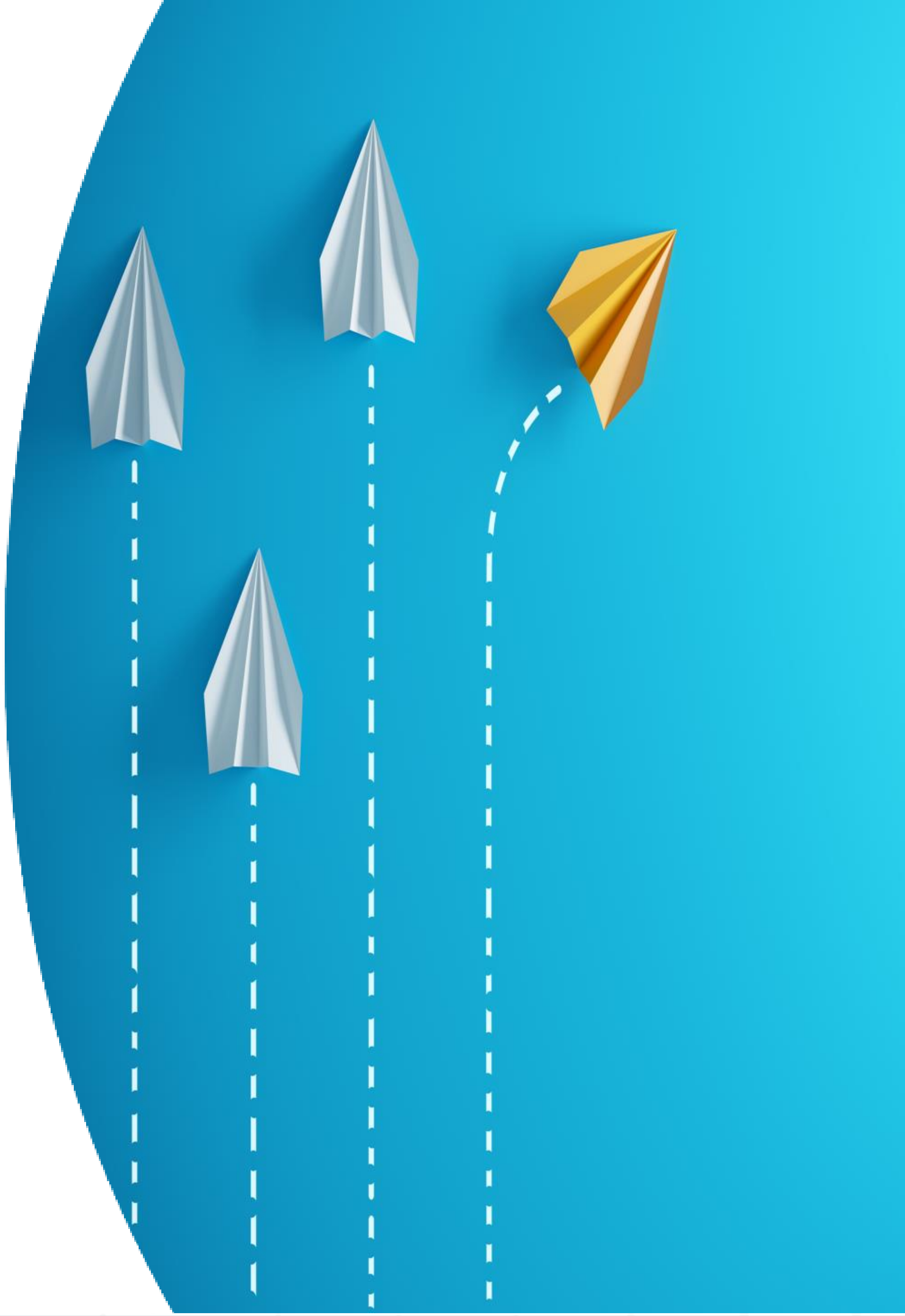
IRF INTERNATIONAL ROAD FEDERATION | WORLD CONGRESS 2024

Study Objective

The aim is to examine the **factors** influencing the **severity** of cyclist injuries in road accidents within Great Britain (1979-2018).

- **CatBoost** machine learning algorithm is used to predict injury severity.

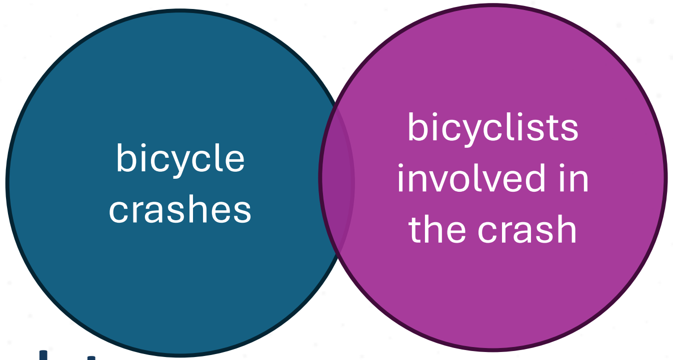
- The identification of the most critical predictors of injury severity aims to provide **valuable insights** that can guide **policy formulation** and enhance safety measures for cyclists.



Data Crashes

- **Dataset**: Bicycle crashes in Great Britain (1979-2018).
- The dataset includes a comprehensive range of variables that provide a detailed account of each bicycle crash:
  - Crash Date and Time
  - Severity of the crash categorized into fatal, serious, and slight injuries
  - Weather Conditions include sunny, rainy, snowy
  - Road Conditions such as dry, wet, icy, etc.
  - Light Conditions daylight, darkness with or without street lighting
  - Vehicle Involvement: the number and types of vehicles involved in the crash
  - Age group and gender of the cyclists
  - Number of casualties

Data Pre-processing



- **Merging** two distinct datasets
- Handling **missing data**
- Identifying the **outliers**
- **Grouping** speed limits into 3 categories:
  - Low Speed (<30mph)
  - Medium Speed (30mph-50mph)
  - High Speed (>50mph)
- Applying **SMOTE** for class imbalance.
- Encoding categorical variables
  - Fatal and Serious Injury (FSI) = 1, Light Injury =0

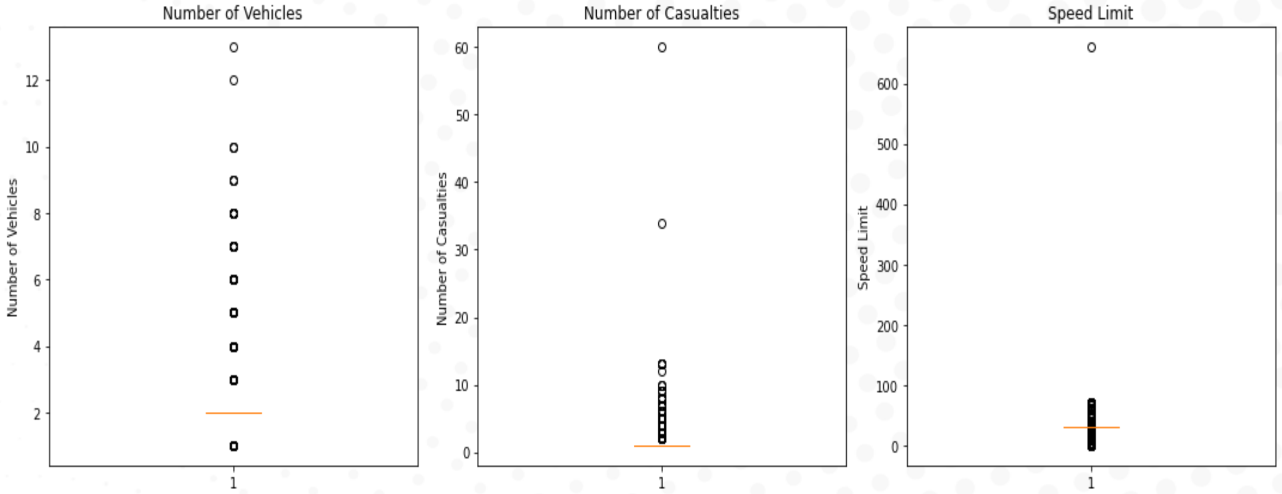


Figure 1: Outliers Boxplots

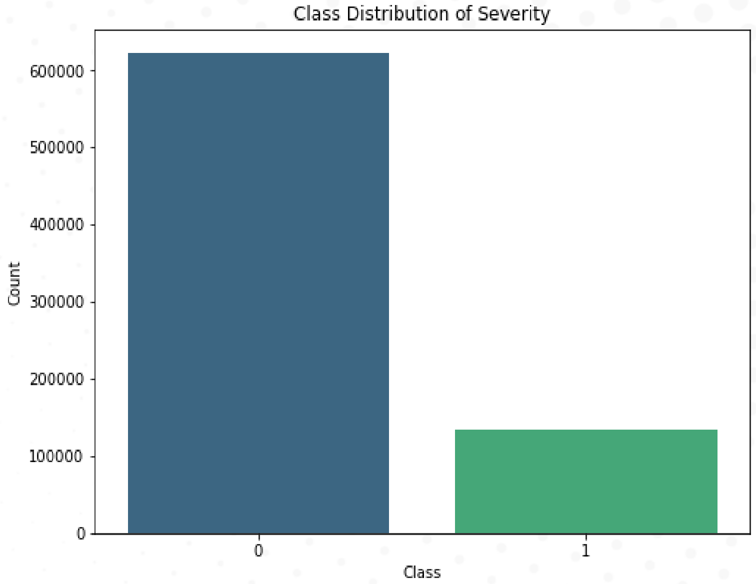


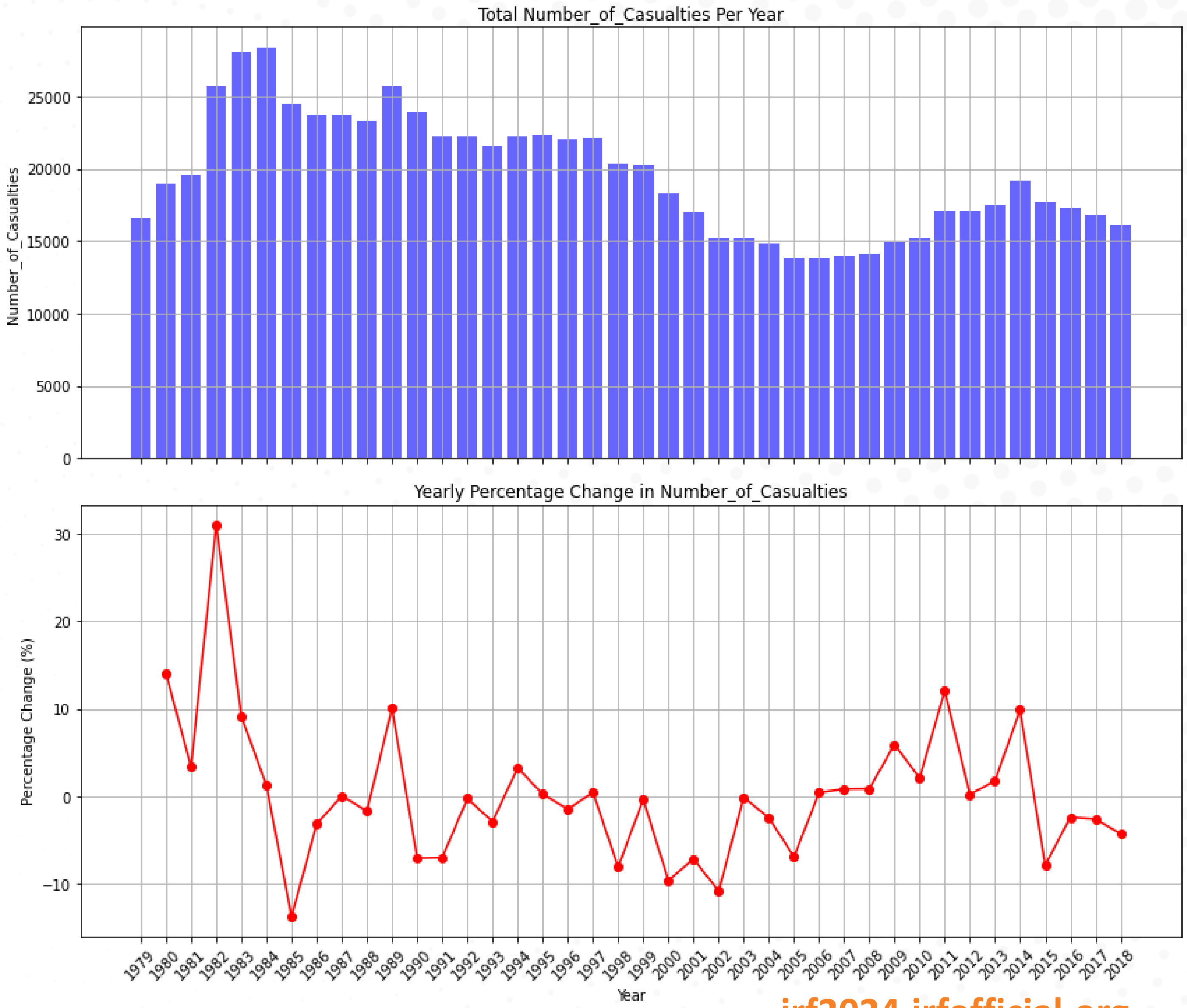
Figure 2: Class Distribution of Severity

Statistics

- The initial rise and peak in casualties (top plot) highlight a period of **increased risk**, followed by a general decline that points to **successful interventions** in improving road safety.

- The recent fluctuations and slight upward trend in both total casualties and percentage changes suggest that while past measures have been effective, **ongoing efforts are required** to address new and emerging factors contributing to casualties.

Figure 3: Number and Percentage Change of Casualties Per Year



Study Objective

- Speed limit 30mph-50mph (category **1**) has the **highest number** of casualties
- Areas with this speed limit might **require targeted interventions** to reduce casualties
- The lower number of casualties in **high-speed roads** (category **2**), is likely due to the fact that there are fewer bicycles on these roads
- The increasing trend in low speed limit category 0 (<30mph) is a cause for concern and warrants **further investigation** to identify and address the underlying causes

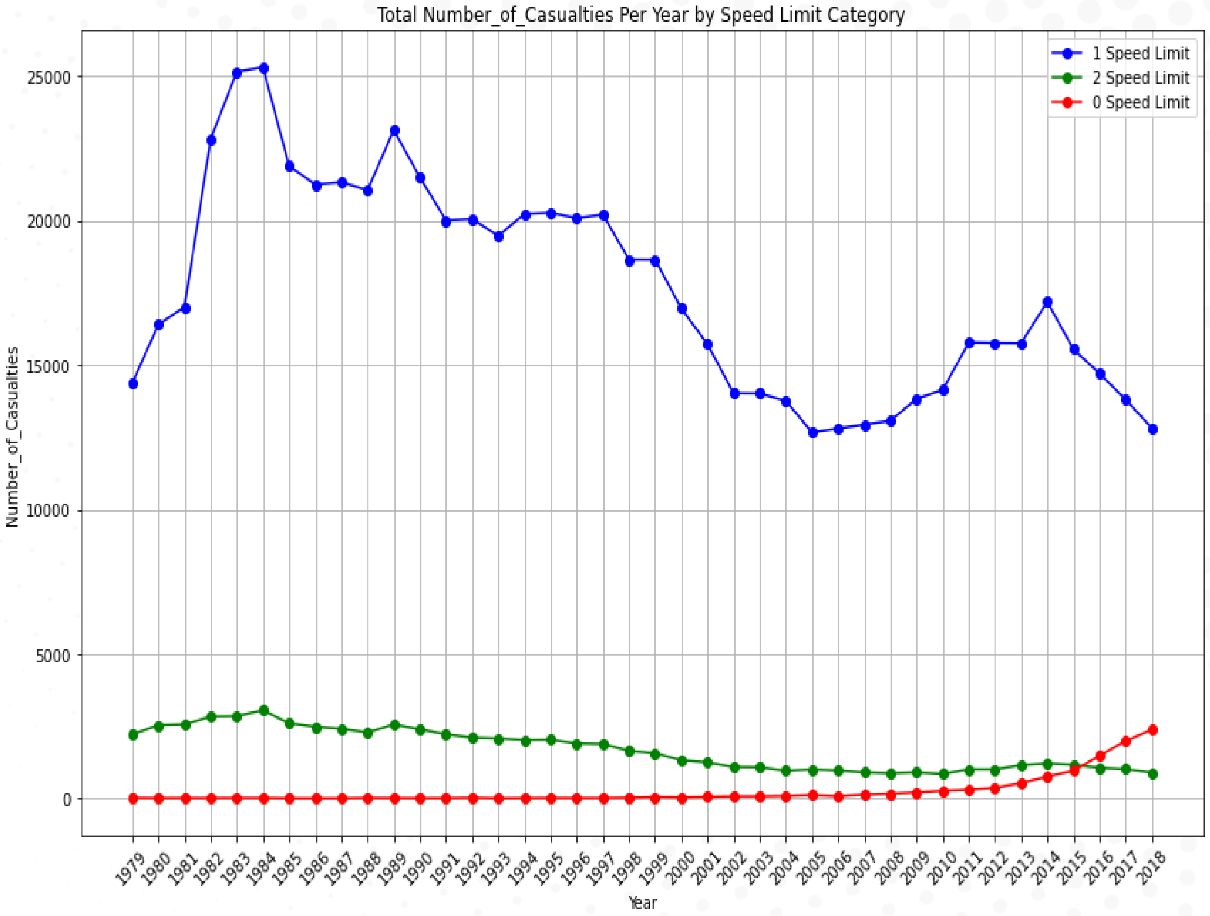


Figure 4: Number of Casualties Per Year based on Speed Limit Categories

Results (1)

- Hyperparameter tuning with **5-fold cross-validation** was carried out to mitigate overfitting and enhance the model's performance with **7 different hyperparameter combinations** tested.

- The **feature importance analysis** provides insights into which variables have the most **significant impact** on predicting the severity of road crashes.
- The **gain metric**, used here, measures the **contribution** of each feature in improving the model's performance.

Hyperparameter	Examined range	Optimized Value
Learning rate	0.01 - 0.29	0.29
Iterations	150 - 450	450
Depth	3 - 11	11
L2 Leaf Reg	1 - 10	10
Border Count	32 - 128	32
Class Weights	{0: 0.5, 1: 50}	{0: 1, 1: 1.15}
Random State	Fixed at 42	42

Table 4: Hyperparameter Tuning Results for CatBoost Model

No.	Feature	Gain
1	Age_Grp	16.826
2	Day	13.993
3	Speed_limit	12.386
4	Number_of_Vehicles	9.866
5	Weather_conditions	9.623
6	Road_type	9.593
7	Light_conditions	9.052
8	Road_conditions	7.838
9	Gender	6.026
10	Number_of_Casualties	4.792

Table 5: CatBoost optimized model feature importance

Results (2) - Feature Importance (Graph)

- The most significant factor is the **age group** as different age groups have varying levels of vulnerability, with younger and older cyclists being potentially more at risk of severe injuries.
- The **day of the week** is also important, reflecting differences in traffic patterns and cycling activities between weekdays and weekends, which can influence injury severity.
- The **speed limit** at the crash location is a critical factor, which directly impacts the severity of injuries due to the increased impact force associated with higher speeds.
- The **number of vehicles** is also significant, suggesting that incidents involving more vehicles tend to be more severe.
- Other significant features include **weather conditions**, which affect the likelihood and severity of injuries, with adverse conditions like rain or snow increasing the risk.
- **Light conditions** an important role, as poor visibility during nighttime can lead to more severe injuries.

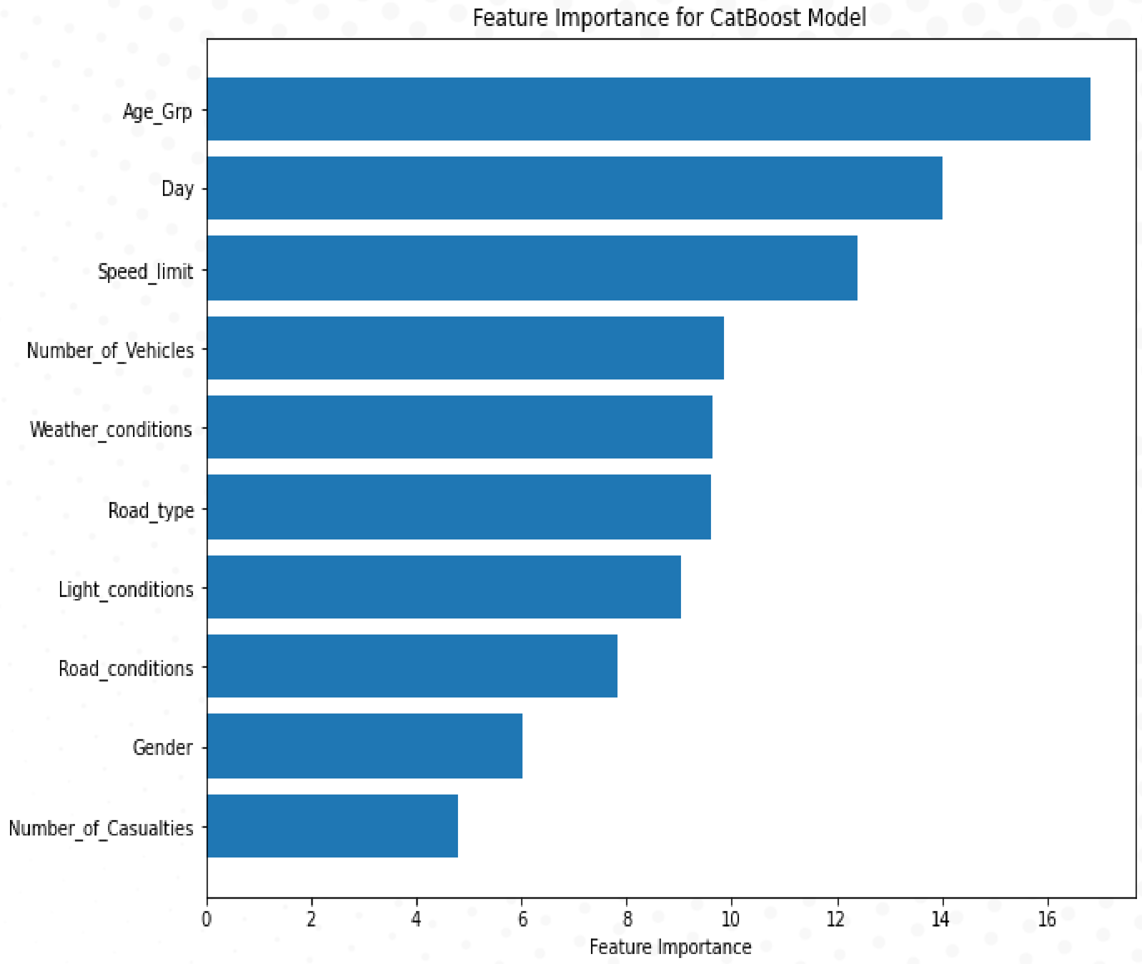


Figure 5: CatBoost Feature Importance Plot

Results (3) – Model Performance Evaluation

- The model's performance on the test set is evaluated using
  - **Precision**
  - **recall**, and
  - **F1-score metrics**

  for each class, along with overall accuracy.

- The overall accuracy of the model is 0.59, as showcased in Table 6, suggesting that the model correctly predicts the severity of cyclist injuries 59% of the time.

Metric	Class 0 (FSI)	Class 1 (Non-injury/Light)
Precision	0.59	0.59
Recall	0.60	0.59
F1-score	0.59	0.59
Accuracy		
Macro avg	0.59	0.59
Weighted avg	0.59	0.59

Table 6: Model Performance Metrics

	Predicted Class 0	Predicted Class 1
Actual Class 0	74095	50291
Actual Class 1	51374	72865

Table 7: Confusion Matrix

Key Findings

- **Machine learning** helps identify **critical factors** influencing cyclist safety.
- **Speed limits** and age group play a **major role** in injury severity.
- **Weather** (rain, snow) and **poor lighting** increase risks.
- **Poor road conditions** and highways increase severity risks.
- Recommendations aim to inform policymakers, urban planners, and transportation authorities to **enhance cyclist safety**.

IRF INTERNATIONAL ROAD FEDERATION | WORLD CONGRESS 2024

Policy Implications & Future Research

- Enforcement of **lower speed** limits is crucial for reducing injury severity.
- Tailored **programs for younger and older cyclists** to promote safety.
- Continuous **policy updates** based on data-driven insights are needed to promote cyclist safety and sustainable urban mobility.
- Focus on **recent data**, behavioral aspects, and emerging technologies (ADAS, smart cities) for future insights.

Thank you for your attention!

Virginia Petraki

Research Associate, PhD Candidate, NTUA

vpetraki@mail.ntua.gr