Predictive Modeling for Driver Insurance Premium Calculation using Advanced Driver Assistance Systems and Contextual Information

Leandro Masello, Barry Sheehan, German Castignani, Montserrat Guillen, Finbarr Murphy.

Abstract—Telematics devices have transformed driver risk assessment, allowing insurers to tailor premiums based on detailed evaluations of driving habits. However, integrating Advanced Driver Assistance Systems (ADAS) and contextualized geolocation data for predictive improvements remains underexplored due to the recent emergence of these technologies. This article introduces a novel risk assessment methodology that periodically updates insurance premiums by incorporating ADAS risk indicators and contextualized geolocation data. Using a naturalistic dataset from a fleet of 354 commercial drivers over a year, we modeled the relationship between past claims and driving data through claims frequency using Poisson regression and claims occurrence probability using machine learning models, including XGBoost and TabNet. The dataset is divided into weekly profiles containing aggregated driving behavior, ADAS events, and contextual attributes. Risk predictions from these models are used to compute weekly premiums for each driver. SHAP is employed to interpret the machine learning model predictions. Results indicate that XGBoost achieved the lowest Log Loss, reducing it from 0.59 to 0.51 with the inclusion of ADAS warnings and driving context. However, these improvements were not consistent across all models and did not show statistically significant differences in ROC AUC values. The proposed methodology computes weekly premiums based on risk predictions from these models, penalizing risky behaviors while incentivizing safe driving behaviors. This dynamic pricing can be incorporated into the insurance lifecycle, enabling tailored policies based on emerging technologies. The study demonstrates the value of integrating diverse data sources for bespoke risk assessment and weekly insurance pricing.

***Index Terms*—advanced driver assistance systems, explainable artificial intelligence, generalized linear models, machine learning, risk assessment.**

I. INTRODUCTION

Motor insurance companies assess their customers' risks to provide coverage for their potential losses. Typical risk factors involve vehicle and driver metadata that divide the portfolio of drivers into different segments according to their crash risk (i.e., risk segmentation) [1]. However, such factors need an essential layer for predicting crashes: driving habits. Usage-based insurance (UBI) addresses this issue by relying on dynamic driving data collected through telematics devices to offer personalized and dynamic risk assessment. The benefits of such an approach are three-fold: it leads to fair pricing, incentivizes safe driving practices, and enables access to risk indicators periodically, even before a crash occurs [2], [3].

The advent of emerging vehicular technologies, such as Advanced Driver Assistance Systems (ADAS), offers opportunities for refining personalized risk assessments. These systems enhance vehicle performance and road safety by assisting drivers with safety-relevant feedback [4], [5]. Depending on the level of automation, some ADAS can control the vehicle's motion (e.g., Automatic Emergency Braking), while warning-based ADAS only triggers alerts about safety-related events (e.g., Forward Collision Warning) [6]. The feedback provided in both cases has direct implications for the frequency and severity of road crashes, therefore modifying the inherent driving risk [7].

Incorporating ADAS data into risk assessment models enables a deeper understanding of driving risk. Such data encompass risk factors related to distraction and risky behaviors that are unavailable with telematics devices. For example, a driver who receives many forward collision warnings per trip has a higher risk appetite than another who keeps a conservative distance from the vehicle ahead. Similarly, driver distraction has several consequences in driving safety, including keeping safe headway distances [8], speed regulation [9], and lane position [10]. Driver inattention from engaging in visually or manually complex tasks has been linked to a three-fold increase in driving risk [11]. Thus, ADAS data presents opportunities for improved risk assessment, constituting relevant information for motor insurers and road safety stakeholders.

The driving context represents another information layer that plays a significant role in automobile risk assessment. Driving behavior cannot be comprehensively assessed in isolation; it must be interpreted within the context of the driving environment, including road types, traffic conditions, road infrastructure, and weather. These contextual factors influence driving risk and the effectiveness of ADAS [7]. For instance, speeding excessively on motorways has a different impact than on urban roads [12]. Similarly, the impact of ADAS can differ based on the complexity of road layouts or weather conditions. Therefore, a comprehensive risk assessment model must integrate data on driving behavior and ADAS use, along with

This project was supported by the Fonds National de la Recherche, Luxembourg (Project Code: 14614423) and the Spanish Ministry of Science and Innovation, NextGenerationEU (Project Codes: TED2021-130187B-I00 and PID2019-105986GB-C21). *(Corresponding author: Barry Sheehan).*

Leandro Masello, Barry Sheehan, and Finbarr Murphy are with the University of Limerick, Limerick KB3-040, Ireland. (email: barry.sheehan@ul.ie).

Leandro Masello is with Motion-S S.A., Mondorf-les-Bains L-5610, Luxembourg.

German Castignani is with the Luxembourg Institute of Science and Technology (LIST), 4362 Esch-sur-Alzette, Luxembourg.

Montserrat Guillen is with the Universitat de Barcelona, 08007 Barcelona, Spain.

contextual information that reflects the driving conditions under scrutiny.

Integrating ADAS data and the driving context into risk assessment models presents several challenges. One significant challenge is these technologies' recent emergence and varied availability, leading to inconsistent data quality, completeness, and a lack of historical correlations with claims. Additionally, the contextual information requires sophisticated methods for accurate data collection and processing. Another challenge is the high dimensionality and heterogeneity of the data, which can complicate the modeling process and require advanced techniques to identify latent relationships. This challenge also influences the willingness to share this data. While many drivers are attracted to potential cost savings from lower premiums for safe driving, privacy concerns remain significant. Insurance companies can address these concerns by ensuring data collection and usage transparency, demonstrating clear benefits, and maintaining trustworthy practices.

This article introduces a risk assessment framework that integrates data concerning driving behavior, ADAS warnings, and the driving context into UBI schemes. It studies predictive modeling for claims frequency and claim occurrence probability in a fleet of commercial vehicles equipped with ADAS. To the best of the authors' knowledge, the article contributes the first risk assessment methodology incorporating weekly contextual information and ADAS warnings into the driver's risk profile. The proposed framework leverages the claims history of a driving fleet to find associations between at-fault claim frequencies and driving patterns through two predictive modeling perspectives: (i) a claims frequency using Poisson Regression and (ii) a claims occurrence probability based on five machine learning techniques – Logistic Regression, Support Vector Machine, Random Forest, XGBoost, and TabNet. The research provides a framework for insurance premium determination that can evolve in tandem with vehicular technologies.

The lack of explainability of machine learning algorithms poses challenges in meeting the transparency requirements set by insurance pricing regulations [13]. Explainable Artificial Intelligence (XAI) appears as an approach to reaching a balance between complex models and regulatory compliance. Shapley Additive Explanations (SHAP), an XAI method based on game theory, can effectively interpret complex models' predictions by analyzing each feature's contribution to the model's output [14]. Despite its application in several domains [15], [16], the adoption of XAI in motor insurance, particularly for interpreting the risk derived from naturalistic driving data, remains limited. The second contribution of this research is the application of SHAP into the insurance lifecycle by analyzing the contribution of each studied feature to the driver risk.

The implications of this research extend to risk management and transportation stakeholders. It offers opportunities for insurers to enhance risk segmentation and pricing strategies for customers with ADAS. Furthermore, weekly pricing tied to the driver's performance incentivizes safe driving patterns since drivers receive feedback with clear implications on how much they pay, as Ellison et al. [17] posited. Thus, the impact of the paper is also relevant for road safety stakeholders.

The article is organized as follows. Section 2 details previous works on risk assessment and UBI. Section 3 presents the collected naturalistic driving dataset from a commercial Light-good Vehicles (LGV) fleet monitored for a year. Section 4 introduces the methodology to evaluate claims frequency and probability with and without ADAS and contextual features. It also introduces approaches to using the resulting models for weekly insurance ratemaking. The respective results of such methods, including model performances and premium implications, are elaborated in Section 5. The article concludes with the main contributions, applications for stakeholders, and future work in Section 6.

II. Related work

Risk assessment for UBI focuses on driving behavior, which can be obtained through vehicle dynamics. The rationale is that risky drivers tend to perform considerable aggressive maneuvers involving acceleration and speeding [18], [19]. Typical values for aggressive acceleration events are absolute magnitudes higher than six m/s^2 [20], [21]. The speeding attitude (i.e., the propensity to violate speed limits) supplements acceleration information by reflecting drivers' negligence and sensation-seeking [22]. Driven distance is another commonly studied factor due to its positive association with claims and driving exposure [23]. However, thorough assessments must understand where policyholders drive, not just how much. The driving context in which certain habits occur addresses this challenge by allowing a deeper understanding of risk attitudes. Ma et al. [24] showed the importance of driving context in a ratemaking scheme based on the relationships between claims history and traditional risk factors, driving habits, and the context in which people drive. Factors like traffic conditions, road layout, road signs, and weather information are associated with road crashes and therefore relevant for insurance premiums [25], [26], [27].

Data from ADAS provides another layer of information for modeling driver risk. These emerging technologies capture driver distraction, a leading cause of crashes, representing around 9.7 % and 7.1 % of fatal crashes in the United States and the United Kingdom [28], [29]. While several factors may cause driver inattention, engaging in cell phone calls is among the most studied causes [27], [30]. Other distraction causes involve fatigue, talking to other passengers, or smoking [31], [32], and their impact on safety makes them worth considering for driver risk assessment.

Integrating ADAS in automobile insurance has received attention in recent literature. Studies suggest that ADAS features, such as automatic emergency braking and lane departure warning, significantly reduce accident frequency, thereby lowering insurance claims [7]. However, while these systems decrease overall collision frequency, they may also increase the severity of claims due to the higher costs of repairing more complex systems. Shannon et al. [33] conducted a comprehensive analysis of various levels of vehicular autonomy and their impact on claims frequency and loss distributions, estimating an increase in large-cost events despite the reduction in claims frequency. In contrast, research by LexisNexis Risk Solutions, which analyzed 11 million vehicles, demonstrated that even though repair costs may rise, reduced

claims frequency leads to overall decreases in claims loss [34]. Their findings showed that ADAS-equipped vehicles had a 23% reduction in bodily injury loss cost, a 14% reduction in property damage loss cost, and an 8% reduction in collision claim loss cost compared to non-ADAS vehicles. Introducing ADAS risk score schemes accentuates the importance of using vehicle safety data in risk assessment frameworks. For instance, Swiss Re's ADAS Risk Score leverages vehicle safety features to offer bespoke premiums to policyholders [35].

Driver risk assessment generally encompasses two modeling approaches. The first consists of the traditional approach used in actuarial sciences based on generalized linear models (GLMs). This approach typically uses Poisson or Negative Binomial distributions to model the claims frequency. For instance, the ratemaking methodology posited by Guillen et al. [20] models claims frequency through vehicle metadata and near-misses (i.e., a situation where an accident is narrowly avoided). The authors illustrate how dynamic risk factors serve to update a traditional insurance premium weekly. The process computes a base premium through traditional risk factors and updates it based on driver behavior events.

The second approach for risk assessment computes the probability that at least one claim happens in a given period, known as claims occurrence probability. This approach is based on Logistic Regression or more complex machine learning algorithms. Paefgen et al. [36] were among the first authors to compare several algorithms from a UBI pricing perspective. The authors concluded that Logistic Regression was the most suitable for insurance due to its interpretability. Ma et al. [24] also studied this algorithm, comparing the results with a Poisson regression for claims frequency, and found that both yielded consistent results. Huang & Meng [37] investigated several machine learning algorithms, and the results showed that ensemble learning, in particular XGBoost, is the method that achieves the best accuracy and robustness. Aiming to compare such algorithms with deep learning models, McDonnell et al. [38] analyzed TabNet for claim occurrence probability, and found that it performs similarly to XGBoost.

In addition to model performance, model interpretation is essential for insurers due to pricing regulations [13]. While GLM models provide straightforward interpretability by reporting attribute coefficients, complex machine learning models pose challenges. Introducing SHAP into predictive modeling for risk assessment is a practical approach to address this issue. Wen et al. [39] found that SHAP was the most effective way to interpret the crash frequency models without limiting the complexity. Li et al. [40] demonstrated the application of SHAP to telematics data within a UBI scheme, providing policyholders with personalized feedback to promote safe driving behavior. Their study, conducted over one year with data from 9,879 vehicles in China, where 14.5% had at least one claim, used SHAP to interpret the claims occurrence probability. This approach helped identify the risk factors contributing to higher claims risk.

This research extends previous works by integrating risk indicators from ADAS and contextual information into UBI models for claims frequency and occurrence probability. The model results are used to compute dynamic weekly insurance premiums, bridging the gap between driving behavior data, risk assessment, and insurance pricing. The proposed methodology leverages machine learning and deep learning techniques to enhance risk segmentation. This approach aims to provide a fair pricing strategy for insurers, reflecting the driving risk of policyholders with emerging technologies. Additionally, using SHAP for model interpretation adds transparency to using complex machine learning models for insurance pricing.

III. DATA

The dataset comprises driving data collected from a fleet of 354 commercial drivers using light-good vehicles in a driving monitoring campaign. As part of their daily job activities, the drivers performed, on average, five daily trips, covering around 143 km. The data collection occurred in the Republic of Ireland between 01/04/2021 and 31/03/2022, encompassing 8,142,896 km from 287,511 trips, where drivers were monitored for 277 days on average. All drivers received feedback about their driving patterns and attended quarterly coaching sessions to meet road safety standards.

The first phase of the data processing pipeline involves collecting driving data. Such data encompasses geolocation samples obtained through a GNSS module and warning-based ADAS, which record timestamped behavior attributes. These timestamps allow us to align events across different data types and augment them with their environmental context using *Motion-S's Contextualizer* service [41]. In this process, geolocation data are augmented with 16 contextual attributes related to road environment, road infrastructure and topology, traffic conditions, road signs, and weather conditions.

Driver behavior events include vehicle dynamics collected from the telematics device and warning-based ADAS events. The former consists of anomalous events of vehicle kinematics that might have led to accidents, including *harsh acceleration* (acceleration greater than 6 m/s^2), *harsh braking* (deceleration greater than 6 m/s^2), and *speeding* [20], [21], [22]. This research classifies speeding into *slight* and *serious speeding* based on events lower or higher than 20 km/h above the speed limit, according to French law [42]. ADAS events involve warnings triggered when the vehicle or driver meets specific criteria, recorded by a driver-facing and road-facing dashcams connected to the vehicle. They include *driver inattention* (when the driver is looking around or talking with a passenger), *making calls*, *smoking*, *fatigue* (when the driver's gaze shows drowsiness), *forward collision* (potential collisions detected against a stopped vehicle when traveling at speeds greater than 20 km/h), *lane departure* (lane changing without using the indicators), and *too close distance* (tailgating events when the vehicle moves at speeds higher than 30 km/h).

The last process of the data processing pipeline consists of getting driver risk profile aggregations using time windows according to the target scheme. This research aligns with [20] by using weekly aggregations. Consequently, the dataset comprises 12,528 driver weeks containing the attributes described in Table I. The studied dataset contains information about the past two years of the fleet claims history. Only at-fault claims are considered, as the goal is to identify relationships between patterns of risky drivers and claims. In the two years, 62 at-fault claims were observed, giving a mean cost per claim of €2,899, where 50 drivers had one claim and six drivers had

two claims. To foster scientific collaboration and ensure the reproducibility of our research, the resulting dataset is publicly available [DOI of the public data – updated after the peer-review].

TABLE I
WEEKLY PROFILE ATTRIBUTES

Category	Attributes[1]
Driving context	mean_speed_limit [km/h], mean_weather_temperature [°C], mean_weather_visibility [m], mean_weather_wind_speed [km/h], prop_clear_weather, prop_congested, prop_more_than_one_lane, prop_motorway, prop_road_quality_moderate, prop_rural, prop_slope_flat, sum_animal_crossing_sign, sum_pedestrian_crossing_sign, sum_roundabout, sum_stop_sign, sum_traffic_light, sum_yield_sign
Driving behavior	sum_harsh_acceleration, sum_harsh_braking, sum_speeding_slight, sum_speeding_serious
ADAS warnings	sum_fatigue_driving, sum_forward_collision, sum_driver_inattention, sum_driver_smoking, sum_driver_making_calls, sum_lane_departure, sum_too_close_distance
Driving exposure	total_distance [km]
Vehicle information	engine_capacity [thousands cc]
Claim information	exposure_in_weeks, claims_count

IV. METHODS

This section presents the risk assessment methodology illustrated in Fig. 1. Using the contextualized weekly profiles described in the previous section, the methodology involves modeling both claim frequency and occurrence probability. Subsequently, it calculates weekly premiums for each driver. These premiums are designed to penalize risky driving behaviors while incentivizing safe driving practices. The models are interpreted using GLM coefficients for claim frequency modeling and SHAP values for machine learning models.

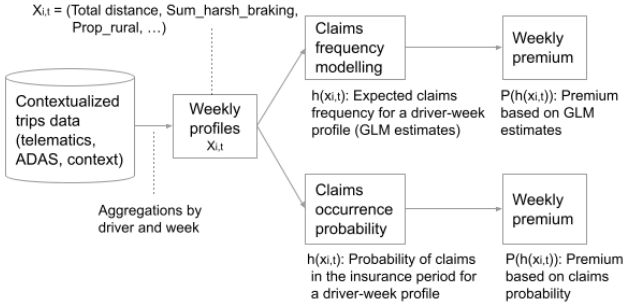

Fig. 1. Methodology for weekly insurance pricing. A weekly profile is a vector *x* with the attributes listed in Table 1 for driver *i* and week *t*. The claims frequency modeling takes a set of weekly profiles and outputs GLM estimates, whereas claims occurrence probability outputs the probability of claims in the insurance period.

Model performance is reported through a 5-fold stratified group cross-validation strategy, considering the nature of claims data and the limited number of drivers in the study. This technique splits the dataset into stratified folds of non-overlapping drivers, preserving the proportion of samples with and without claims. In each iteration, four folds are used as a training set, leaving 20% of the data for evaluation. This process is repeated five times, with each iteration reporting a performance metric according to the modeling approach. The average of these metrics, along with the respective standard deviation over the folds, is then reported to provide a comprehensive evaluation of the model's performance.

A. Claims frequency modeling

The claims frequency modeling aims to find the relationship between the number of claims in a given period and driver profile attributes. GLMs are used to model such relationships, assuming that the number of claims follows a Poisson or Negative Binomial distribution [1]. This research investigates a Poisson regression to model claims frequency weekly, following the methodology posited by Guillen et al. [20].

Poisson regression is particularly suitable for modeling count data and is widely used in insurance for claim frequency modeling due to its simplicity and interpretability. The model assumes that the number of claims Y_i for a given insurance policy i over a period T_i, represented by the duration of the contract, follows a Poisson distribution. This assumption holds because insurance claims are rare events that occur independently.

The number of claims can be modelled through a vector of k risk factors $x_i=(x_{i1}, x_{i2}, \ldots, x_{ik})$. The parameter λ_i represents the expected claim frequency and is a function of the linear combination of risk factors $x_i^T\beta$, where β is the vector of parameters resulting from the Poisson model. Such a relationship is determined by (1). The exponential term gives the predicted claims number by the exposure unit, while the period T_i allows capturing different contract durations. In this research T_i denotes the number of weeks of the policy.

$$E(Y_i|x_i) = \lambda_i = exp(ln\, T_i + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}) \quad (1)$$
$$= T_i \cdot exp(\beta_1 x_{i1} + \ldots + \beta_k x_{ik})$$

Traditional insurance models generally collect the risk factor vector x_i when the policy is underwritten. In UBI models, this vector also encompasses attributes about the driving profile. Equation 2 rewrites (1) by using two risk factor sets – static and telematics-based – through vectors x_i and E_{it}. The telematics risk vector E_{it} represents the driver behavior events of the i^{th} policyholder at week t. However, the number of telematics observations is T_i, whereas the number of observations for the static risk factor is one per driver. Following [20], [24], a possible approach to overcome this challenge is replicating past claims data T_i times (i.e., $Y_i = Y_{it}$ and $T_i = T_{it}$ for every driver-week t).

$$E(Y_i|x_i) = T_i \cdot exp(x_i^T\beta) \cdot exp\left(E_{it}^T\alpha\right) \quad (2)$$

[1] The prefix refers to the respective aggregation operation: mean, proportion, sum. Proportions are given in the range 0-1 and represent the exposure to that condition within the week. Summations denote the total count of the respective event over the week.

The model is measured through several metrics. The mean Poisson deviance (MPD) represents the mean of Poisson unit deviances. The root-mean-squared error (RMSE) is also considered to compare the performance of previous work [37]. The Pearson's Chi-squared test compares the observed distribution of data and the expected distribution if the variables are independent [43]. These metrics are complemented by the Akaike information criterion (AIC), which measures the goodness-of-fit and penalizes large number of model parameters.

B. Claims occurrence probability modeling

The second approach predicts the claim probability for each driver. In contrast to the previous approach, the dependent variable represents whether a given driver had any claims in the insurance period. Four machine learning algorithms are considered through their Python implementations: Logistic Regression, Support Vector Machine, Random Forest, and XGBoost. These models were chosen for their distinct characteristics and prevalence in similar works, as outlined in the Related Work section. Additionally, TabNet, a deep learning architecture specifically designed for tabular data, is included for a benchmark comparison against deep learning techniques.

Logistic Regression, grounded in maximum likelihood estimation, provides straightforward interpretations of coefficients. This model uses a logit link function to determine the conditional probability of a claim given the linear combination of risk factors. It serves as a reference model due to its traditional use in insurance models [36], [44]. Support Vector Machine (SVM) constructs hyperplanes in a high-dimensional space to classify observations based on a vector of risk factors. As a non-probabilistic model, claim occurrence probabilities are derived using Platt scaling [45], [46]. Random Forest leverages an ensemble of decision trees, enhancing prediction accuracy and robustness by introducing randomness in the tree growth process, resulting in independent predictors making uncorrelated errors [47]. XGBoost is built on the theory of gradient boosting, iteratively improving model accuracy by optimizing residual errors [48]. It consistently scores among the top performers in similar works. TabNet employs sequential attention mechanisms to identify and focus on the most relevant features at each decision step [49]. It has achieved comparable results to XGBoost in previous research, demonstrating its predictive performance in risk assessment [38].

Given a set of driver-week vectors x_{it}, the models aim to learn the model hypothesis $h(x_{it})$ that predicts the probability of having a claim in the insurance period (two years). Scikit-learn's MaxAbsScaler is implemented to scale the input vector.

An inner loop of 5-fold stratified group cross-validation is employed to select the best hyperparameters, creating a nested cross-validation framework in conjunction with the outer loop used to evaluate model performance. The primary metrics for choosing the optimal models are Log Loss and the area under the receiver operating characteristic curve (ROC AUC), which aligns with previous research [37]. The former compares predicted probabilities with ground truth classes, where low values represent good predictions. The ROC AUC measures the probability that a randomly chosen driver with claims is ranked higher than a randomly chosen driver without claims [50]. The models implement a balanced strategy for setting class weights needed for the imbalanced nature of the data.

C. From risk assessment to insurance premiums

1) Premium based on claims frequency

The coefficients resulting from (2) are the basis for the weekly premium computation. The proposed premium, detailed in (3), is given by the expected claims frequency multiplied by the expected claim cost, which for simplicity is the average cost C. Equation 4 separates the telematics risk factor vector E_{it} into B_{it} and C_{it} to represent behavioral and contextual factors. Thus, the weekly premium is composed by base, behavioral, and contextual premiums. The upper bound of a linear rate approximation is used to penalize event counts instead of a percent increase of the base premium.

$$\begin{aligned} P_{it} &= C \cdot T_i \cdot \exp(x_i{}^T \beta) \cdot \exp\left(E_{it}{}^T \alpha\right) \\ &= P_{base-i} \cdot \exp\left(E_{it}{}^T \alpha\right) \end{aligned} \tag{3}$$

$$\begin{aligned} P_{it} &= P_{base-i} \cdot exp\left(B_{it}{}^T \alpha\right) \cdot exp\left(C_{it}{}^T \gamma\right) \\ &\approx P_{base-i} \cdot \left(1 + B_{it}{}^T \alpha + C_{it}{}^T \gamma\right) \\ &\leq P_{base-i} + B_{it}{}^T \alpha_{max} + C_{it}{}^T \gamma_{max} \end{aligned} \tag{4}$$

2) Premium based on claims occurrence probability

The second weekly premium approach is estimated by multiplying the average claim cost by the predicted probability of having a claim and dividing by 104 weeks (i.e., two years). Equation 5 describes the weekly premium of driver i at week t, where C is the average cost of a claim in the training set, $h(x_{it})$ is the model output, and T is the insurance period.

$$P_{it} = C \cdot \frac{h(x_{it})}{T} \tag{5}$$

The outputs of machine learning models are processed through SHAP to interpret the predictions. SHAP is a model-agnostic technique based on cooperative game theory with Shapley values [14]. It aims to fairly distribute the contribution of each feature to the model predictions through an additive feature attribution method where the most important features receive the highest absolute Shapley value. After considering all possible feature combinations, this value represents a feature's average expected marginal contribution to the model output. The Shapley value of feature j is obtained through (6), where S is the subset of features of the entire set N; v is a characteristic function that assigns values to feature subsets, and $v(S)$ describes the total expected sum of contributions that the features belonging to S can obtain by cooperation.

$$\Phi_j(v) = \sum_{S \subseteq N\{j\}} \frac{|S|!\ (n - |S| - 1)!}{n!} [v(S \cup \{j\}) - \ v(S)] \tag{6}$$

V. Results and Discussion

A. Claims frequency modeling

The Poisson regression model for the weekly rate of at-fault claims, detailed by (2), estimates the parameters listed in Table II. There are no statistically significant differences between the

model with all attributes and with only traditional telematics. The AIC values show slightly better performance of the former. The results are similar to those of the models in [37], which reported slight variance in the RMSE for Poisson GLM with only traditional attributes and adding driving behavior. Pearson's chi-squared tests show no evidence of a lack of fit.

TABLE II
POISSON MODELS FOR THE WEEKLY RATE OF CLAIMS

Attribute	Estimate[2]	
	All attributes	Traditional telematics
const	-5.2555***	-4.81249***
engine_capacity	-0.86439***	-1.00335***
prop_road_quality_moderate	-0.35186	
prop_slope_flat	1.51666**	
prop_motorway	2.73525	
prop_rural	2.77197***	
prop_more_than_one_lane	0.81547*	
prop_clear_weather	0.21007	
prop_congested	4.98569***	
mean_speed_limit	-0.03271	
mean_weather_temperature	-0.02009*	
mean_weather_wind_speed	0.00221	
mean_weather_visibility	-0.00004	
sum_roundabout	-0.0066***	
sum_traffic_signal	0.00502***	
sum_stop_sign	-0.03172***	
sum_yield_sign	0.01185*	
sum_pedestrian_crossing_sign	0.00974	
sum_animal_crossing_sign	0.00508	
sum_speeding_serious	0.00143*	0.00078*
sum_harsh_acceleration	-0.03936***	-0.03826***
sum_harsh_braking	0.02004	0.01575
sum_forward_collision	0.00062	
sum_driver_inattention	0.00078*	
sum_too_close_distance	0.00001	
sum_lane_departure	0.00094**	
sum_driver_making_calls	-0.00832	
sum_driver_smoking	-0.00071	
sum_fatigue_driving	0.00043	
total_distance	0.00042*	0.00025*
Model performance with 5-fold cross-validation		
Mean Poisson deviance	0.62 (std: 0.11)	0.60 (std: 0.11)
RMSE	0.39 (std: 0.06)	0.38 (std: 0.06)
Goodness-of-fit (Chi-squared)	8,831 (std: 469)	8,789 (std: 303)
Akaike information criterion	8,468 (std: 456)	8,666 (std: 464)

The estimates reflect the mean coefficient over the five folds. Several contextual and behavioral attributes have significant effects, denoted with the * symbol, and latent interactions might affect the coefficients. For instance, while *prop_motorway* presents a high estimate, the prediction is influenced by *mean_speed_limit*, which has a negative coefficient. As expected, the *total distance* and heavy traffic conditions (i.e., *prop_congested)* increase the claim frequency, aligning with the literature. Weather conditions do not present considerable effects on claim predictions, in contrast to [51], who found that windy conditions increase the expected frequency for drivers in Spain. The only weather-related attribute with a significant effect is the temperature, which decreases the predicted claims frequency as the temperature increases.

As for driver behavior attributes, *sum_speeding_serious*, *sum_driver_inattention*, and *sum_lane_departure* report significant values with positive coefficients. This finding indicates their effect in representing driver negligence and risk appetite, supporting road safety efforts in avoiding speeding violations and distraction. Other attributes with positive coefficients encompass *forward collision*, *too close distance*, and *harsh braking*, although lacking significant effects. The negative values observed for *harsh acceleration* might be due to the expertise of the commercial drivers.

B. Claims occurrence probability modeling

Table III details the comparison of model performance for claims occurrence probability resulting from nested Cross-Validation. XGBoost reported the lowest log loss, and Random Forest reported the best ROC AUC. The former indicates that XGBoost assigns more probability to profiles with claims and less probability to profiles without claims (i.e., safe drivers), which is the objective for setting technical premiums. The ROC AUC values resemble the claims probability values reported by those of [37], where the best model achieved an ROC AUC of 0.613. The benchmark established by Logistic Regression shows that this traditional modeling achieves competitive results while providing straightforward interpretation. TabNet achieved competitive results, similar to XGBoost and Random Forest, with lower tuning efforts, as highlighted in [38].

Incorporating ADAS and context reduces the log loss for XGBoost, going from 0.59 (std: 0.03) to 0.51 (std: 0.03). However, this result is not consistent with the other models and there are no statistically significant differences in ROC AUC values when considering the model without ADAS and context. This finding might be due to the professional nature of the drivers, with considerable expertise and levels of safe driving set by the fleet company. Moreover, as part of their operations, they might be exposed to various driving contexts without a particular setting that distinguishes them.

Fig. 2 presents the contribution of features for model predictions. Each point represents a particular instance of the training set (i.e., a weekly profile), where the color serves to identify the feature value, from low (blue) to high (red). The value indicates whether the feature decreases or increases the claim probability on a particular instance, given by the SHAP. High SHAP magnitudes represent high feature impacts on the predicted probability, where positive values increase the probability and negative ones decrease it. The attributes are sorted by their global importance, given by the mean absolute of a feature's SHAP values.

[2] Stars refer to p-values (***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.1$)

TABLE III
CLAIMS OCCURRENCE PROBABILITY MODELS COMPARISON WITH 5-FOLD CROSS-VALIDATION

Model	Log Loss	ROC AUC
Performance with all attributes		
XGBoost	0.51 (std: 0.03)	0.58 (std: 0.07)
SVC	0.52 (std: 0.10)	0.54 (std: 0.04)
TabNet	0.60 (std: 0.42)	0.58 (std: 0.04)
Random Forest	1.02 (std: 0.10)	0.59 (std: 0.07)
Logistic Regression	1.52 (std: 0.07)	0.56 (std: 0.05)
Performance without ADAS and context		
XGBoost	0.59 (std: 0.03)	0.61 (std: 0.05)
SVC	0.44 (std: 0.07)	0.48 (std: 0.05)
TabNet	0.47 (std: 0.11)	0.60 (std: 0.10)
Random Forest	1.29 (std: 0.13)	0.63 (std: 0.06)
Logistic Regression	1.55 (std: 0.08)	0.58 (std: 0.08)

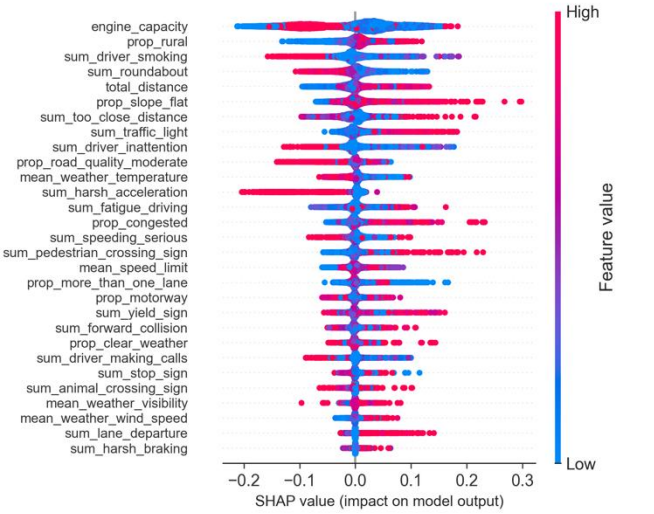

Fig. 2. Feature importance of the XGBoost model. The figure shows SHAP values for instances of the training set, where each point represents one weekly profile. The color scale represents the feature value from low (blue) to high (red). The y-axis is sorted according to the feature importance.

The SHAP analysis shows that many features do not linearly affect the model output. For instance, low engine capacity values have positive and negative impacts, which depend on interactions with other features. Aligned with the results of the Poisson model, moderate road qualities, roundabouts, and high temperatures tend to reduce the claim probability. Rural areas, distance traveled, traffic lights, congested traffic conditions, and flat roads positively impact claims prediction. In contrast to the Poisson model, high-speed limits increase the risk. As for driver behavior events, the ones with the most direct impacts on increased risk probabilities are too-close distance, fatigue, and lane departure. Harsh acceleration indicates reductions in the claim probability, which may be due to driver expertise, as [20] argued. Smoking and inattention tend to reduce the model outputs. Other behavioral events, including severe speeding and forward collisions, have mixed effects on the model output and may vary according to the interactions with other attributes.

C. From risk assessment to weekly insurance premiums

1) Premium distribution

The studied premiums represent the technical premium without incorporating additional insurance components such as operational costs. Fig. 3 compares the weekly premiums from the two approaches using an evaluation set of 71 drivers listed over the vertical axis. Each bar represents the interquartile range of the driver's weekly premium, where the darker color indicates whether the driver has any claim. The figure shows that the Poisson model has a wider premium spread, although three drivers with claims have mean premiums below the global mean (€5.87). In contrast, the XGBoost model presents a more compacted distribution, but only two drivers out of 11 have premiums below the global mean (€3.86) Furthermore, XGBoost gives lower variability concerning per-driver weekly premiums, identified by shorter bars.

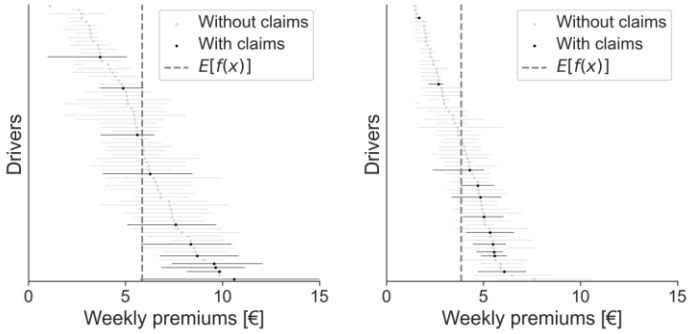

Fig. 3. Comparison of weekly premium distribution per driver using Poisson and XGBoost models on a test set. Each bar represents the weekly premium distribution per driver, given by the interquartile range, where the central point is the driver's mean weekly premium.

The impact on the whole driver portfolio is detailed in Table IV, showing the mean weekly premium per driver and its respective annual value. The training and evaluation sets correspond to one fold of the cross-validation process. The former has 283 drivers and 49 claims, whereas the latter has 71 drivers and 13 claims. The difference in the resulting premiums arises from the distinct approaches taken by the two models in both their hypothesis and premium computation methods. The Poisson model tends to provide more consistent and stable premium estimates across the train and test sets. On the other hand, the XGBoost model assigns high-risk probabilities to driver profiles with claims in the training set. This results in heavily penalizing these profiles, leading to greater variability in the premium estimates. However, this penalization is reduced in the test set due to the lower model outputs, indicating that the model is better at identifying high-risk profiles during training but more conservative during testing. The GLM also allows for the disaggregation of the resulting weekly premium into three premium components. Based on the claims history of such drivers, with a mean claim cost C of €2,316, the mean base premium is €4.06. This value represents the initial rate without considering the driving context or behavioral events and depends on the information collected at the creation of the

policy. Then, such a premium is adjusted according to the number of behavioral and contextual factors.

2) Risk profile examples for low and high risk

This section compares the billing processes using two different driver profiles taken from Fig. 3, one for a risky driver (i.e., with claims in the claims history) and another for a safe driver (i.e., without claims and with a low mean premium). The risky driver experienced more driver behavior events, particularly driver inattention and forward collision, drove significantly more on rural roads (i.e., 27% of the weekly distance, compared to 4% of the safe driver), and drove 664 km in contrast to 192 km. Applying the linear combination (4) with the Poisson coefficients of Table II, these driving profiles have a weekly premium of €5.60 for the risky driver and €2.56 for the safe driver.

TABLE IV
WEEKLY AND YEARLY DRIVER PREMIUMS

Observation	Fleet	Poisson	XGBoost
Mean weekly premium per driver (std dev)	Train set	€5.87 (€2.40)	€9.47 (€2.04)
	Validation set	€5.90 (€2.56)	€3.85 (€1.51)
Mean yearly premium per driver (std dev)	Train set	€306.80 (€105.8)	€494.31 (€365.3)
	Validation set	€307.92 (€107.8)	€200.57 (€74.5)

With the claims probability model, the risky profile has a weekly premium of €6.02 and the safe profile has a premium of €1.49. Fig. 4 shows the composition of such predictions using SHAP's force plots. The plots indicate how different features of the weekly profile increase (red) or decrease (blue) the probability of having a claim. For the risky driver's profile (top), the main features are the vehicle's engine capacity, the proportion of rural roads, total distance and moderate road qualities, and events involving inattention and smoking. In contrast, for the safe driver's profile (bottom), low values of the mean speed limit and proportion of rural roads and the engine capacity decrease the predicted probability.

One of the principal differences between claims frequency and claims probability is that the feature contribution of the former is fixed on the model's coefficients. In the latter, it varies depending on feature interactions. That variability stems from the mathematical characteristics of SHAP, where feature contributions are computed considering feature coalitions. However, the static nature of the GLM model makes it more suitable for regulatory requirements and transparency within the insurer ratemaking process.

3) Using both approaches in the insurance lifecycle

While machine learning models enable an enhanced segmentation of the fleet's risk, their application in the insurance domain is challenged by insurance regulations. SHAP could solve this challenge by allowing explanations of model predictions to the customer, although with considerable effort. This article proposes the complementary use of both approaches, where complex models and XAI are used internally to enhance and validate risk segmentation.

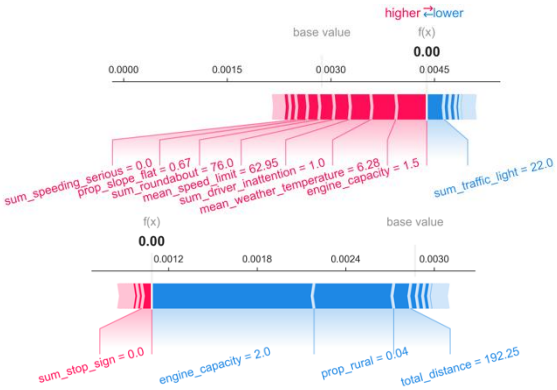

Fig. 4. Predicted claims probability for risky (top) and safe (bottom) weekly driver profiles. The figure gives the contribution of each feature to the predicted probability of having any claims in a week.

Fig. 5 illustrates the insurance lifecycle incorporating the two modeling techniques. After issuing the base policy with traditional insurance factors, the vehicle reports telematics data, which is contextualized and aggregated into the driver's weekly profile. Consequently, the premium is updated with the GLM coefficients, satisfying the interpretability requirements for motor insurers. In parallel, the insurer can use the machine learning-based process internally to validate the portfolio risk segmentation and exploit feature interactions through XAI. For example, model interpretation analyses could highlight that speeding patterns should be coupled with exposure to different road types and use such interaction in the GLM for the pricing update as a separate attribute.

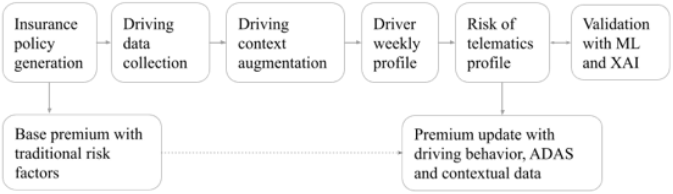

Fig. 5. Proposed risk assessment for the usage-based insurance lifecycle. Abbreviations: Machine Learning (ML), Explainable Artificial Intelligence (XAI).

VI. CONCLUSION

This article proposes a driver risk assessment methodology integrating risk indicators from ADAS, the driving context, and vehicle dynamics. To the authors' knowledge, this is the first research exploring insurance pricing of a commercial fleet equipped with warning-based ADAS and contextual attributes (e.g., road types, weather, and traffic conditions). Such a dynamic risk assessment enables a comprehensive periodic premium where drivers are charged according to how much, how safe, and where they drive, incentivizing safe driving through the insurance bill.

The proposed methodology is designed to assist stakeholders in the motor insurance industry by enabling fair premiums through incorporating emerging driving assistance technologies and driving context. Insurers can implement these findings by augmenting their existing telematics data with contextualization services and beginning to collect information about ADAS usage. With their extensive data records, insurers could replicate the driver aggregations according to their needs and target dynamic billing schemes, whether weekly or

monthly. The Poisson and XGBoost modeling results show consistent attribute impacts on driver risk, given by the expected claim frequency or probability, respectively. Moderate road qualities, roundabouts, and high temperatures tend to reduce driver risk, while rural areas, distance traveled, traffic lights, congested traffic conditions, and flat roads are positively associated with higher risk levels. Results showed that incorporating ADAS and contextual information improves the log loss of XGBoost, although there were no significant differences concerning the ROC AUC or the Poisson model. This finding might be due to the professional nature of the drivers, with considerable levels of safe driving set by the fleet company and lack of clear differences between them in their operational driving contexts. Thus, future work could explore another fleet of drivers. Moreover, complementing modeling efforts with explainable AI methods, such as SHAP, can support insurers and policymakers in using state-of-the-art models that meet interpretability requirements. By leveraging these insights through periodic premiums, insurers can offer tailored policies and incentivize safe driving behaviors, while policymakers can develop informed regulations that encourage the use of ADAS and contextual data in risk assessment.

This study has its limitations. The dataset used is specific to a commercial fleet in Ireland, which may restrict the applicability of the findings to other regions or driver types. Additionally, the study focuses on warning-based ADAS captured by dashcams connected to the vehicle; future research could delve into the impact of integrated ADAS and more autonomous driving systems (e.g., autopilot). An important assumption in our methodology involves replicating claims over weekly profiles to identify risky driving behaviors. This is due to the limited timeframe of the study and the number of drivers relative to the infrequent occurrence of at-fault claims. Future studies could address this by including more drivers and extending the study period, thereby allowing for a more comprehensive assessment of the model's scalability. Finally, there is a need for further research to evaluate the long-term effectiveness of dynamic billing schemes on driver behavior. These areas present opportunities for future research and development, which could enhance the applicability and effectiveness of the proposed methodology.

REFERENCES

[1] A. Charpentier, “Statistique de l’assurance,” *Univ. Rennes 1 Univ. Montr. 2010 Pp133*, vol. 3rd cycle, p. 114, Dec. 2010.

[2] M. Guillen, J. P. Nielsen, M. Ayuso, and A. M. Pérez-Marín, “The use of telematics devices to improve automobile insurance rates,” *Risk Anal.*, vol. 39, no. 3, pp. 662–672, 2019, doi: 10.1111/risa.13172.

[3] D. I. Tselentis, G. Yannis, and E. I. Vlahogianni, “Innovative motor insurance schemes: A review of current practices and emerging challenges,” *Accid. Anal. Prev.*, vol. 98, pp. 139–148, Jan. 2017, doi: 10.1016/j.aap.2016.10.006.

[4] M. M. Antony and R. Whenish, “Advanced Driver Assistance Systems (ADAS),” in *Automotive Embedded Systems: Key Technologies, Innovations, and Applications*, M. Kathiresh and R. Neelaveni, Eds., in EAI/Springer Innovations in Communication and Computing. , Cham: Springer International Publishing, 2021, pp. 165–181. doi: 10.1007/978-3-030-59897-6_9.

[5] D. W. Eby *et al.*, “Prevalence, attitudes, and knowledge of in-vehicle technologies and vehicle adaptations among older drivers,” *Accid. Anal. Prev.*, vol. 113, pp. 54–62, Apr. 2018, doi: 10.1016/j.aap.2018.01.022.

[6] J. M. Scanlon, R. Sherony, and H. C. Gabler, “Injury mitigation estimates for an intersection driver assistance system in straight crossing path crashes in the United States,” *Traffic Inj. Prev.*, vol. 18, no. sup1, pp. S9–S17, May 2017, doi: 10.1080/15389588.2017.1300257.

[7] L. Masello, G. Castignani, B. Sheehan, F. Murphy, and K. McDonnell, “On the road safety benefits of advanced driver assistance systems in different driving contexts,” *Transp. Res. Interdiscip. Perspect.*, vol. 15, p. 100670, Sep. 2022, doi: 10.1016/j.trip.2022.100670.

[8] R. R. Knipling *et al.*, “Assessment of IVHS countermeasures for collision avoidance : rear-end crashes,” DOT-HS-807-995, May 1993. Accessed: Dec. 15, 2022. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/4276

[9] A. Bamney, S. Sonduru Pantangi, H. Jashami, and P. Savolainen, “How do the type and duration of distraction affect speed selection and crash risk? An evaluation using naturalistic driving data,” *Accid. Anal. Prev.*, vol. 178, p. 106854, Dec. 2022, doi: 10.1016/j.aap.2022.106854.

[10] T. Seacrist, E. C. Douglas, C. Hannan, R. Rogers, A. Belwadi, and H. Loeb, “Near crash characteristics among risky drivers using the SHRP2 naturalistic driving study,” *J. Safety Res.*, vol. 73, pp. 263–269, Jun. 2020, doi: 10.1016/j.jsr.2020.03.012.

[11] S. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey, “The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data,” Apr. 2006, Accessed: Mar. 29, 2022. [Online]. Available: https://vtechworks.lib.vt.edu/handle/10919/55090

[12] L. Aarts and I. van Schagen, “Driving speed and the risk of road crashes: A review,” *Accid. Anal. Prev.*, vol. 38, no. 2, pp. 215–224, Mar. 2006, doi: 10.1016/j.aap.2005.07.004.

[13] P. Regulation, “Regulation (EU) 2016/679 of the European Parliament and of the Council,” *Regul. Eu*, vol. 679, p. 2016, 2016.

[14] S. Lundberg, “Shap.” Accessed: Jan. 05, 2023. [Online]. Available: https://github.com/slundberg/shap

[15] K. Lin and Y. Gao, “Model interpretability of financial fraud detection by group SHAP,” *Expert Syst. Appl.*, vol. 210, p. 118354, Dec. 2022, doi: 10.1016/j.eswa.2022.118354.

[16] I. Chang, H. Park, E. Hong, J. Lee, and N. Kwon, “Predicting effects of built environment on fatal pedestrian accidents at location-specific level: Application of XGBoost and SHAP,” *Accid. Anal. Prev.*, vol. 166, p. 106545, Mar. 2022, doi: 10.1016/j.aap.2021.106545.

[17] A. B. Ellison, M. C. J. Bliemer, and S. P. Greaves, “Evaluating changes in driver behaviour: A risk profiling approach,” *Accid. Anal. Prev.*, vol. 75, pp. 298–309, Feb. 2015, doi: 10.1016/j.aap.2014.12.018.

[18] S. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey, “Comparing real-world behaviors of drivers with high versus low rates of crashes and near crashes,” 2009.

[19] C. Ryan, F. Murphy, and M. Mullins, “Semiautonomous Vehicle Risk Analysis: A Telematics-Based Anomaly Detection Approach,” *Risk Anal.*, vol. 39, no. 5, pp. 1125–1140, 2019, doi: 10.1111/risa.13217.

[20] M. Guillen, J. P. Nielsen, and A. M. Pérez-Marín, “Near-miss telematics in motor insurance,” *J. Risk Insur.*, vol. 88, no. 3, pp. 569–589, 2021, doi: 10.1111/jori.12340.

[21] S. E. Lee, B. G. Simons-Morton, S. Klauer, M. C. Ouimet, and T. A. Dingus, “Naturalistic assessment of novice teenage crash experience,” *Accid. Anal. Prev.*, vol. 43, no. 4, pp. 1472–1479, Jul. 2011, doi: 10.1016/j.aap.2011.02.026.

[22] B. A. Jonah, R. Thiessen, and E. Au-Yeung, “Sensation seeking, risky driving and behavioral adaptation,” *Accid. Anal. Prev.*, vol. 33, no. 5, pp. 679–684, Sep. 2001, doi: 10.1016/S0001-4575(00)00085-3.

[23] J.-P. Boucher, S. Côté, and M. Guillen, “Exposure as Duration and Distance in Telematics Motor Insurance Using Generalized Additive Models,” *Risks*, vol. 5, no. 4, p. 54, Dec. 2017, doi: 10.3390/risks5040054.

[24] Y.-L. Ma, X. Zhu, X. Hu, and Y.-C. Chiu, “The use of context-sensitive insurance telematics data in auto insurance rate making,” *Transp. Res. Part Policy Pract.*, vol. 113, pp. 243–258, Jul. 2018, doi: 10.1016/j.tra.2018.04.013.

[25] J. Jun, R. Guensler, and J. Ogle, “Differences in observed speed patterns between crash-involved and crash-not-involved drivers: Application of in-vehicle monitoring technology,” *Transp. Res. Part C Emerg. Technol.*, vol. 19, no. 4, pp. 569–578, Aug. 2011, doi: 10.1016/j.trc.2010.09.005.

[26] X. Hu, X. Zhu, Y.-L. Ma, Y.-C. Chiu, and Q. Tang, “Advancing usage-based insurance – a contextual driving risk modelling and analysis approach,” *IET Intell. Transp. Syst.*, vol. 13, no. 3, pp. 453–460, 2019, doi: 10.1049/iet-its.2018.5194.

[27] Y. Peng, G. Song, M. Guo, L. Wu, and L. Yu, “Investigating the impact of environmental and temporal features on mobile phone distracted

driving behavior using phone use data," *Accid. Anal. Prev.*, vol. 180, p. 106925, Feb. 2023, doi: 10.1016/j.aap.2022.106925.
[28] NHTSA, "Overview of Motor Vehicle Crashes in 2019," Overview of Motor Vehicle Crashes in 2019. Accessed: Mar. 29, 2022. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813060
[29] UK Department for Transport, "Reported road accidents, vehicles and casualties tables for Great Britain," GOV.UK. Accessed: Mar. 29, 2022. [Online]. Available: https://www.gov.uk/government/statistical-data-sets/reported-road-accidents-vehicles-and-casualties-tables-for-great-britain
[30] T. Jannusch, D. Shannon, M. Völler, F. Murphy, and M. Mullins, "Smartphone Use While Driving: An Investigation of Young Novice Driver (YND) Behaviour," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 77, pp. 209–220, Feb. 2021, doi: 10.1016/j.trf.2020.12.013.
[31] T. A. Ranney, "Driver Distraction: A Review of the Current State-of-Knowledge," Art. no. HS-810 787, Apr. 2008, Accessed: Aug. 21, 2022. [Online]. Available: https://trid.trb.org/view/868221
[32] B. Öz, T. Özkan, and T. Lajunen, "Professional and non-professional drivers' stress reactions and risky driving," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 13, no. 1, pp. 32–40, Jan. 2010, doi: 10.1016/j.trf.2009.10.001.
[33] D. Shannon, T. Jannusch, F. David-Spickermann, M. Mullins, M. Cunneen, and F. Murphy, "Connected and autonomous vehicle injury loss events: Potential risk and actuarial considerations for primary insurers," *Risk Manag. Insur. Rev.*, vol. 24, no. 1, pp. 5–35, 2021, doi: 10.1111/rmir.12168.
[34] J. Kanet and G. Hinton, "True Impact of ADAS Features on Insurance Claim Severity Revealed," LexisNexis Risk Solutions. Accessed: May 19, 2024. [Online]. Available: https://risk.lexisnexis.com/insights-resources/white-paper/true-impact-of-adas-features-on-insurance-claim-severity-revealed
[35] SwissRe, "ADAS Risk Score." 2019. Accessed: May 19, 2024. [Online]. Available: https://www.swissre.com/dam/jcr:217d3421-a203-471e-b296-91ac5e97ee26/ADAS_PitchBook_4-3%20high%20res.pdf
[36] J. Paefgen, T. Staake, and F. Thiesse, "Evaluation and aggregation of pay-as-you-drive insurance rate factors: A classification analysis approach," *Decis. Support Syst.*, vol. 56, pp. 192–201, Dec. 2013, doi: 10.1016/j.dss.2013.06.001.
[37] Y. Huang and S. Meng, "Automobile insurance classification ratemaking based on telematics driving data," *Decis. Support Syst.*, vol. 127, p. 113156, Dec. 2019, doi: 10.1016/j.dss.2019.113156.
[38] K. McDonnell, F. Murphy, B. Sheehan, L. Masello, and G. Castignani, "Deep learning in insurance: Accuracy and model interpretability using TabNet," *Expert Syst. Appl.*, vol. 217, p. 119543, May 2023, doi: 10.1016/j.eswa.2023.119543.
[39] X. Wen, Y. Xie, L. Jiang, Y. Li, and T. Ge, "On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development," *Accid. Anal. Prev.*, vol. 168, p. 106617, Apr. 2022, doi: 10.1016/j.aap.2022.106617.
[40] H.-J. Li, X.-G. Luo, Z.-L. Zhang, W. Jiang, and S.-W. Huang, "Driving risk prevention in usage-based insurance services based on interpretable machine learning and telematics data," *Decis. Support Syst.*, vol. 172, p. 113985, Sep. 2023, doi: 10.1016/j.dss.2023.113985.
[41] Motion-S, "Augmenting Locations In Real-Time," Motion-S. Accessed: Mar. 08, 2022. [Online]. Available: https://developer.motion-s.com/docs/augmenting-locations-in-real-time
[42] Sécurité Routière, "Réglementation de la vitesse au volant." Accessed: Mar. 10, 2023. [Online]. Available: https://www.securite-routiere.gouv.fr/reglementation-liee-aux-risques/reglementation-de-la-vitesse-au-volant
[43] N. S. Turhan, "Karl Pearson's Chi-Square Tests.," *Educ. Res. Rev.*, vol. 16, no. 9, pp. 575–580, 2020.
[44] L. Brühwiler, C. Fu, H. Huang, L. Longhi, and R. Weibel, "Predicting individuals' car accident risk by trajectory, driving events, and geographical context," *Comput. Environ. Urban Syst.*, vol. 93, p. 101760, Apr. 2022, doi: 10.1016/j.compenvurbsys.2022.101760.
[45] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
[46] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classif.*, vol. 10, no. 3, pp. 61–74, 1999.
[47] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[48] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
[49] S. Ö. Arik and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning," *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 8, Art. no. 8, May 2021, doi: 10.1609/aaai.v35i8.16826.
[50] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
[51] J. Reig Torra, M. Guillen, A. M. Pérez-Marín, L. Rey Gámez, and G. Aguer, "Weather Conditions and Telematics Panel Data in Monthly Motor Insurance Claim Frequency Models," *Risks*, vol. 11, no. 3, Art. no. 3, Mar. 2023, doi: 10.3390/risks11030057.

Leandro Masello received his Ph.D. degree from the University of Limerick in 2023 as part of the Emerging Risk Group (ERG) and Motion-S. His industrial research project centered on risk assessment and data-driven pricing models for connected and automated vehicles. Leandro received the Engineering degree in Computer Engineering from the University of Buenos Aires in 2019, with a thesis in objective driver risk profiling using road safety statistics for insurance telematics applications. His research interests include actuarial sciences, risk assessment, machine learning, and road safety.

Barry Sheehan is an associate professor of risk management and insurance at the Kemmy Business School at the University of Limerick. He is Head of the Department of Accounting and Finance and program director for a cluster of award-winning interdisciplinary programs, including the MSc in Machine Learning for Finance. With a professional background in actuarial science, his research uses machine-learning techniques to estimate the changing risk profile produced by emerging technologies. He is a senior researcher within the Emerging Risk Group (ERG) and Lero, which has long-established expertise in insurance and risk management and has continued success within large research consortia, including EU H2020, Interreg Europe, and Science Foundation Ireland research projects.

German Castignani received a master's degree in computer science engineering from the University of Buenos Aires in Argentina in 2009 and a Ph.D. in computer science from Institut Mines-Télécom in France, in 2012. He has served as a Research Associate at the University of Luxembourg / SnT where he has been involved in several research projects and related publications in the fields of vehicular networking, mobile sensing and mobility management. Dr. Castignani has also co-founded and led Motion-S, a spin-off of his research work at the University of Luxembourg in mobility and driving behavior analysis, delivering services in the area of data-driven solutions for insurance and automotive industries. In 2023, he joined the Luxembourg Institute of Science and Technology (LIST), where he leads the Digital Twin Innovation Centre and the AI and Data Analytics (AIDA) platform.

Montserrat Guillen is full professor and Director of the Riskcenter, University of Barcelona, the Research Group on Risk in Insurance and Finance. Department of Econometrics, Statistics and Applied Economics. ICREA Academia 2011 and 2018 awardee. Montserrat Guillén was born in Barcelona in 1964. She received a Master of Science in Mathematics and Mathematical Statistics in 1987 and a Ph.D. degree in Economics from UB in 1992. She received a MSc degree in Data Analysis from the University of Essex (United Kingdom).She is currently Honorary Visiting Professor in the Faculty of Actuarial Science and Insurance at the Bayes Business School, City, University of London. She was Visiting Research faculty at the University of Texas at Austin (USA) in 1994. She was also Visiting Professor at the University of Paris II. Her research focuses on actuarial statistics and quantitative risk management.

Finbarr Murphy is Executive Dean and Professor in Quantitative Finance and Emerging Risk at the University of Limerick. A computer engineering graduate, Finbarr worked for over ten years in investment banking before returning to academia and completing his Ph.D. degree in 2010. A former Fulbright Scholar, Finbarr has delivered numerous guest lectures across the globe. His research interests include quantitative finance and more recently, emerging technological risk using machine learning tools. He is currently engaged in several EU H2020 and Irish Science Foundation Ireland (SFI) projects.