Short-term future prediction of driving risk using XGBoost with Personal, Physiological, Biological, and Driving Behavior Data

# Short-term future prediction of Driving risk using XGBoost with Personal, Physiological, Biological, and Driving Behavior Data

Michele Guagnano[1], Yecan Wang[2], Hiroki Taniguchi[2], Nozomi Nagatani[2], Hiroshi Ono[2], Satoru Shinkawa[2], Sumie Miyamoto[3], Katsuhito Fujiu[3], Madoka Takai[4], Massimo Violante[1], Shigenobu Mitsuzawa[2]

[1]*Dept. of Control and Computer Engineering*
*Politecnico di Torino*
Turin, Italy
[2]*Honda Motor R&D Co., Ltd.*
Wako, 351-0113, Saitama, Japan
[3]*Department of Medical the*
*University of Tokyo*
113-8654 Tokyo, Japan
[4]*Department of Bioengineering*
*the University of Tokyo* 113-8654 Tokyo, Japan

***Abstract*—This study investigated the use of a combination of driving, physiological, and personal features to predict unsafe driving behaviors with large advances. The 54 subjects involved were monitored for one week before the beginning of the experiment. Their personal information was collected, as well as their sleeping quality, stress state, and daily activity. Afterwards, they underwent 12 driving simulation sessions, each divided into 4 laps. During sessions on the driving simulator, saliva samples, physiological data, and driving-based measures were collected. The salivary cortisol level was measured before and after each lap, serving as a stress biomarker. An XGBoost model was trained and used to predict the driving safety level of a lap based on data from the previous lap (10 minutes earlier). Driving safety is categorized based on a 3-level scale. The model identified safe driving, moderately unsafe driving, and severely unsafe driving with precision of 98%, 81%, and 96%.**

***Index Terms*—driving risk prediction, in-cabin monitoring, wearable devices, cortisol, XGBoost**

## I. Introduction

According to the World Health Organization (WHO), around 1.2 million deaths and more than 25 million severe injuries occur every year due to road traffic crashes [1], [2]. These tragic events, which are more frequent in low and middle-income countries, have also a high economic impact on national health systems. There is a wide variety of accident causes; they can range from environmental factors to individual driver's characteristics [3]. The most common are

- Distracted driving, which can reduce the ability to react on time due to some sources of inattention, like the smartphone.
- Impaired driving, which can be caused by alcohol, drugs, or fatigue. It leds to wrong judgements and high reaction time.
- Aggressive driving, like speeding and tailgating.
- Health conditions, like sleep disorders and cardiovascular problems.
- Emotional issues and stress.

As the mentioned factors can progressively lead to unsafe driving behaviors and then to a crash, it is important to recognize them with a certain time advance. Nowadays, Advanced Driver Assistance Systems (ADAS) are used to predict various potential collisions [4]. They catch information about risky conditions in the traffic environment, and alerts about imminent risks as early as possible. However, they mainly focus on external factors driven by vehicle dynamics and surrounding environments, and ignore the driver's state, leading to a reduction in risk prediction. In order to predict these factors, in-cabin monitoring has been employed. These systems consists in cameras and sensors inside the vehicle able to track facial and driving features. Furthermore, wearable devices such as smartwatches have proven to be valuable tools for monitoring a driver's physiological state [5] and can be used in combination with in-cabin monitoring systems. Moreover, Jafarpoor and Rahimi-Movaghar stated that demographics affected the number of traffic accidents [3]. Gender, age, driving experience, driver's physical and mental abilities and psychological factors like personality type, temperament, mood, and emotions, distraction by outside or inside stimuli, socioeconomic context and the individual income, sociocultural backgrounds, level of governance and law enforcement as well as internalization of legality and fidelity to the values of law in the society are all influences, they found to affect driving behavior. For this reason, we decided to take in account also personal information.

In our work, we monitored 54 subjects on a driving simulator, collecting eye tracking metrics, voice, driving metrics, smartwatch features, and personal data, and using XGBoost to predict unsafe behaviors. The subjects provided us with age, gender, and driving experience. Also, their sleep time, daily activity, and stress (from smartwatch) were monitored for one week. Rather than performing real-time predictions, we utilized data from one lap to predict the unsafe behavior in the following lap. Before and after each lap, saliva samples of the volunteers were collected for cortisol measurement, which is known as the ‘stress biomarker’. Cortisol is secreted by the hypothalamic pituitary-adrenal (HPA) axis, and regulates different physiological mechanisms. Our model, trained using features such as physiological states and driving performance, demonstrated promising results, particularly in identifying safe driving and severe unsafe driving behavior (respectively levels 0 and 2). However, there was a slight reduction in performance when predicting mid-level unsafe behavior (level 1), which was occasionally misclassified as level 0.


Fig. 1. Experimental setup

## II. Methodology

### A. Data Collection

A total of 54 subjects participated in the study. Their personal one-week information, such as sleep quality, stress index, and Heart Rate (HR), was consecutively collected before each test.. Then, they underwent 12 days of testing on a driving simulation, which replicated urban and mountain driving scenarios. In every test, they performed 4 laps, each lasting 10 minutes. Before and after every lap, saliva were sampling using collection vials for further cortisol level measurement. In scenarios, the features corresponding to potential causes of unsafe driving were collected as following:

- Eye-tracking metrics, which include the eyes’ position and the pupil dimension, are collected at 100Hz.
- Voice emotion, quantifying normal state, anxiety, and tiredness for each lap
- Driving metrics, such as the yaw, the speed, and the acceleration, were acquired at 100Hz.
- Smartwatch metrics, including Heart Rate (HR), Beat to Beat Interval (BBI), and 3-axis acceleration. BBI refers to the variation in time between successive heartbeats, also known as the RR interval, which is critical for understanding heart rate variability (HRV). The BBI provides a more detailed picture of the variability within an individual’s heart rate pattern. Cardiac features were extracted at 1Hz, and 3-axis accelerations were extrapolated at 25Hz.

The data acquired during one lap, combining with personal information collected in advance, were used to predict the driving behavior of the following one. The experimental setup, complete of driving simulator, eye-tracker, smartwatch, and salivary cortisol detector, is shown in Fig. 1.

### B. Data Preprocessing

Recordings with more than 20% of missing values in a lap were excluded from the analysis. From the remained ones, measures of different nature were considered. From the eyetracker, The pupil size was collected from eye-tracker due to its variation linked to the neural activity [6]. In particular, it is used as an indication of fatigue, attention load, and emotional arousal [7]. From the smartwatch, HR and BBI, which is commonly used to derive heart rate variability (HRV) metrics and reflect the influence of the autonomic nervous system (ANS) on cardiac activity, were collected [8]. In particular, pNN50 was calculated. This represents the percentage of successive RR intervals differing by more than 50 milliseconds and is a key index of HRV [9]. It specifically reflects the parasympathetic nervous system activity, and its decrease under stress or tiredness [10], [11]. Regarding driving features, car acceleration in x-axis, z-axis, and yaw were considered. Acceleration in the X-axis (the longitudinal acceleration) represents the change in the speed of vehicle in the forward or backward direction. This metric is also crucial for understanding how the driver controls the vehicle in terms of acceleration and braking. The acceleration in the Z-axis captures movements vertically, which helps assess the influence of road conditions (e.g., bumps, potholes, or sharp curves) on the vehicle’s motion, as well as the driver’s response to them [12]. Yaw refers to the rotational movement of the vehicle around its vertical axis, indicating the driver’s ability of maintaining control of the vehicle, particularly in dynamic driving situations, such as sharp turns, lane changes, or navigating curves [13].

To track emotional issues, the driver’s voice was recorded and analyzed for voice emotion recognition [14]. Thus, the level of normality, tiredness, and anxiety was quantified in each lap. In order to reduce dimensionality and extract meaningful emotional patterns, t-distribuited Stochastic Neighbor Embedding (t-SNE) was applied, resulting in a single feature called “combined voice”. t-SNE is a nonlinear dimensionality reduction technique particularly effective for visualizing highdimensional data and capturing subtle local structures and clusters, unlike linear methods such as principal component analysis (PCA) [15]. By applying t-SNE, the multi-dimensional emotional data were compressed into a single parameter per lap, providing a more manageable representation that retained the emotional context of the driver's voice. The combined voice feature was then used as one of the inputs to the machine learning model for the prediction of unsafe behavior, with the improvement of the model efficiency and a more simplified interpretation of the emotional influence on driving behavior.

In addition, personal information, particularly age, gender, driving age, daily driving frequency, and total sleep time, was also included as model inputs. Finally, the cortisol level before the beginning of the lap was obtained by competitive ELISA, a traditional methodology to measure biomarkers. Given that with factors such as emotion, stress, distraction, and fatigue, cortisol level was used as one essential feature.

The entire dataset was normalized through StandardScaler to ensure that every feature contributes equally.

*C. Defining Unsafe Behaviors*

Eboli, Mazzulla, and Pungillo [16] used an approach based on acceleration and speed to define unsafe driving behaviors. Two thresholds are derived from (1), where V is the speed, $a_x$ is the lateral acceleration, and $a_y$ is the longitudinal acceleration.

$$\sqrt{a_x^2 + a_y^2} = g \cdot [0.198 \cdot (\frac{V^2}{100}) - 0.592 \cdot (\frac{V}{100}) + 0.569] \quad (1)$$

An example of a driving performed in our test is shown in Fig. 2, where red lines are the thresholds and the points are instantaneous values of vehicle positioning in the plane (V, a). A specific moment is considered safe if the point falls within the thresholds, and unsafe vice versa.

Being the interval in the driving simulator 10 milliseconds, unsafe behaviors are classified as the proportion of unsafe driving behaviors within each one second. The goal of this work is to predict the safety level of the next step using data from the current lap. In this situation, real-time unsafe classification is not required, but a system able to label the entire lap instead. Eboli, Mazzulla, and Pungillo, based on their previously explained classification, proposed a method to label an entire driving session [17]. They calculated real-time safety levels and then classified the driving session as follows: Low risk, with less than 5% of points outside the thresholds; Medium risk, with between 5% and 8% points outside the thresholds; High risk, with above 8% of points outside the thresholds.

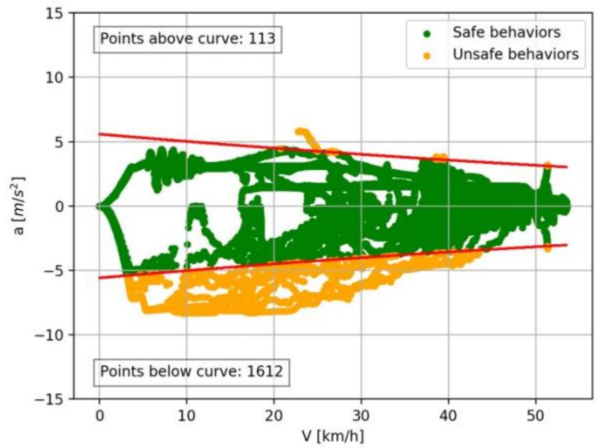


Fig. 2. Example of driving behavior definition

*D. Model Training*

Data collected from a specific lap was used to predict the behavior of the following lap through Extreme Gradient Boosting (XGBoost), a machine learning technique combining the output of decision trees to develop a more reliable model. For each test, the first three laps were used as input, and laps from the second to the fourth were used as respective output. The dataset was split into a training set (70%) and a test set (30%).

## III. Results and discussion

Personal information, driving features, eye tracking data, voice emotion, physiological features, and cortisol level from a lap were used to predict the safety level of the following lap. The safety of a driving lap was classified into 3 categories: safe driving labeled as 0, moderately unsafe labeled as 1, and unsafe driving labeled as 2. The achieved results are shown in Table 1 in terms of Precision, Recall, and F1-score, and the relative confusion matrix is shown in Table 2.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\text{–}score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (4)$$

Where TP, FP, and FN are respectively True Positives, False Positives, and False Negatives. Precision tells how often the model correctly predicts a positive value, while recall reflects how many of the total positives were identified. F1score indicates the balance between precision and recall. Given that the dataset was imbalanced, as the unsafe behaviors were less frequent than safe ones, these metrics provided us a clear understanding of the performance of class recognition, rather than relying solely on accuracy. The model showed an excellent performance on the classification of safe driving (level 0) and severe unsafe driving (level 2), while distinguishing level 1 from level 0 was challenging.

TABLE I
XGBOOST RESULTS

| Unsafe Levels | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| 0 | 0.98 | 0.96 | 0.97 | 149 |
| 1 | 0.81 | 0.89 | 0.85 | 28 |
| 2 | 0.96 | 0.96 | 0.96 | 28 |

TABLE II
CONFUSION MATRIX

| True / Predicted | Predicted 0 | Predicted 1 | Predicted 2 |
|---|---|---|---|
| **True 0** | 143 | 5 | 1 |
| **True 1** | 3 | 25 | 0 |
| **True 2** | 0 | 1 | 27 |

Something else that was evaluated was how our results matched the relationship between risky behaviors and cortisol level. At the lower end of the cortisol spectrum, drivers tend to exhibit signs consistent with fatigue-induced impairment. Low cortisol levels often reflect insufficient arousal or alertness, potentially linked to poor sleep quality or circadian misalignment. In this state, drivers may suffer from delayed reaction times, reduced vigilance, and diminished situational awareness, all of which elevate the risk of unsafe actions behind the wheel [18]. Conversely, high cortisol levels are typically indicative of increased physiological stress. While moderate stress can sometimes sharpen focus temporarily, excessive cortisol can lead to cognitive overload, emotional reactivity, and impaired judgment. Drivers with elevated cortisol may display riskier behaviors such as aggressive driving, poor attention allocation, or hurried decision-making in complex traffic scenarios. Cortisol can be identified in multiple biological samples, such as serum, sweat, hair, urine, and saliva [19]. For the sake of this work, is necessary a system to detect cortisol level while driving, and this is possible thanks to the Biocompatible core shell microneedle sensor discussed by Zhou et al. [20].

In Fig. 3 is shown the frequency distribution of all the cortisol level measured in the analysis. The test dataset was divided based on the achieved unsafe driving behavior, and the frequency distribution of cortisol level was plotted for each class. Results are shown in Fig. 4.

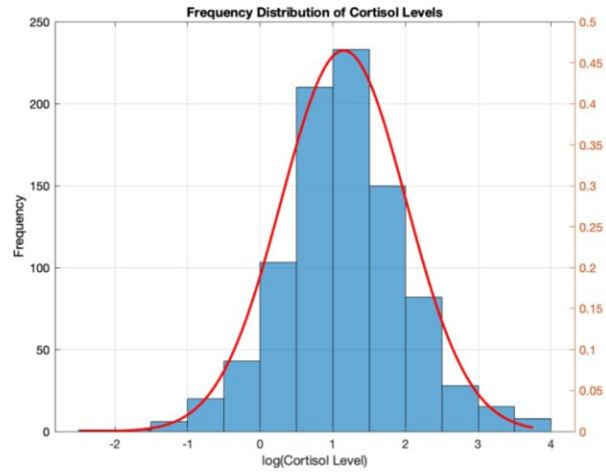


Fig. 3. Frequency distribution of cortisol levels

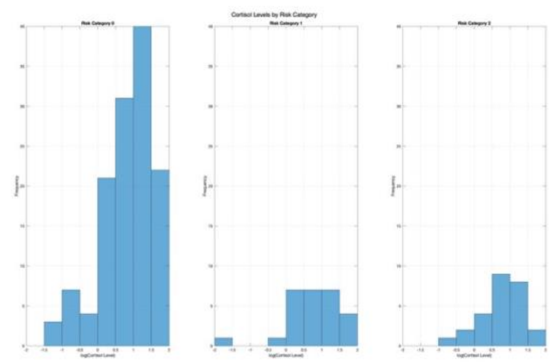

Fig. 4. Frequency distribution of cortisol levels for each unsafe driving class

In risky driving (level 1 and level 2), the distribution seems to be more concentrated in area of relatively high cortisol levels. For level 1, a single lap exists at the extreme left tail of the cortisol distribution, where a poor vigilance level is expected. In the future, the relationship between cortisol level and driving risk requires further study due to the selected limited data points of 205.

XGBoost was also compared with other commonly used machine learning techniques. Table 3 shows the average precision, recall, and F1-score obtained from Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbors (kNN), and Naïve Bayes (NB) on the considered data. All the models tested in this study demonstrated strong performance in identifying safe driving (level 0), which consitutes the majority of the dataset. However, the effectiveness of the models significantly declined when detecting level 1 and level 2, the underrepresented classes, highlighting the challenge of class imbalance. Notably, among the evaluated models, XGBoost stood out for its ability to better capture these minority classes.

TABLE III
MODELS COMPARISON

| Models | Precision | Recall | F1-score |
|---|---|---|---|
| DT | 78% | 80% | 79% |
| RF | 88% | 86% | 86% |
| LR | 77% | 76% | 76% |
| SVM | 82% | 74% | 78% |
| kNN | 78% | 73% | 74% |
| NB | 67% | 73% | 70% |
| XGBoost | 92% | 94% | 93% |

## IV. CONCLUSION

In this study, the combination of driving, physiological and personal features was tested to predict unsafe driving behaviors in a driving simulation scenario. XGBoost was used to train a predictive model able to perform this task. This model performed particularly well in predicting safe and highly unsafe driving behaviors, which were crucial for early intervention. However, a slight reduction in performance was observed in predicting moderately unsafe behavior (level 1). This misclassification, where level 1 was occasionally predicted as level 0, highlighted the challenge of distinguishing between moderate unsafe behavior and safe driving based on the selected features. Cortisol levels emerged as a valuable physiological indicator in this analysis, as its variation was consistently associated with risky driving behavior. Both high and low cortisol levels, reflecting stress and fatigue states, correlate with increased unsafe driving patterns. This highlights cortisol as a meaningful biomarker for monitoring driver state. Importantly, thanks to wearable biosensors able to detect it, it is now feasible to monitor cortisol during driving. This enables early detection of potentially risky physiological conditions, paving the way for proactive interventions to enhance road safety. Future work will be operated to improve level 1 identification performance and to study unsafe driving in real driving scenarios.

## REFERENCES


[1] M. Grimm and C. Treibich, “Socio-economic determinants of road traffic accident fatalities in low and middle income countries,” 2010.

[2] World Health Organization, “Global status report on road safety 2023.”

[3] S. Jafarpour and V. Rahimi-Movaghar, “Determinants of risky driving behavior: a narrative review,” *Medical Journal of the Islamic Republic of Iran*, vol. 28, p. 142, 12 2014.

[4] K. Damsara and A. de Barros, “A Systematic Review on User Acceptance of Advanced Driver Assistance Systems (ADAS),” *Transportation Research Procedia*, vol. 82, pp. 3472–3482, 2025.

[5] R. E. Barka and I. Politis, “Driving into the future: A scoping review of smartwatch use for real-time driver monitoring,” *Transportation Research Interdisciplinary Perspectives*, vol. 25, p. 101098, 5 2024.

[6] D. Filipa Ferreira, S. Ferreira, C. Mateus, N. Barbosa-Rocha, L. Coelho, and M. A. Rodrigues, “Advancing the understanding of pupil size variation in occupational safety and health: A systematic review and evaluation of open-source methodologies,” *Safety Science*, vol. 175, p. 106490, 7 2024.

[7] K. T. Nguyen, W. K. Liang, C. H. Juan, and C. A. Wang, “Timefrequency analysis of pupil size modulated by global luminance, arousal, and saccade preparation signals using Hilbert-Huang transform,” *International Journal of Psychophysiology*, vol. 176, pp. 89–99, 6 2022.

[8] N. Gullett, Z. Zajkowska, A. Walsh, R. Harper, and V. Mondelli, “Heart rate variability (HRV) as a way to understand associations between the autonomic nervous system (ANS) and affective states: A critical review of the literature,” *International Journal of Psychophysiology*, vol. 192, pp. 35–42, 10 2023.

[9] A. R. Banu and V. Nagaveni, “Assessment of Sympathetic and Parasympathetic Activities of Nervous System from Heart Rate Variability Using Machine Learning Techniques,” *SN Computer Science*, vol. 4, pp. 1–10, 9 2023.

[10] H. Luo, J. Wei, Y. Yasin, S. J. Wu, A. Barszczyk, Z. P. Feng, and K. Lee, “Stress Determined through Heart Rate Variability Predicts Immune Function,” *Neuroimmunomodulation*, vol. 26, pp. 167–173, 10 2019.

[11] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, and L. Pecchia, “Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis,” *Biomedical Signal Processing and Control*, vol. 18, pp. 370–377, 4 2015.

[12] L. Langle and R. Dantu, “Are you a safe driver,” *Proceedings 12th IEEE International Conference on Computational Science and Engineering, CSE 2009*, vol. 2, pp. 502–507, 2009. [13] T. Gordon, M. A. Barnes, D. Blower, L. P. Kostyniuk, T. Gordon, A. Blankespoor, M. Barnes, D. Blower, P. Green, and L. Kostyniuk,
“YAW RATE ERROR-A DYNAMIC MEASURE OF LANE KEEPING CONTROL PERFORMANCE FOR THE RETROSPECTIVE ANALYSIS OF NATURALISTIC DRIVING DATA,”

[14] Y. Lim, K.-W. Ng, P. Naveen, and S.-C. Haw, “Emotion Recognition by Facial Expression and Voice: Review and Analysis,” *Journal of Informatics and Web Engineering*, vol. 1, pp. 45–54, 9 2022.

[15] L. Van Der Maaten and G. Hinton, “Visualizing Data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[16] L. Eboli, G. Mazzulla, and G. Pungillo, “Combining speed and acceleration to define car users’ safe or unsafe driving behaviour,” *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 113–125, 7 2016.

[17] L. Eboli, G. Mazzulla, and G. Pungillo, “How to define the accident risk level of car drivers by combining objective and subjective measures of driving style,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 49, pp. 29–38, 8 2017.

[18] M. Kumari, E. Badrick, T. Chandola, E. K. Adam, M. Stafford, M. G. Marmot, C. Kirschbaum, and M. Kivimaki, “Cortisol secretion and fatigue: Associations in a community based cohort,” *Psychoneuroendocrinology*, vol. 34, pp. 1476–1485, 11 2009.

[19] Y. Wang, H. Murakami, T. Kasama, S. Mitsuzawa, S. Shinkawa, R. Miyake, and M. Takai, “An automatic immuno-microfluidic system integrating electrospun polystyrene microfibrous reactors for rapid detection of salivary cortisol,” *iScience*, vol. 26, p. 107820, 10 2023.

[20] S. Zhou, Y. Chino, T. Kasama, R. Miyake, S. Mitsuzawa, Y. Luan, N. B. Ahmad, H. Hibino, and M. Takai, “Biocompatible Core-Shell Microneedle Sensor Filled with Zwitterionic Polymer Hydrogel for Rapid Continuous Transdermal Monitoring,” *ACS nano*, vol. 18, pp. 26541–26559, 10 2024.