Impaired-driving Detection using multi-sensor Data

MIS 6940: Project Seminar

Submitted to:
Prof. Venugopal Balijepally

Submitted By:
Vishaka Sharma

Table of Contents

Abstract.......... 2
Organization/Industry Description 3
Research Questions 3
Sponsors.......... 3
Method.......... 4
Project Deliverables/Findings:.......... 5
Research Evaluation.......... 6
Lessons Learned 6
Limitations & Future Work.......... 7
References 7
Appendices 8

Abstract

Driving under the influence has remained a serious societal issue, and alcohol and fatigue have a significant impact on driver's performance. This project examines the behavior of impaired driving using multimodal data recorded in a controlled driving simulator experiment conducted at Oakland University. Phase 1 evaluated the eligibility screening response rates to determine the eligible respondents and general screening exclusion criteria, like motion sickness and health concerns. Phase 2 examined simulator performance, BAC levels, physiological indicators (ECG), sleepiness (KSS), and DMS measures across four impairment stages. The study intends to find quantifiable signs of impairment and determine whether or not machine learning models can distinguish high-risk states of driving. The statistical analysis and Pearson correlations of the study were performed using Python, regression, and ML algorithms (Logistic Regression, Random Forest, Decision Tree), and a Power BI dashboard, revealing that BAC is a strong predictor of impairment classification, and physiological responses can further contribute to the predictive ability.

Organization/Industry Description

The study falls into the framework of the automotive safety and transportation technology sector, which is becoming increasingly impacted by the development of advanced sensing, driver-monitoring, and predictive analytics. In developing preventive safety systems, automotive companies, transportation regulators, and ADAS developers are highly dependent on the insights provided by data to comprehend human impairment.

This research study used a driving simulator donated by Mitsubishi Electric Automotive America (MEAA) and performed at the Oakland University research center under the Industrial and Systems Engineering Department. Impairment studies with simulation are common among OEMs and Tier-1 suppliers to come up with drowsiness detection, alcohol impairment monitoring, and driver assistance systems to enhance road safety.

Research Questions

The project will answer critical research questions in the industry:

1. What is the effect of alcohol consumption (BAC) on physiological responses and driving behavior in a simulator environment?

2. What signals, ECG, steering variability, speed, KSS sleepiness exhibit any statistically meaningful relationship with impairment?

3. Are multi-modal data reliable in categorizing the state of High Impairment vs Low Impairment with machine learning models?

4. What could the results do to assist the real-life ADAS, driver monitoring systems, or industry compliance?

Sponsors

Industry Sponsor: Mitsubishi Electric Automotive America (MEAA)
Provided the driving simulator, cameras, and biosensors.

Academic Advisor: Prof. Vijitashwa Pandey & Prof. Venugopal Balijepally
Research Team: Graduate assistants involved in participant coordination, data collection, and monitoring.

Method

Data Collection & Study Design.
The research followed a two-phase structure combining eligibility screening and simulator-based impaired driving analysis.

Phase 1- Screening & Eligibility Analysis: Participants completed a detailed eligibility form capturing demographics, drinking habits, health conditions, motion-sickness sensitivity, and other risk factors.
Phase 2 - Impaired-Driving
The participants drove in the following two conditions:
1. Baseline (Level 1) - No alcohol
2. Rising alcohol levels (Levels 2-4) - administered to reach approx 0.079% BAC

Each session included:
1. Simulator logs: Steering angle, Speed, Throttle, brake, Autopilot
2. BAC & Sleepiness: BAC readings (breathalyzer) + detox measurements & KSS sleepiness scale (1-9)
3. Physiological (BITalino): ECG
4. Driver Monitoring System (DMS): Face mask signal, Glasses detection, Head tracking state
5. Demographics: Age, gender, Height, Weight

All variables and sensor definitions are documented in **Appendix A.**

Data Preparation
These steps were implemented using the following steps in Python with the use of pandas, NumPy, SciPy, and scikit-learn:

Consolidated 106 participants' folders with 4 levels each
Standardized and cleaned variable names
Extracted Mean/standard deviation/Min/Max for simulator, DMS, and ECG data
Removed invalid or incomplete participants
Built a Master dataset (Final_Master_Dataset)
Created statistical and correlation tables (Pearson r, p-values, N)
Built Machine learning dataset with binary target:

HighImpair = 1 (Levels 3-4)
LowImpair = 0 (Levels 1-2)

Analytics & Modeling Techniques
To respond to research questions, the following analyses were conducted:

- Descriptive Statistical Means, standard deviations, and ranges.
- Pearson Correlation Analysis: R-values, Np-values, N.
- Regression Models: Three OLS models investigating the prediction of steering variability by using BAC, KSS, speed, and ECG.
- Machine Learning Models: Logistic Regression, Random Forest, and Decision Tree to classify impairment.
- ROC/AUC Assessment: Model performance comparison.

- Tools used: Python, Power BI, OpenPyXL, Jupyter Notebook

Project Deliverables/Findings:

1. Correlations and Descriptive Statistics

The results of the descriptive statistics and the correlation analysis combined reveal evidence of impairment-related changes across behavioral and physiological indicators. BAC value and KSS score increase consistently from Level 1 to Level 4, which validates the effects of alcohol and alertness. The control variability and speed fluctuations are more dispersed at moderate and high impairment levels, indicating fewer changes in driving stability. The correlations also indicate that there are strong correlations between BAC value and KSS score and steering variability, whereas there are moderate correlations between speed and ECG measures. Overall, these findings confirm that a few critical variables are BAC, sleepiness, steering variability, speed variability, and ECG fluctuations, as they can reflect most impairment-related changes in the data. Full descriptive statistics and the correlation matrix are provided in **Appendix B**

2. Regression Models

In all regression models, BAC was the only predictor of steering variability that was consistent and statistically significant ($p < 0.05$), indicating that the alcohol level is the best linear predictor of impairment in this dataset. The addition of KSS (sleepiness) in Model 2 did not significantly increase the prediction (R2 only changed between 0.015 and 0.016), and KSS was not statistically significant. The complete multimodal Model 3 (BAC + KSS + Speed + ECG + Throttle + Brake) also did not raise the R2 by more than 0.026, suggesting that further physiological and driving signals do not exhibit linear explanatory values of steering impairment. Such low R^2 values are anticipated in real-world human factors data, where impairment effects are non-linear and differ among individuals. Comprehensively, regression findings support the hypothesis that BAC is the main linear predictor, whereas the non-linear ML models should be used to describe the multimodal patterns of impairments. Complete regression tables, including coefficients and standard errors, are included in **Appendix C.**

3. ML/DL Prediction Models

The performance of the two Logistic Regression models and the Random Forest is high in all three ML models, but the Random Forest has the highest results (Accuracy = 0.870, AUC = 0.912). Decision Tree depicts less accuracy (0.826) and AUC (0.819), which is congruent with its propensity to overfit. The importance of features in the Random Forest has validated that BAC_value is the most important predictor of high impairment, then throttle behavior, speed signals, and KSS. A moderately significant role belongs to physiological signals (ECG variability). In general, impairment is reliably classified using the ML methods, and the Random Forest with the best balance between predictive power and comprehensibility. Confusion matrices, classification reports, and feature importances are provided in **Appendix D**

Research Evaluation

Impaired-driving detection is a key industrial problem for automotive OEMs, insurance companies, transportation safety departments (NHTSA), and emerging ADAS/AV. The fact is that the conducted analysis shows that the measures of behavioral driving and physiological indicators can be used reliably to indicate alcohol-impairment. The close relation among BAC, KSS, steering variability, throttle consistency, speed variability, and ECG noise confirms the trend of the industry toward the multi-sensor driver monitoring system (DMS/DMS+).

Regression outcomes indicate that BAC_value is a statistically significant predictor of poor steering stability, and the machine-learning models (Random Forest with the accuracy of 0.87 and AUC of 0.91) indicate that impairment can be predicted with high reliability based on a few measurable signals. These results facilitate the creation of in-vehicle impairment detection in real-time, driver safety interventions, as well as insurer risk-based pricing. They further demonstrate that physiological indicators have a slight positive impact on predictive performance, which aligns with the existing trends in camera-based and wearable-based safety analytics. Overall, the deliverables strengthen the industry's understanding of how cognitive, behavioral, and physiological indicators interact under alcohol influence and have an empirical foundation for the next-generation driver-safety systems.

Lessons Learned

Throughout this project, several significant research and analytical lessons emerged:

- Cleaning of data has a strong influence on the accuracy of final models. Bad cases of missing values, duplicated values of participants, and inconsistency of demographic fields had a strong effect on the correlations and regressions until they were cleaned effectively.
- Physiological information is very sensitive. ECG characteristics needed close interpretation as a result of variability being inflated by outliers and sensor noise.
- Simple models can outperform complex ones when the variables are strong. Logistic regression performed near to that of the random forest, which indicates that even in situations where there is a high correlation between the features and the impairment, a linear model is still able to perform well.
- Human impairment is multi-dimensional. KSS (subjective sleepiness) and BAC (objective alcohol concentration) as a combination enhanced the knowledge, whereas neither of them could explain all the behavioral change, which highlights the importance of multi-sensor fusion.
- Visualization is essential. The use of trends to plot the steering, throttle, and speed trends helped interpret patterns more clearly than numeric tables alone.

Through these lessons, the rigor of analysis has been enhanced, as well as the knowledge of the researchers on the real-world data challenges in human-factor safety analytics.

Limitations & Future Work

Limitations of the study:

- Single Experimental environment. The data was collected in a simulator, which restricted generalizability to the real world.
- Sample size and the diversity of the participants. Only a small number of participants contributed to the study, which limits the data; results might vary according to age group, gender, and level of driving experience.
- Potential multicollinearity. The driving measures, including speed, throttle, and brake variability, have a natural correlation, which decreases the interpretability of the regression model.
- Subjective bias in KSS scores. The self-reported alertness is not very consistent and might not accurately represent cognitive impairment.
- Sensor variability. ECG readings showed large standard deviations due to sensor placement, motion artifacts, and individual physiological differences.

Future Work should explore:

- On-road testing of the models based on real-world conditions.
- More advanced ML models, e.g., Gradient Boosting, XGBoost, or time-series deep-learning models.
- Cross-participant generalization, testing whether a model trained on one driver can predict impairment for another.
- Feature engineering of temporal patterns - such as second-by-second variability instead of summary statistics.

This would make the models more robust and extend to OEMs, fleet management companies, and insurance technology providers.

References

Driver Alcohol Detection System for Safety (DADSS). (2020). *National Highway Traffic Safety Administration*. https://www.nhtsa.gov

Hartman, R., Huestis, M. A., Brown, T. L., & Stout, R. (2019). Alcohol effects on driving performance. *Journal of Safety Research, 69*, 157–165.

International Organization for Standardization. (2019). *ISO 20077-1: Road vehicles — Extended vehicle (ExVe) methodology — Part 1: General information*.

Bitbrain. (2021). *Wearable physiological sensor specifications*. https://www.bitbrain.com

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Appendices

Appendix A - Dataset Description & Data Dictionary

This appendix provides a detailed description of the full dataset used in the impaired-driving experiment, including participant demographics, sensor modalities, and variable definitions.

A1. Participant Information

Variable & its description

Participant_ID: Unique participant number (1–100+). Each participant completed 3–4 driving trials across increasing impairment levels.

Levels & Experimental alcohol impairment condition:

- 1 = Baseline (0 BAC)
- 2 = Low impairment
- 3 = Moderate impairment
- 4 = High impairment

Date: Timestamp when each driving trial was recorded.

Age, Gender, Weight, Height: Demographic data collected during onboarding.

A2. Alcohol Measurement Variables

BAC_value: Blood Alcohol Concentration measured before each driving trial (0.000–0.160).

A3. Karolinska Sleepiness Scale (KSS)

KSS_score Numeric sleepiness score (1–9). Higher = more sleepy.

A4. Driving Simulator (SIM) Telemetry Variables

Steering Behavior:

SIM_Steering_mean: Average steering wheel angle. Negative = left; positive = right.

SIM_Steering_std: Steering variability (higher = more unstable/wobbling).

SIM_Steering_min / max: Extreme left/right steering positions during the trial.

Throttle Input

SIM_Throttle_mean / std/max: Accelerator pedal input (0–1).

- 0 = no throttle
- 1 = full throttle

Brake Input

SIM_Brake_mean / std/min/max: Brake pedal pressure (0–1).
• 0 = no brake
• 1 = full brake

Vehicle Speed
SIM_Speed_mean / std/min/max: Vehicle speed measurements (km/h). Mean ≈ 80 km/h expected.

A5. BITalino Physiological Sensor (ECG)

BIT_ECG_mean Average electrical heart-signal level (0–1023).
BIT_ECG_std Heart-signal variability indicating physiological stress or arousal.
BIT_ECG_min / max Minimum and maximum ECG readings during the trial.

Interpretation:

Higher std = more variability = sympathetic activation (stress, cognitive load).

1023 = upper sensor limit.

Appendix B - Correlation & Descriptive Statistics

No.	Variable	Mean	Std. Dev.	1	2	3	4	5	6	7
1	BAC_value	0.029	0.025	1.000						
2	KSS_score	3.304	1.517	0.443***	1.000					
3	Steering_Std	0.018	0.013	-0.141**	-0.037	1.000				
4	Speed_mean	80.633	10.818	0.347***	0.220***	-0.253***	1.000			
5	ECG_Std	62.315	42.343	-0.084	-0.047	-0.052	0.089	1.000		
6	Throttle_mean	0.481	0.051	0.387***	0.252***	-0.286***	0.968***	-0.118	1.000	
7	Brake_Std	0.052	0.023	0.105	0.139***	0.089	-0.014	0.004	0.049	1.000

N = 411; *p < 0.10, **p<0.05, ***p<0.01

	Participant_	Level	BAC_value	KSS_score	_Steering_	Throttle_r	_Brake_m	M_Brake_s	_Speed_m	M_Speed_	_Speed_r	T_ECG_me	IT_ECG_st	IT_ECG_mi	IT_ECG_ma	Age_x	Gender_x	Weight_x	Height_x	Age_y	Weight_y	Height_y
Pearson r																						
Participant	1	0	-0.023	0.006	0.079	-0.025	0.341	0.047	-0.068	-0.319	-0.172	0.184	-0.161	0.126	-0.088	-0.194	0.132	-0.022	-0.161	-0.194	-0.022	-0.161
Level	0	1	0.905	0.45	-0.155	0.429	-0.027	0.079	0.379	-0.536	0.008	-0.044	-0.071	0.18	-0.207	0	0	0	0	0	0	0
BAC_value	-0.023	0.905	1	0.442	-0.141	0.387	-0.013	0.105	0.347	-0.463	0.011	-0.035	-0.071	0.167	-0.176	0.071	0.017	0.075	0.04	0.071	0.075	0.04
KSS_score	0.006	0.45	0.442	1	-0.037	0.252	0.006	0.139	0.22	-0.173	0.14	-0.004	-0.038	0.02	-0.048	-0.06	0.059	-0.023	0.083	-0.06	-0.023	0.083
SIM_Steer	0.079	-0.155	-0.141	-0.037	1	-0.068	0.114	0.076	-0.06	0.232	0.084	0.032	-0.027	-0.005	0.035	0.052	-0.011	0.038	0.01	0.052	0.038	0.01
SIM_Throt	-0.025	0.429	0.387	0.252	-0.068	1	-0.163	0.05	0.968	-0.286	0.546	0.08	-0.045	0.082	-0.118	-0.064	0.027	-0.067	0.027	-0.064	-0.067	0.027
SIM_Brake	0.341	-0.027	-0.013	0.006	0.114	-0.163	1	0.349	-0.218	-0.001	-0.13	0.063	-0.106	0.055	-0.018	-0.069	-0.178	0.084	0.085	-0.069	0.084	0.085
SIM_Brake	0.047	0.079	0.105	0.139	0.076	0.05	0.349	1	-0.014	0.032	0.094	0.004	-0.008	0.025	-0.068	0.119	0.002	0.022	0.037	0.119	0.022	0.037
SIM_Speec	-0.068	0.379	0.347	0.22	-0.06	0.968	-0.218	-0.014	1	-0.253	0.48	0.076	-0.052	0.099	-0.127	-0.054	0.012	-0.073	0.054	-0.054	-0.073	0.054
SIM_Speec	-0.319	-0.536	-0.463	-0.173	0.232	-0.286	-0.001	0.032	-0.253	1	0.407	-0.071	0.089	-0.177	0.224	0.048	0.009	0.054	0.064	0.048	0.054	0.064
SIM_Speec	-0.172	0.008	0.011	0.14	0.084	0.546	-0.13	0.094	0.48	0.407	1	-0.039	0.107	-0.135	0.168	0.007	0.092	-0.012	0.001	0.007	-0.012	0.001
BIT_ECG_r	0.184	-0.044	-0.035	-0.004	0.032	0.08	0.063	0.004	0.076	-0.071	-0.039	1	-0.716	0.221	-0.104	0.028	-0.047	0.067	0.049	0.028	0.067	0.049
BIT_ECG_s	-0.161	-0.071	-0.071	-0.038	-0.027	-0.045	-0.106	-0.008	-0.052	0.089	0.107	-0.716	1	-0.468	0.345	0.001	-0.077	-0.068	0.046	0.001	-0.068	0.046
BIT_ECG_r	0.126	0.18	0.167	0.02	-0.005	0.082	0.055	0.025	0.099	-0.177	-0.135	0.221	-0.468	1	-0.558	-0.081	0.054	-0.058	-0.049	-0.081	-0.058	-0.049
BIT_ECG_r	-0.088	-0.207	-0.176	-0.048	0.035	-0.118	-0.018	-0.068	-0.127	0.224	0.168	-0.104	0.345	-0.558	1	-0.043	-0.109	0.04	0.062	-0.043	0.04	0.062
Age_x	-0.194	0	0.071	-0.06	0.052	-0.064	-0.069	0.119	-0.054	0.048	0.007	0.028	0.001	-0.081	-0.043	1	-0.017	0.214	0.069	1	0.214	0.069
Gender_x	0.132	0	0.017	0.059	-0.011	0.027	-0.178	0.002	0.012	0.009	0.092	-0.047	-0.077	0.054	-0.109	-0.017	1	-0.107	-0.556	-0.017	-0.107	-0.556
Weight_x	-0.022	0	0.075	-0.023	0.038	-0.067	0.084	0.022	-0.073	0.054	-0.012	0.067	-0.068	-0.058	0.04	0.214	-0.107	1	0.315	0.214	1	0.315
Height_x	-0.161	0	0.04	0.083	0.01	0.027	0.085	0.037	0.054	0.064	0.001	0.049	0.046	-0.049	0.062	0.069	-0.556	0.315	1	0.069	0.315	1
Age_y	-0.194	0	0.071	-0.06	0.052	-0.064	-0.069	0.119	-0.054	0.048	0.007	0.028	0.001	-0.081	-0.043	1	-0.017	0.214	0.069	1	0.214	0.069
Weight_y	-0.022	0	0.075	-0.023	0.038	-0.067	0.084	0.022	-0.073	0.054	-0.012	0.067	-0.068	-0.058	0.04	0.214	-0.107	1	0.315	0.214	1	0.315
Height_y	-0.161	0	0.04	0.083	0.01	0.027	0.085	0.037	0.054	0.064	0.001	0.049	0.046	-0.049	0.062	0.069	-0.556	0.315	1	0.069	0.315	1

	Participant_	Level	BAC_value	KSS_score	_Steering_	Throttle_r	_Brake_m	M_Brake_s	_Speed_m	M_Speed_	_Speed_r	T_ECG_me	IT_ECG_st	IT_ECG_mi	IT_ECG_ma	Age_x	Gender_x	Weight_x	Height_x	Age_y	Weight_y	Height_y
p-value																						
Participant	0	1	0.6543	0.9014	0.1198	0.6276	0	0.3509	0.1769	0	0.0006	0.0003	0.0016	0.0134	0.0837	0.0001	0.009	0.664	0.0014	0.0001	0.664	0.0014
Level	1	0	0	0	0.0021	0	0.5944	0.1178	0	0	0.872	0.3895	0.1662	0.0004	0	1	1	1	1	1	1	1
BAC_value	0.6543	0	0	0	0.0053	0	0.7948	0.0387	0	0	0.8328	0.4997	0.1636	0.001	0.0005	0.1579	0.7442	0.1373	0.4343	0.1579	0.1373	0.4343
KSS_score	0.9014	0	0	0	0.4626	0	0.9095	0.0062	0	0.0006	0.0058	0.9409	0.4612	0.7021	0.3497	0.2351	0.2509	0.6458	0.1051	0.2351	0.6458	0.1051
SIM_Steer	0.1198	0.0021	0.0053	0.4626	0	0.1802	0.0238	0.1323	0.2339	0	0.0988	0.5317	0.602	0.9239	0.4891	0.3084	0.8324	0.4556	0.8507	0.3084	0.4556	0.8507
SIM_Throt	0.6276	0	0	0	0.1802	0	0.0012	0.3237	0	0	0	0.1171	0.3818	0.1085	0.021	0.2094	0.5995	0.1865	0.5886	0.2094	0.1865	0.5886
SIM_Brake	0	0.5944	0.7948	0.9095	0.0238	0.0012	0	0	0	0.983	0.0099	0.2168	0.0389	0.2845	0.724	0.1722	0.0004	0.098	0.0941	0.1722	0.098	0.0941
SIM_Brake	0.3509	0.1178	0.0387	0.0062	0.1323	0.3237	0	0	0.7753	0.5316	0.0642	0.9391	0.877	0.6208	0.1861	0.019	0.9698	0.6648	0.4685	0.019	0.6648	0.4685
SIM_Speec	0.1769	0	0	0	0.2339	0	0	0.7753	0	0	0	0.1359	0.3139	0.0539	0.0129	0.2868	0.8182	0.1489	0.2889	0.2868	0.1489	0.2889
SIM_Speec	0	0	0	0.0006	0	0	0.983	0.5316	0	0	0	0.1627	0.0818	0.0005	0	0.346	0.8528	0.2864	0.2102	0.346	0.2864	0.2102
SIM_Speec	0.0006	0.872	0.8328	0.0058	0.0988	0	0.0099	0.0642	0	0	0	0.4435	0.0361	0.0081	0.0009	0.8894	0.0695	0.8127	0.9767	0.8894	0.8127	0.9767
BIT_ECG_r	0.0003	0.3895	0.4997	0.9409	0.5317	0.1171	0.2168	0.9391	0.1359	0.1627	0.4435	0	0	0	0.0426	0.5831	0.3547	0.1933	0.3361	0.5831	0.1933	0.3361
BIT_ECG_s	0.0016	0.1662	0.1636	0.4612	0.602	0.3818	0.0389	0.877	0.3139	0.0818	0.0361	0	0	0	0	0.978	0.1307	0.1845	0.3706	0.978	0.1845	0.3706
BIT_ECG_r	0.0134	0.0004	0.001	0.7021	0.9239	0.1085	0.2845	0.6208	0.0539	0.0005	0.0081	0	0	0	0	0.1119	0.2876	0.2548	0.3418	0.1119	0.2548	0.3418
BIT_ECG_r	0.0837	0	0.0005	0.3497	0.4891	0.021	0.724	0.1861	0.0129	0	0.0009	0.0426	0	0	0	0.3972	0.0324	0.4312	0.2269	0.3972	0.4312	0.2269
Age_x	0.0001	1	0.1579	0.2351	0.3084	0.2094	0.1722	0.019	0.2868	0.346	0.8894	0.5831	0.978	0.1119	0.3972	0	0.731	0	0.1746	0	0	0.1746
Gender_x	0.009	1	0.7442	0.2509	0.8324	0.5995	0.0004	0.9698	0.8182	0.8528	0.0695	0.3547	0.1307	0.2876	0.0324	0.731	0	0.0343	0	0.731	0.0343	0
Weight_x	0.664	1	0.1373	0.6458	0.4556	0.1865	0.098	0.6648	0.1489	0.2864	0.8127	0.1933	0.1845	0.2548	0.4312	0	0.0343	0	0	0	0	0
Height_x	0.0014	1	0.4343	0.1051	0.8507	0.5886	0.0941	0.4685	0.2889	0.2102	0.9767	0.3361	0.3706	0.3418	0.2269	0.1746	0	0	0	0.1746	0	0
Age_y	0.0001	1	0.1579	0.2351	0.3084	0.2094	0.1722	0.019	0.2868	0.346	0.8894	0.5831	0.978	0.1119	0.3972	0	0.731	0	0.1746	0	0	0.1746
Weight_y	0.664	1	0.1373	0.6458	0.4556	0.1865	0.098	0.6648	0.1489	0.2864	0.8127	0.1933	0.1845	0.2548	0.4312	0	0.0343	0	0	0	0	0
Height_y	0.0014	1	0.4343	0.1051	0.8507	0.5886	0.0941	0.4685	0.2889	0.2102	0.9767	0.3361	0.3706	0.3418	0.2269	0.1746	0	0	0	0.1746	0	0

Appendix C - Regression Models

	Model 1	**Model 2**	**Model 3**
Constant	0.0204*** (0.001)	0.0198*** (0.002)	0.0240*** (0.008)
Independent Variables			
BAC_value	-0.0669** (0.027)	-0.0734** (0.031)	-0.0720** (0.032)
KSS_score	—	0.0002 (0.000)	0.0002 (0.001)
SIM_Speed_mean	—	—	0.0001 (0.000)

BIT_ECG_std	—	—	-0.0000118 (0.000016)
SIM_Throttle_mean	—	—	-0.0369 (0.055)
SIM_Brake_std	—	—	0.0523* (0.030)
R^2	0.015	0.016	0.026
F	5.985**	3.099**	1.659
Adj-R^2	0.013	0.011	0.010

*p < 0.10, ** p < 0.05, ***p < 0.01

Appendix D - Machine Learning Results

Independent Variables
X1 = BAC_value
X2 = KSS_score
X3 = SIM_Steering_std
X4 = SIM_Speed_mean
X5 = BIT_ECG_std
X6 = SIM_Throttle_mean
X7 = SIM_Brake_std
X8 = SIM_Brake_mean
X9 = BIT_ECG_mean
Gender_bin

Model Summary

	Model 1 – Logistic Reg				**Model 2 – Random Forest**			
Class	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
0	0.836	0.879	0.857	58	0.877	0.862	0.870	58
1	0.870	0.825	0.847	57	0.862	0.877	0.870	57
Accuracy	0.852				0.870			

Micro Avg	0.853	0.852	0.852	115	0.870	0.870	0.870	115
Weighted Avg.	0.853	0.852	0.852	115	0.870	0.870	0.870	115
AUC	0.920				0.912			

	Model 3 – Decision Tree			
Class	Precision	Recall	F1-Score	Support
0	0.839	0.810	0.825	58
1	0.814	0.842	0.828	57
Accuracy	0.826			
Micro Avg	0.826	0.826	0.826	115
Weighted Avg.	0.827	0.826	0.826	115
AUC	0.819			

```
==============================
Logistic Regression - Confusion Matrix
==============================
```

Prediction	**0**	**1**
TRUE		
0	51	7
1	10	47

```
==============================
Random Forest - Confusion Matrix
==============================
```

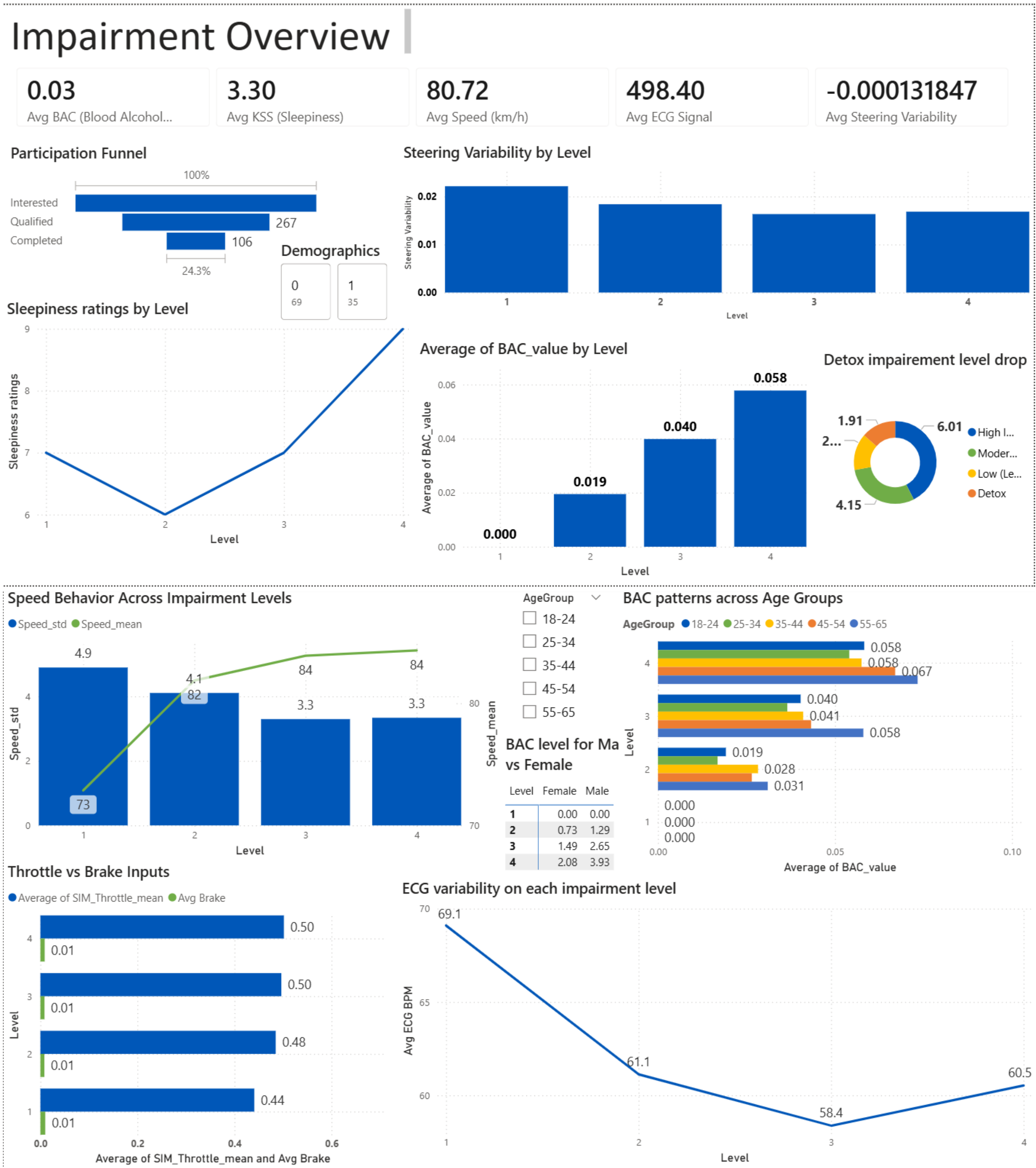
Prediction	**0**	**1**
TRUE		
0	50	8
1	7	50

```
==============================
Decision Tree - Confusion Matrix
==============================
```

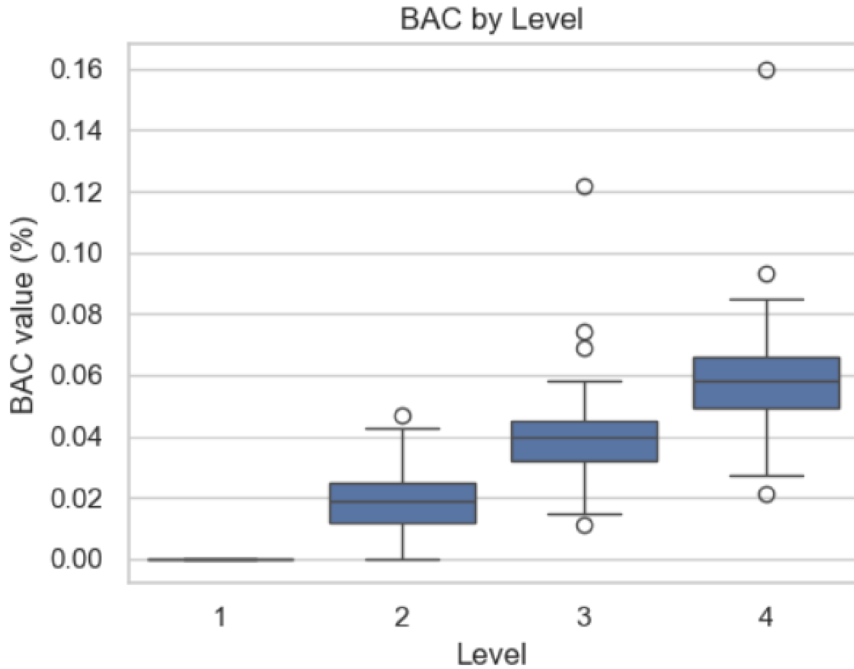
Prediction	**0**	**1**
TRUE		
0	47	11
1	9	48

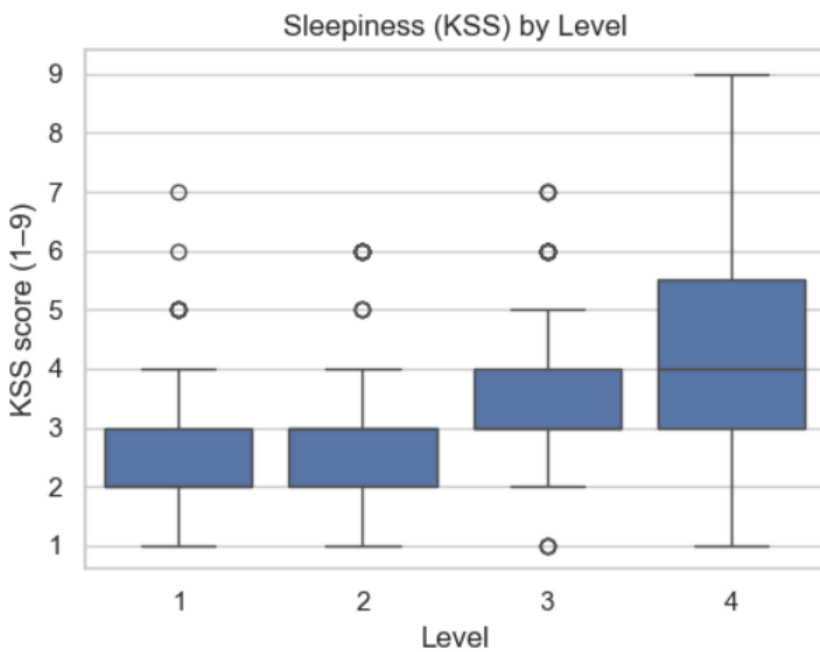
Appendix E - Visualizations

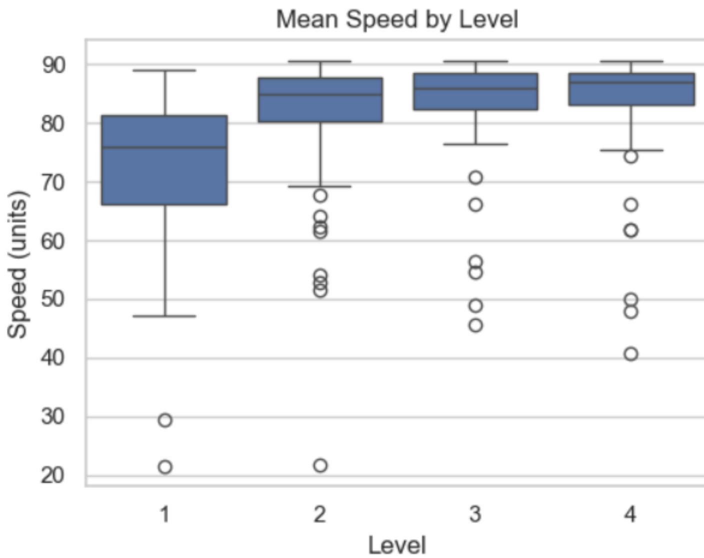
All charts

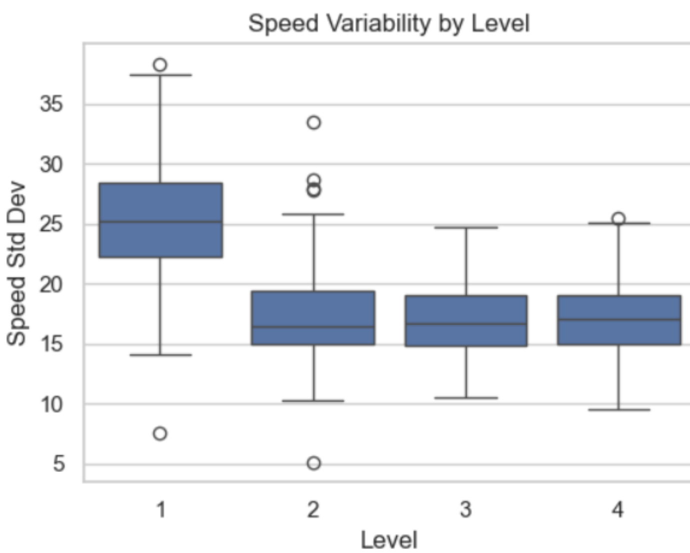


Level	Female	Male
1	0.00	0.00
2	0.73	1.29
3	1.49	2.65
4	2.08	3.93


BAC by Level
BAC value (%)
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1
2
3
4
Level


Sleepiness (KSS) by Level
KSS score (1–9)
9
8
7
6
5
4
3
2
1
1
2
3
4
Level


Mean Speed by Level
Speed (units)
90
80
70
60
50
40
30
20
1
2
3
4
Level


Speed Variability by Level
Speed Std Dev
35
30
25
20
15
10
5
1
2
3
4
Level