Classification of motion sickness conditions in autonomous driving applications

Srikanth NOUDURI[1], Nikolas SCHIERHORST[1], Paul SEIWERT[2], Lotte WAGNER-DOUGLAS[3], Alexander KIRMAS[3], Nikica HENNES[4], Gudrun VOß[3], Stefan LADWIG[3], Kai REWITZ[2], Lutz ECKSTEIN[3], Dirk MÜLLER[2], Alexander MERTENS[1], Verena NITSCH[1]

[1] Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Eilfschornsteinstraße 18, D-52062 Aachen, Germany
[2] E.ON Energy Research Center, RWTH Aachen University, Mathieustraße 10, D-52074 Aachen, Germany
[3] Institute for Automotive Engineering, RWTH Aachen University, Steinbachstraße 7, D-52074 Aachen, Germany
[4] Research and Innovation Center, Ford-Werke GmbH, Henry-Ford-Straße 1, 50735, Köln, Germany

Abstract: With the introduction of novel automated vehicles (AVs), drivers are becoming passengers who could be susceptible to motion sickness symptoms in cars (carsickness). Therefore, it is advisable to adapt driving characteristics like accelerations, braking, lane change behaviors based on the sensitivity of the passengers. Physiological markers play an essential role in empirical evaluations and development of potential countermeasures that can appropriately respond to a passenger’s state with a possibility to detect early signs of carsickness at an individual level. In this work, we examined 48 healthy participants’ data from an experimental study spanned over several laps and developed an algorithm to address reliable, objective detection of carsickness symptoms prior to their subjective assessments. As a result, models with Logistic Regression Classifier recorded highest weighted accuracy of (80 %) and F1 scores of (0.78) over the laps.

Key Words: Carsickness, Physiological Measurements, Misery Scores, Machine Learning, Multi-Class Classification, Algorithm Development.

1. Introduction

Carsickness is a syndrome associated with a wide variety of unpleasant symptoms. This phenomenon might become highly relevant when passengers travelling in autonomous vehicles are engaged in non-driving, visual/visuomotor tasks and while potentially facing rearwards. In recent years, many researchers incorporated biological signals for various mental health recognition tasks such as classification of stress levels in automobile drivers (Lopez-Martinez et al. 2019), detection of simulator sickness (Tauscher et al. 2020) and Virtual Reality (VR) induced motion sickness (Keshavarz et al. 2022). However, only a limited number of studies have specifically addressed carsickness detection by employing multimodal learning approaches in real-life driving scenarios. While Tan et al. (2022) addressed motion sickness detection with traditional machine learning models, Hwang et al. (2022) examined unimodal and multimodal datasets (EEG, ECG, PPG, GSR etc.) with deep learning models, but the real-time applicability of such algorithms was not discussed. As carsickness symptoms might not be indicated by instantaneous changes in biological signals, these algorithms will have progressively accurate predictions over the time, when they are trained with datasets obtained from passengers being continuously monitored. Hence, we propose a real-time applicable algorithm, whereby machine learning models appropriately perceive passengers while examining them over a longer course of duration in the form of consecutive laps. The following algorithm pipeline (Figure 1) was implemented as a part of this work, with an objective to classify participants into different levels of carsickness conditions (null, low, and high) prior to their subjective assessments.

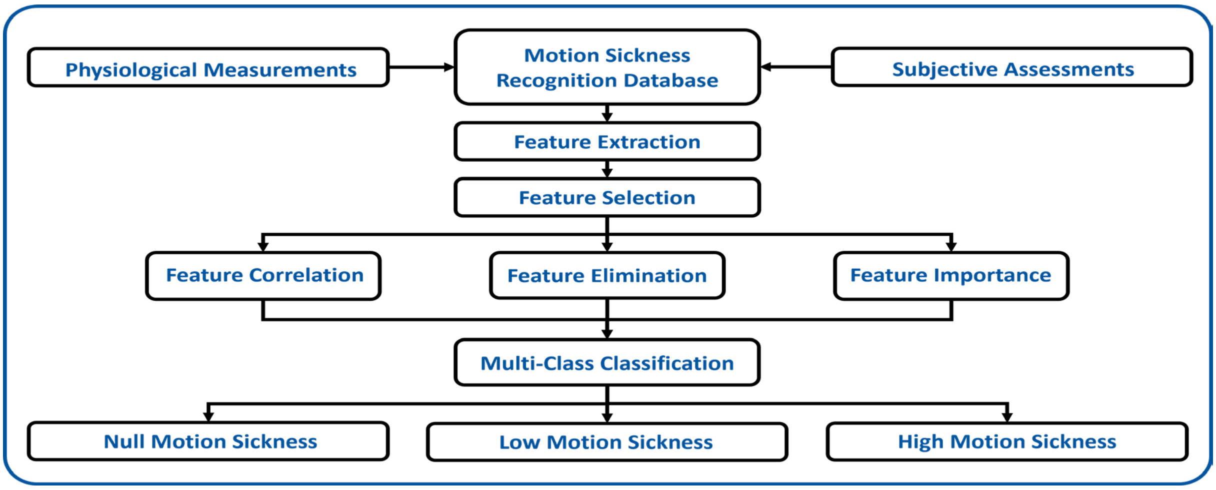

***Figure 1**:* *Algorithm Pipeline*

2. Methodology

This section details the evaluation of objective measures to detect carsickness symptoms, and the proposed methodology comprises the following intermediate steps:

2.1 Motion Sickness Recognition Database:

To formulate an input database as a part of an algorithm development for model training and testing purposes, subjective assessments (MISC Scores: Misery Scale) along with objective measures (physiological measurements) were obtained from an experimental study with a group of healthy participants (*N* = 48 [*n* = 25 male, *n* = 23 female]; *M* = 32.91 years, *SD* = 12.10 years), when they were driven on a prescribed route over several laps under driving conditions similar to that in autonomous driving applications. During these experimental evaluations, physiological measurements (EDA, ECG and EDR) were recorded at a frequency of 512 Hz by using a sensor set from *g.tec medical engineering GmbH.* Similarly, a system of thermography-based infrared (IR) sensor and an optical camera were used for the evaluation of skin temperature measurements. The FLIR Lepton 3.5 IR sensor was positioned at approximately 1m in front of the participants and was continuously calibrated to improve accuracy. A more detailed explanation of the experimental design, objective measures and subjective assessments is described by Wagner-Douglas et al. (2024).

2.2 Feature Extraction:

Raw sensor data prone to artifacts (environmental, body, motion), noise, outliers and missing values were processed with cleaning techniques such as filtering, smoothing, and outlier detection. Pre-processing of cleaned sensor data also involved normalization of the data range, standardization, and dimensionality reduction techniques to reduce data complexities, enhance the quality, integrity, and consistency of the extracted features. Overall, 12 indicators from ECG, EDA, IR measurements were extracted, and the database got accumulated with 13 new features (including MISC) corresponding to every finished lap. Specifically, indicators from ECG measures were obtained using the software tool ‘Kubios-HRV’ by Tarvainen et al. (2014). This incorporated R-R interval detection and beat correction algorithms by Lipponen et al. (2019) for Time- and Frequency domain analysis of the Heart Rate Variability (HRV). From EDA measures, raw signal was decomposed into Phasic, Tonic components by a deconvolution approach proposed by Benedek et al. (2010). Later, indicators such as mean amplitudes from raw signal, area under the curve and skin conductance responses (SCRs) from phasic driver components were derived. While skin temperature was derived from a rectangular window of (7 x 5) pixels on the forehead, breath temperature was analysed from cabin temperature recordings using MATLAB.

These indicators were initially analysed using a Linear-Mixed-Effects Model (LMEM) individually for each sensor set. Here, MISC score was the dependent variable, while the number of driven laps, testing condition (baseline or parcours), the season (summer or winter), and the appointment (first or second) were the fixed effects. With the extracted features as the predictors used in this model, all the participant-IDs were considered as the random effect to respect interindividual influences. As a result, significant effects were shown with the interactions between the lap-count and the testing condition. While indicators HR, EDR from ECG data and mean amplitude of the raw signal from EDA data followed an expected trend, indicators from IR data did not result in a significant effect (Wagner-Douglas et al. 2023). Although, these unimodal analyses confirmed the relevant literature findings, the relationships between individual indictors and MISC scores indicated the relevance of multimodal learning models to detect carsickness symptoms in real-time. As this work aimed at an integrative analysis of multiple biological signals, Machine Learning (ML) models with highest performance ratings in classification tasks were employed on data obtained after every lap.

2.3 Feature Selection:

The proposed framework follows the same design as that of the experimental evaluations, where participants data after every lap was introduced into the input database and the MISC score to be obtained (MISC_n, after n^{th} lap) was predicted. Along with all the primary features, the model also considered secondary features such as differences in measurements from the previous lap and the difference from the acclimatization phase (lap: 0). As features were progressively accumulating after every lap, Recursive Feature Elimination and Cross Validation (RFECV) was implemented as a part of feature selection module. All the selected features were then fed into the model in the form of a feed-forward loop (Figure 2) after every lap and this helped models with consistency and overall improvement in their performances over the laps.

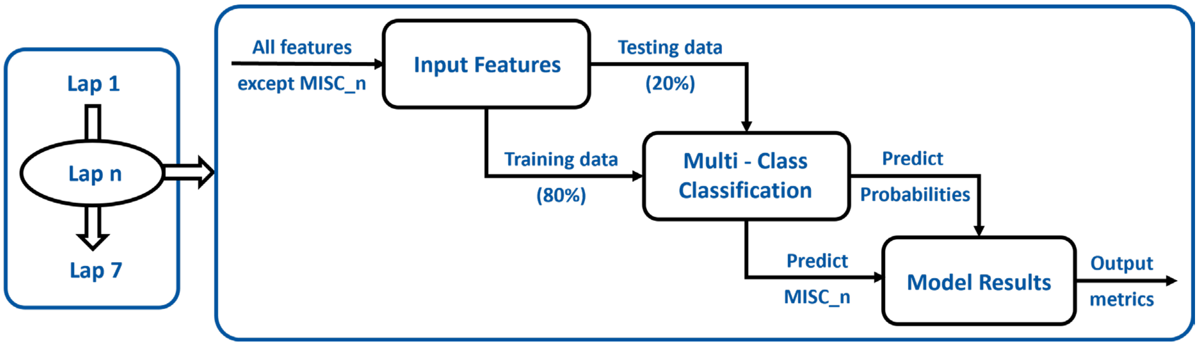

***Figure 2**:* *Model Implementation.*

2.4 Multi-Class Classification:

The selected features (except MISC_n) were randomly split into test (80 %), train (20 %) datasets and the estimator models were trained to predict output variable i.e. MISC score of the finished lap. Estimators such as Logistic Regression Classifier (LR), Random Forests Classifier (RF), Gradient Boost Classifier (GB) and Decision Trees Classifier (DT) were used in this approach. Since this was a multiclass classification problem, the predicted classes were mapped with MISC scores: (class 0: {0}; class 1: {1, 2, 3}; class 2: {4, 5, 6, 7}). Since these models were trained with sparse training datasets that recorded high misery scores (MISC = 5+), they were expected to have a decline in their performance with increasing MISC scores as the laps progressed. To improve this, estimators with RFECV along with hyperparameter tuning techniques such as Grid-Search-CV (GS) were implemented with a balanced weighting method across the predicted classes. Model results in the form of actual and predicted classes are presented with the help of a confusion matrix corresponding to each lap (Figure 3).

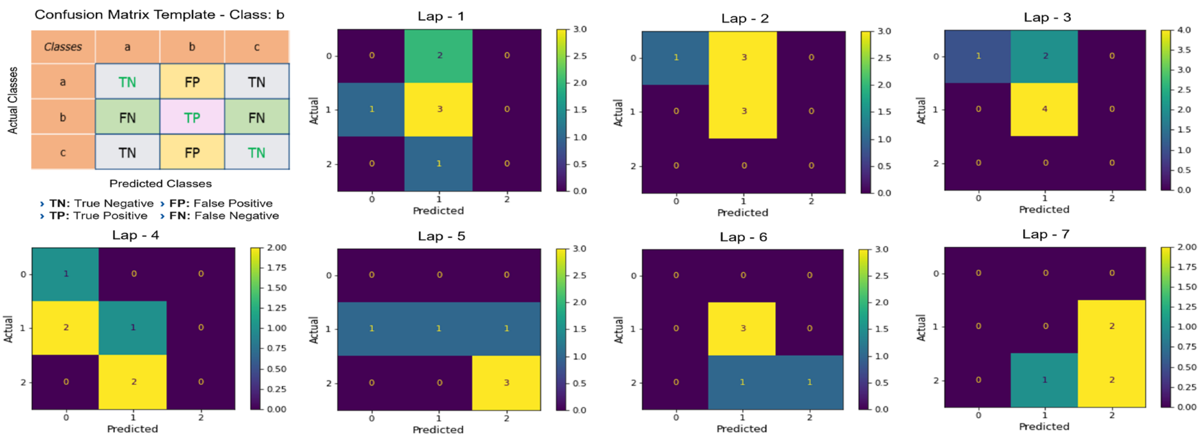

***Figure 3**:* *Confusion Matrices – Logistics Regression (LR)*

3. Results and Discussions

As the laps progressed, the interplay between the attributes (TP, TN, FP, FN) of these confusion matrices, indicate that the predicted classes did match an increasing trend of the actual classes. While accuracy scores were defined as the fraction of correct predictions (TP+TN), the F1 scores were defined as the harmonic mean of precision and recall. Although these ML models showed promising results, sparse training datasets with minimal count of predicted classes caused limitations in the cross-validation steps and affected the quality of model training process. As a result, performance metrics of all these models declined in the middle and in the last laps. However, accuracy, F1 scores were progressive over the laps and all the performance metrics recorded their highest scores in the second last lap (Figure 4).

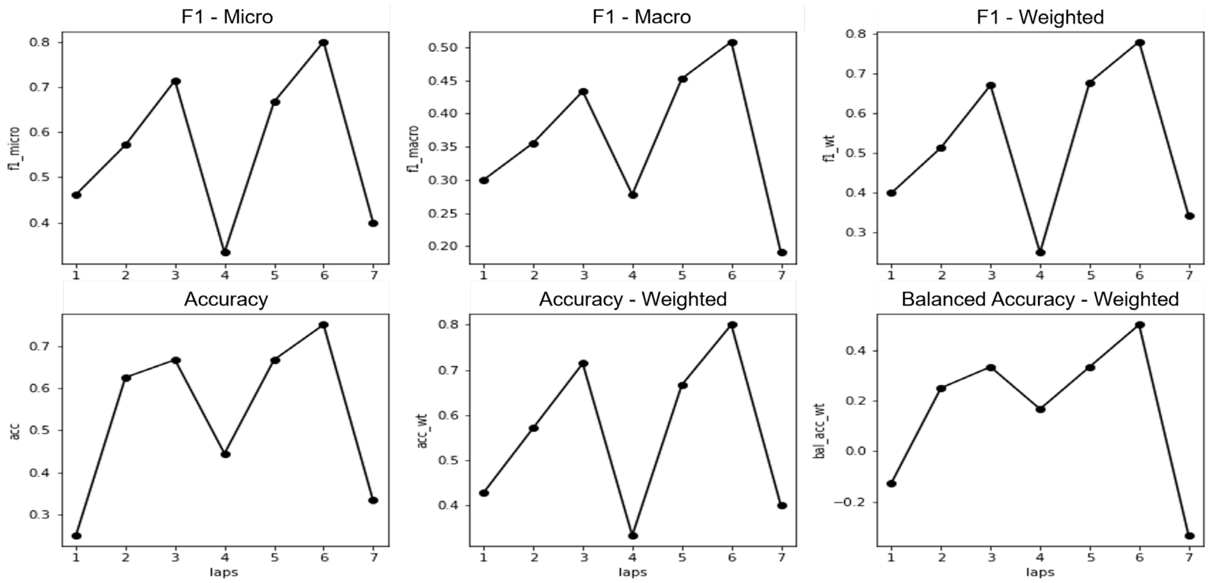

***Figure 4**: Performance Metrics – Logistics Regression (LR)*

Among the evaluated classifiers, Logistic Regression outperformed others with highest weighted accuracy of (80 %) and F1 scores of (0.78) in the second last lap. This model results are presented with the help of test and predicted samples along with their corresponding probability distributions and ROC – AUC curves (Figure 5).

Model Results (LR) – Lap: 6

Test Sample	Predicted Sample
2	2
1	1
1	1
1	1
2	1

Classes	Precision	Recall	F1-Score
0	0.00	0.00	0.00
1	0.75	1.00	0.85
2	1.00	0.50	0.66
Micro Avg	0.80	0.80	0.80
Macro Avg	0.58	0.50	0.50
Weighted Avg	0.85	0.80	0.78

Probability Distribution (PD) and ROC_AUC Graph – Lap :6

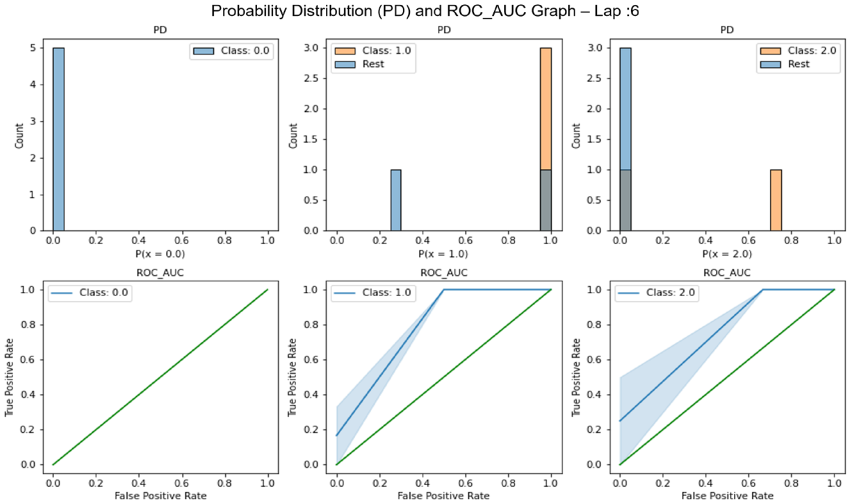

***Figure 5**: Model Results after Lap:6 – Logistics Regression (LR)*

The proposed algorithm has been proven to be efficient in multi-class classification approaches for predicting if the passenger was motion sick prior to their subjective assessments. LR model results are comparable with the state-of-the-art approaches of Hwang et. al. (2022) having an accuracy of (76.26 %) and F1-Score of (0.7350).

4. Conclusions

Indicators from ECG, EDA, and IR sensor sets having profound correlation with MISC scores served as input features and the proposed algorithm classified passengers into different car sickness levels irrespective of the testing conditions. Real-time classification of participants would be susceptible to limitations caused by the feature extraction and model training steps. However, all the performance metrics indicate that this approach can be improved further with high quality training data. Stratified sampling of training datasets would be beneficial for predicting target classes with heterogeneous subgroups. The presented approach demonstrates the potential of multimodal learning models to detect early signs of carsickness symptoms in passengers at an individual level. In-series applicability of such technologies is still in research phase as the evaluation criteria being the accuracy of the sensor systems can be strongly influenced by environmental, body artifacts. In the future, with advancements in wearable devices and having these models incorporated into the decision-making algorithms of driverless autonomous vehicles, driving characteristics could be adapted to enhance ride comfort, passenger well-being and safety.

5. References

Benedek M & Kaernbach C (2010). A continuous measure of phasic electrodermal activity. Journal of Neuroscience Methods, 190(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Hwang J-U, Bang J-S, Lee S-W (2022). Classification of motion sickness levels using multimodal biosignals in real driving conditions, 2022 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1304-1309, doi: 10.1109/SMC53654.2022.9945559.

Keshavarz B, Peck K, Rezaei S, Taati, B. (2022), Detecting and predicting visually induced motion sickness with physiological measures in combination with machine learning techniques. Int J Psychophysiol.; 176:14-26. doi: 10.1016/j.ijpsycho.2022.03.006.

Lipponen JA & Tarvainen MP (2019). A robust algorithm for heart rate variability time series artefact correction using novel beat classification. Journal of Medical Engineering & Technology, 43(3), 173–181. https://doi.org/10.1080/03091902.2019.1640306

Lopez-Martinez D & Elhaouij N & Picard R (2019). Detection of real-world driving-induced affective state using physiological signals and multi-view multi-task machine learning, pp. 356-361, doi: 10.1109/ACIIW.2019.8925190.

Tarvainen MP, Niskanen J-P, Lipponen J A, Ranta-aho PO & Karjalainen PA (2014). Kubios HRV – Heart rate variability analysis software. Computer Methods and Programs in Biomedicine, 113(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Tan R et al. (2022). Motion sickness detection for intelligent vehicles: A wearable-device-based approach, 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), pp. 4355-4362, doi: 10.1109/ITSC55140.2022.9922392

Tauscher J-P, Witt A, Bosse S, Schottky F, Grogorick S, Castillo S, Magnor M (2020). Exploring neural and peripheral physiological correlates of simulator sickness. Computer Animation and Virtual Worlds. 10.1002/cav.1953.

Wagner-Douglas L et al. (2024), Detection of motion sickness in participants through subjective and objective measurement. 70. GfA-Frühjahrskongress 2024, Stuttgart.

Acknowledgments: Special thanks to RWTH Aachen University, Ford Motor Company, and Ford - RWTH Aachen Alliance for their constructive cooperation.

Gesellschaft für Arbeitswissenschaft e.V.

Arbeitswissenschaft in-the-loop

Mensch-Technologie-Integration und ihre Auswirkung auf Mensch, Arbeit und Arbeitsgestaltung

70. Kongress der
Gesellschaft für Arbeitswissenschaft e.V.

Institut für Arbeitswissenschaft und
Technologiemanagement IAT
Universität Stuttgart

In Zusammenarbeit mit dem Fraunhofer-Institut für
Arbeitswirtschaft und Organisation IAO

06. – 08. März 2024

GfA-Press

Bericht zum 70. Arbeitswissenschaftlichen Kongress vom 06. – 08. März 2024

Institut für Arbeitswissenschaft und Technologiemanagement (IAT), Universität Stuttgart
In Zusammenarbeit mit: Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO), Stuttgart

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Sankt Augustin: GfA-Press, 2024
ISBN 978-3-936804-34-8

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle (s. u.) erhältlich.

Geschäftsstelle der GfA
Simone John, Tel.: +49 (0)30 1300-13003, Alte Heerstraße 111, D-53757 Sankt Augustin
info@gesellschaft-fuer-arbeitswissenschaft.de · www.gesellschaft-fuer-arbeitswissenschaft.de

Screen design und Umsetzung
© 2024 fröse multimedia, Frank Fröse,
office@internetkundenservice.de, www.internetkundenservice.de