Journal on Interactive Systems, 2026, 17:1
ISSN: 2763-7719 • doi: 10.5753/jis.2026.5597
This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Paper

Prediction and Analysis of Cybersickness in VR Games Using Symbolic Machine Learning

Wedrey Nunes da Silva [University of Brasilia | *wedrey.silva@aluno.unb.br*]
Thiago Porcino [National Laboratory for Scientific Computing - LNCC | *thiago@lncc.br*]
Carla Denise Castanho [University of Brasilia | *carlacastanho@unb.br*]
Ricardo Pezzuol Jacobi [University of Brasilia | *jacobi@unb.br*]

Department of Computer Science, University of Brasilia (UnB), Campus Darcy Ribeiro, Asa Norte, Brasilia, DF, 70910-900, Brazil.

Abstract. Cybersickness (CS) is one of the main challenges for the adoption of Virtual Reality (VR), manifesting through symptoms such as nausea, dizziness, and eye strain, particularly in Head-Mounted Display (HMD) devices. Although subjective measures, such as questionnaires, are widely used to assess CS, they do not allow for real-time user feedback. This study investigates the role of biosignals in identifying the causes and indicators of CS in VR games, employing Symbolic Machine Learning (SML) to classify the most relevant factors. Our approach combines Electrocardiogram (ECG), Electrodermal Activity (EDA), and Accelerometer (ACC) data with game metrics and user profile attributes. Data were collected from two VR games: a car game and a flying game. Decision Trees and Random Forests were used to build interpretable models, and the results showed that integrating biosignals and game data significantly improves CS prediction, with Random Forest achieving an AUC of 0.95. The findings highlight that exposure time, motion intensity, and electrodermal activity are among the key predictors of CS, reinforcing the importance of physiological monitoring in VR research. Furthermore, the study demonstrates the potential of SML in creating explainable models, contributing to more effective strategies for mitigating CS.

Keywords: Virtual Reality, Cybersickness, Biosignals, HMD Devices, Symbolic Machine Learning, Decision Tree, Random Forest.

Edited by: Joao Marcelo Teixeira | **Received:** 21 February 2025 • **Accepted:** 17 October 2025 • **Published:** 01 January 2026

1 Introduction

Virtual Reality (VR) has emerged as an innovative technology, offering new possibilities for human-computer interaction [LaViola Jr, 2000; Vince, 2004]. However, prolonged exposure to virtual environments can trigger Cybersickness (CS), a condition characterized by symptoms such as nausea, dizziness, eye strain, and headaches. The duration of these symptoms varies from a few minutes to several hours after VR exposure, depending on individual user characteristics. CS not only compromises the quality of the experience but also represents a major challenge for the widespread adoption of VR, often leading to the abandonment of the technology due to the discomfort it causes [LaViola Jr, 2000]. Therefore, it is crucial for developers to implement effective strategies to prevent and mitigate CS from the early design phases, reducing its impact and optimizing user experience [Stanney and Hash, 1998].

To effectively understand and address CS, researchers have focused on developing strategies and methods to assess its occurrence and severity. According to LaViola Jr [2000], three main theories explain the causes of CS: (1) the *Sensory Conflict Theory*, which attributes CS to the discrepancy between sensory information sent to the brain, especially between visual signals and those from the vestibular system [Reason and Brand, 1975; Caserman *et al*., 2021]; (2) the *Postural Instability Theory*, which describes CS as a physiological response resulting from difficulty in maintaining balance and postural control [Riccio and Stoffregen, 1991]; and (3) the *Poison Theory*, which suggests that CS symptoms in virtual environments occur due to adverse stimulation affecting the visual and vestibular systems, leading the body to misinterpret the situation as a possible ingestion of toxic substances, triggering discomfort reactions [LaViola Jr, 2000].

CS is a real problem with medical and safety implications, limiting the use of VR in fields such as education, entertainment, engineering, and video games, where its potential is significant [Kennedy and Lilienthal, 1995]. Studies indicate that between 40% and 60% of VR users may experience moderate to severe symptoms, while approximately 5% remain unable to adapt even after multiple exposures [Kolasinski, 1995; Garcia-Agundez *et al*., 2019]. This condition is particularly common in devices such as Head-Mounted Displays (HMDs), which provide an immersive experience by displaying content directly in the user's eyes. Although recent advances in HMD technology have improved realism and immersion, they may also intensify the discomfort associated with CS [Saredakis *et al*., 2020]. In this context, Kourtesis *et al*. [2019] highlight the importance of considering both hardware advancements and software features as complementary strategies to reduce the occurrence of CS.

Although several theories attempt to explain the causes of CS, there is still no systematic and direct method for measuring it [Krokos and Varshney, 2022]. Researchers often rely on subjective measures, such as self-reported questionnaires, to assess discomfort intensity, providing an overview of symptoms, including nausea, visual discomfort, and disorientation, which vary according to the individual and the virtual environment [Kennedy *et al*., 1993; Islam *et al*., 2020].

However, these approaches, generally applied after the VR experience, limit the accurate monitoring of the condition during immersion. Studies such as the one from Young *et al*. [2007] show that symptom reports are significantly higher when assessments are conducted both before and after exposure, compared to a single questionnaire applied after the experience, highlighting the limitations of this method. These findings emphasize the need for more objective approaches, such as the use of biosignals, which enable the identification of causative factors and indicators of CS, allowing for real-time analysis with greater accuracy.

This article is an extended and revised version of the study titled "Analysis of Cybersickness through Biosignals: an approach with Symbolic Machine Learning," published in the proceedings of the *28th Symposium on Virtual and Augmented Reality* (SVR 2024) [Nunes da Silva *et al*., 2024], organized by the ACM Digital Library and held in Manaus, Brazil, from September 30 to October 3, 2024.

The primary objective of this study is to analyze the influence of biosignals on understanding the factors associated with CS in VR games, aiming to demonstrate a significant correlation between the occurrence of this condition and the recorded signals. The central hypothesis is that integrating quantitative assessments (biosignals and game data) with subjective assessments (data from the Cybersickness Profile Questionnaire – CSPQ), using Symbolic Machine Learning (SML) techniques, represents an effective approach to identifying, classifying, and understanding the causative factors and indicators of CS.

The data collected through the CSPQ questionnaires, along with game information and biosignals such as Electrocardiogram (ECG), Electrodermal Activity (EDA), and body movements captured by an Accelerometer (ACC), were used for training and evaluating the model. During the experiments, participants interacted with two VR games: a car racing and a flying game. The selection of these games was motivated by the possibility of conducting the experiments more efficiently, allowing for the collection and analysis of key metrics such as speed, acceleration, and rotation, which are essential for identifying the factors associated with CS. In contrast, Role-Playing Games (RPGs) frequently interrupt character movements with dialogues, combats, and other interactions, making it difficult to collect continuous and relevant data. Additionally, RPGs would require significantly more time to conduct the experiments.

The transparency provided by symbolic methods is particularly valuable in contexts that require clear and explainable models, such as this research, which aims to identify and understand the elements that influence the manifestation of symptoms related to CS in VR games. To achieve this, we implemented an algorithm based on SML techniques, capable of classifying the main causative and indicative factors of CS. In summary, our contributions include:

- Investigation of the correlation between the levels of CS reported by users and the captured biosignals;
- Generation of rankings of the main indicators and causative factors of CS;
- Comparison of the results obtained with the PDI (*Potential Discomfort Indicator*) method, proposed in this study, and the PCS (*Potential Cause Score*) method, presented in the reference study.

This article is organized as follows: Section 2 presents previous studies that explore the use of biosignals to detect and analyze CS and *Visually Induced Motion Sickness* (VIMS) in virtual environments. Section 3 provides a detailed description of the methodology used, while Section 4 focuses on the approach based on SML. The results, including the scores from the *Virtual Reality Sickness Questionnaire* (VRSQ), physiological changes, and generated rankings, are presented in Section 5. Finally, Section 6 discusses the limitations of this study and presents conclusions and perspectives for future research.

2 Related Work

In addition to subjective measures, it is essential to complement analyses with additional methods, such as quantitative assessments based on participants' physiological signals during exposure to CS. Among the most common biosignals are EEG, ECG (used to calculate Heart Rate - HR and Heart Rate Variability - HRV), Galvanic Skin Response (GSR), and Electrogastrography (EGG). CS analysis through SML, using decision tree-based models, provides greater clarity and reliability in predictions, facilitating the identification of the main factors associated with this condition.

2.1 Analysis with Biosignals

Islam *et al*. [2020] conducted a study that analyzed the detection and prediction of CS in a VR roller coaster simulation. During the experiment, physiological data such as HR, GSR, breathing rate (BR), and HRV were collected. The virtual environment was developed in Unity 3D, and participants were exposed to both rest and activity periods while their biosignals were recorded. Initially, the data were analyzed using a Support Vector Machine (SVM) model, followed by an architecture combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). The study identified that most participants experienced CS, with strong correlations between self-reported scores on the Fast Motion Sickness (FMS) scale and HR, GSR, and HRV data. The CNN-LSTM model outperformed the others, demonstrating efficiency in detecting and predicting CS based on physiological biosignals.

The *Simulator Sickness Questionnaire (SSQ)* presents limitations as it interrupts the user during the experience, compromising results and making real-time quantitative measurements unfeasible. To overcome this challenge, [Krokos and Varshney, 2022] investigated CS induced by vection in immersive virtual environments using EEG signals. With 44 participants, the study revealed a correlation between CS symptoms reported via joystick and increases in Delta, Theta, and Alpha waves. The experiment consisted of a 61-second simulation in a virtual environment with both abrupt and smooth camera movements, using an HTC Vive head-mounted display synchronized with a 14-channel EEG. Participants reported greater discomfort with abrupt changes and difficulty in anticipating movements, suggesting that direct control could mitigate symptoms. Approximately 70% of participants tilted their bodies according to movements, and 32% had prior expe-

rience with immersive devices. The authors emphasized the need for evaluation standards for hardware (headsets, trackers, and monitors) as well as content (games, performances, and other immersive experiences), along with further investigations into factors such as specific tasks, immersion duration, age, and sex in CS.

The study by Garcia-Agundez *et al*. [2019] investigated the use of a two-lead ECG to detect CS in 13 VR players during 15-minute sessions, recording ECG data and applying the SSQ questionnaire. Four participants interrupted the experiment due to CS, and data from two were discarded due to artifacts. The results showed statistically significant differences in SSQ and ECG scores, with an average increase of 8 points in the SSQ. Surprisingly, participants with higher levels of CS exhibited lower heart rates. The analysis indicated a higher correlation between Standard Deviation of the Normal-to-Normal intervals (SDTNN) and SSQ, particularly in oculomotor symptoms (r = 0.47) and disorientation (r = 0.38), highlighting the complex relationship between CS, ECG, and specific symptoms, in contrast with previous studies.

Keshavarz *et al*. [2022] explored the use of physiological data such as ECG, EDA, EGG, respiration, body temperature, skin temperature, and body movements, combined with Machine Learning (ML) techniques, to identify and predict the onset of VIMS. The experiment included 43 participants who watched a video designed to induce VIMS for 15 minutes. The analysis with ML models showed a moderate correlation between physiological measures and the scores on the FMS Scale. The authors emphasized that, although useful, physiological measures alone are not sufficient for reliable real-time detection or prediction of VIMS. Among the evaluated models, Random Forest (RF) stood out, achieving an Area Under the Curve (AUC) of 0.75, effectively distinguishing participants who experienced symptoms from those who did not.

2.2 CS Analysis with Symbolic ML

Porcino [2021] proposed an experimental analysis based on SML to identify the possible causes of CS in VR games. The research is based on the hypothesis that symbolic models, by offering a clear and interpretable representation of the results, can help game designers identify and implement more effective strategies to mitigate CS symptoms. To collect the data, two games were developed in Unity 3D: a racing game and a flying game. During the experiments, participants engaged in sessions of 5 or 20 minutes and answered the CSPQ and VRSQ questionnaires before and after the sessions, reporting their level of discomfort.

In addition to subjective evaluations, real-time data were collected during the gameplay, including acceleration, head orientation, and scene position. These data, combined with up to eight factors attributed to CS (acceleration, exposure time, frame rate, age, rotation, gender, speed, and prior VR experience), were analyzed using two SML algorithms: Decision Tree and Random Forest. These algorithms generated rankings of the possible causes of CS, facilitating the interpretation and identification of strategies to mitigate symptoms. The authors also reviewed previous studies that proposed approaches for factors such as exposure time, acceleration, speed, frame rate, and camera rotation on the z-axis [Bouyer *et al*., 2017; Budhiraja *et al*., 2017], but noted that aspects related to user profile, such as gender, age, and experience, still lack clear strategies.

3 Method and Experiments

This section describes the procedures used to conduct the experiments in this study, covering the biosignals and games employed, the data collection methods through qualitative questionnaires, the techniques applied to obtain quantitative data, and the preprocessing stage of the collected data.

3.1 Software and Equipment

The simulation was conducted using the HTC Vive HMD, which features a resolution of 1080 x 1200 pixels per eye, a 90Hz refresh rate, and a 110-degree field of view. The operating system used was Windows 10 Pro (x64), running on a computer equipped with an Intel Core i7 processor, 16 GB of RAM, and an NVIDIA GeForce GTX 1050 Ti graphics card, capable of running games developed in Unity 3D, SteamVR, and biosignal collection software. As shown in Figure 1, two VR games were used: a car racing game and a flying game, both implemented in Unity 3D as part of the study conducted by Porcino [2021]. Biosignal collection was performed using the BITalino (r)evolution Plugged, a portable, modular, and cost-effective device widely used in scientific research. This equipment enables the precise acquisition of physiological signals such as ECG, EMG, EEG, EGG, and ACC, with sampling rates of 1, 10, 100, or 1000Hz and communication via Bluetooth/BLE.

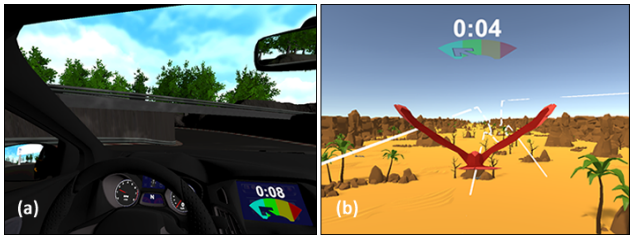


Figure 1. Car Racing Game on the left (a) and Flying Game on the right (b) [Porcino, 2021].

The Biosignal Collector application (BC) (Figure 2) was designed to enhance biosignal collection and allow participants to report CS levels using voice commands, based on a scale from 0 (no symptoms) to 3 (severe). Integrated with the OpenSignals software via the TCP/IP protocol, the BC collects data from the BITalino and generates output in JSON format, supporting frequencies of 10Hz, 100Hz, or 1000Hz, as well as enabling customized configuration of channels (1 to 6) and sensors. The tool organizes the data by sensor, including the reported CS levels, and allows for visualization and saving of customized graphs, suitable for analysis or CNN model training. Written in C# for the Windows Desktop environment, the BC features a user-friendly and efficient interface, simplifying the preprocessing and analysis of the collected data.

3.2 Qualitative Data and Game Data

Qualitative data were collected during participants' interaction with the game software. The CSPQ questionnaire, con-

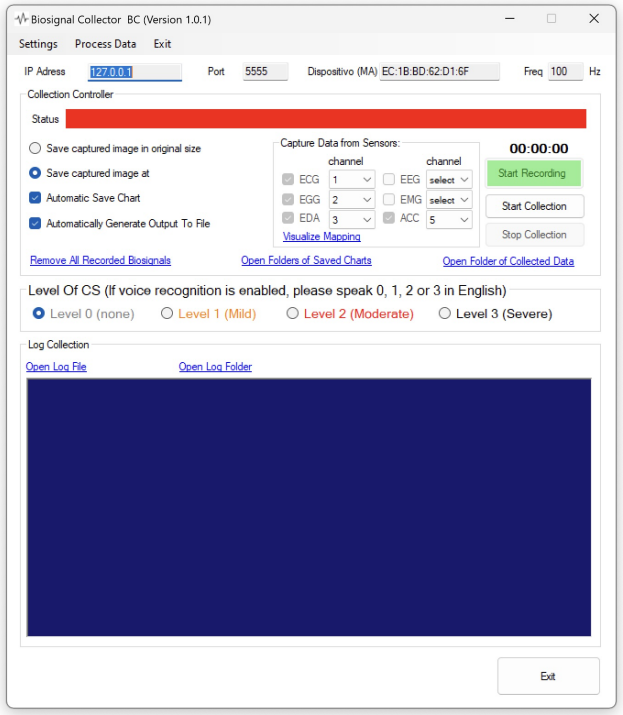


Figure 2. BC Main Screen

sisting of nine questions, explored aspects such as gender, age, prior experience, pre-existing symptoms, flicker sensitivity, glasses usage, visual impairments, posture during gameplay, and dominant eye, as described by Porcino [2021]; Porcino *et al*. [2022]. On the other hand, the VRSQ questionnaire evaluated nine symptoms related to CS, including general discomfort, fatigue, eye strain, difficulty concentrating, headaches, a "heavy head" sensation, blurred vision, dizziness, and vertigo. Additionally, game data were continuously recorded throughout the experiment at a rate of one instance per second, capturing information such as speed, acceleration, rotations, regions of interest, and other variables.

3.3 Biosignals

As previously mentioned, our study utilized biosignal data from ECG and EDA, as well as body movements captured by an ACC. The literature [Garcia-Agundez *et al*., 2019; Qu *et al*., 2022; Islam *et al*., 2020] highlights that EDA and ECG biosignals are widely recognized as important indicators of CS. Furthermore, according to LaViola Jr [2000], abrupt user movements can be useful in identifying CS and are frequently used in studies on the topic, as exemplified in the work of Jeong and Han [2024].

3.3.1 Electrocardiogram

The ECG is the recording of the electrical activity generated by the heart on the body's surface [Alberdi *et al*., 2016]. It is widely used for the assessment of heart diseases and cardiac rhythm disorders. This recording provides valuable information about heart rate, rhythm, and potential arrhythmias, enabling a more precise analysis of the effects of CS on the cardiovascular system. Typically, the heart rate range is between 60 and 100 beats per minute (bpm), corresponding to an RR interval of 0.6 to 1 second [Qu *et al*., 2022].

ECG signals allow for the analysis of HRV, providing essential information about the heart's electrical activity. In this study, the RR interval was used as an input value every second in the dataset employed for the SML model.

The RR interval (RR-I) is one of the HRV parameters and corresponds to the time between two consecutive heartbeats. This value is obtained by measuring the distance between successive RR peaks in the ECG signal, reflecting the duration of the cardiac cycle, as illustrated in Figure 3. This metric is widely used to assess heartbeat regularity and temporal variation. A shorter RR-I is associated with an increased heart rate, while a longer RR-I indicates a reduced heart rate.

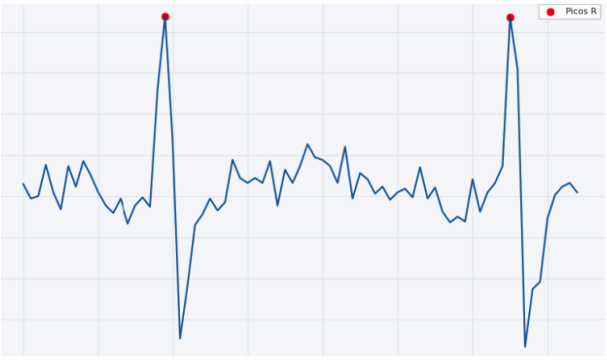


Figure 3. Example of an ECG with RR Interval highlighted.

3.3.2 Electrodermal Activity

CS can trigger emotional and psychological responses that manifest as changes in electrodermal activity. In the study conducted by Qu *et al*. [2022], the relationship between CS symptoms and EDA was investigated, revealing that individuals experiencing more intense symptoms exhibited a significant increase in skin conductance. These findings suggest that electrodermal response may serve as a robust indicator for assessing the severity of CS and its effects on the body.

EDA corresponds to variations in the skin's electrical conductance in response to emotional or psychological stimuli and is widely used as a physiological measure to assess autonomic nervous system activity. It is recorded through electrodes placed on the skin's surface, preferably in regions with higher sweat gland activity, such as the palms of the hands, fingers, or the soles of the feet. The unit of measurement for EDA is microsiemens (μS), with values typically ranging between 5 and 50 [Qu *et al*., 2022].

3.3.3 Accelerometer

The ACC is an electromechanical device used to measure acceleration forces, playing a crucial role in various technological applications. Since its popularization in the 1990s with the advancement of Microelectromechanical Systems (MEMS), these devices have revolutionized industries such as automotive and consumer electronics by enabling functionalities like hard drive protection and motion control in games [Andrejašic, 2008]. In addition to measuring displacements and providing data on the angle and magnitude of movements, ACCs are used to analyze rapid and abrupt motions during exposure to virtual environments, allowing researchers to understand their relationship with CS. These devices capture acceleration data in units of gravitational force (g) or meters per second squared ($\mathrm{m/s^2}$) [PLUX Biosignals, 2024], making a significant contribution to studying discomfort associated with human movement [Bassett and John, 2010].

3.4 Biosignal Acquisition

As illustrated in Figure 4, the electrodes used in the Einthoven lead configuration for ECG were positioned with IN+ (red) and IN- (black) attached to the clavicles, while the REF (white) was placed over the iliac crest. For EDA data collection, the electrodes were applied to the participant's fingers. The ACC sensor was fixed at the suprasternal notch using adhesive tape to ensure stability during the experiments.

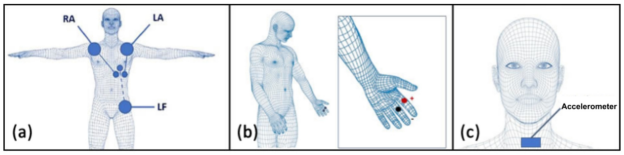


Figure 4. a) Electrode placement for ECG [PLUX Biosignals, 2023a]; b) EDA on fingers [PLUX Biosignals, 2023b]; c) ACC on the user's suprasternal notch.

3.5 Participants

The selection criteria adopted in this experiment were designed to ensure the safety of participants, excluding those with severe health conditions, such as extreme vestibular disorders and severe balance problems, who are more susceptible to intense CS. The recruitment process utilized various outreach methods, including sending messages in WhatsApp groups and individual approaches. As expected, most participants were students from the University of Brasilia (UnB), since the dissemination efforts were primarily targeted at this audience.

3.6 Procedure

To conduct this experiment, the participation of all volunteers was ensured through explicit consent. Each participant confirmed their agreement to participate anonymously by signing the informed consent form.

Upon arrival at the laboratory, participants were given time to read the consent form and receive basic instructions about the experiment. These instructions included information about the game, the recommended posture for greater comfort, and the recommendation to avoid conversations during the experience, except in case of the need to stop due to discomfort. Participants were instructed to describe their initial state using a symptom scale (in English): 0 for no CS symptoms, 1 for mild symptoms, 2 for moderate symptoms, and 3 for severe symptoms. After this, they were equipped with the HMD and the electrodes/sensors.

To minimize the impact on participants, considering the use of VR devices and body electrodes, an exposure time of 10 minutes was chosen, different from the 5 and 20 minute intervals proposed by Porcino [2021]. This time was divided into 5 minutes for baseline data collection and 5 minutes for the VR game experience. Additionally, 2 to 3 minutes were allocated for completing the questionnaires at the end of the session.

Figure 5 presents the experiment flow, organized into four distinct phases, providing a clear overview of the steps followed by the participants. Initially, a five-minute collection of physiological signals was conducted during the resting phase. Next, participants completed the CSPQ and Pre-VRSQ questionnaires, providing profile information before the experience. During the five-minute interaction with the game, game data, biosignals, and continuous verbal reports on CS symptoms were collected. Finally, after the VR experience ended, participants completed the Post-VRSQ questionnaire.

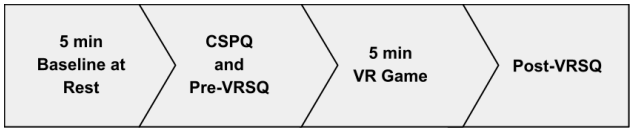


Figure 5. Experiment Flow.

3.7 Collected Data

As presented in Table 1, the study involved 17 participants distributed across two games, with 8 in the car racing game and 9 in the flying game. Among the participants, 8 were female and 9 were male, with ages ranging from 18 to 50 years.

Table 1. Experiments conducted and characteristics of the collected data.

Game	Gender	Age	Exposure	Users
Car	Female	37 to 50	5 (min)	2
Car	Female	18 to 36	5 (min)	2
Car	Male	18 to 36	5 (min)	4
Flight	Male	18 to 36	5 (min)	5
Flight	Female	18 to 36	5 (min)	4
Total				**17**

During the experiments, technical issues with Bluetooth connectivity between the Bitalino board and the computer running the OpenSignals software resulted in the exclusion of data from participants 2 and 4, as it was impossible to synchronize the biosignals, game information, and discomfort reports. Additionally, participant 6 was excluded from the analyses since, despite scoring on the VRSQ, he/she did not report discomfort during the experiment. In total, 2,309 instances were collected in the car racing game and 2,408 in the flying game.

3.7.1 Structure of Data Obtained from the Games

Table 2 presents the organization of the data collected during the experiments conducted in the car racing and flying games. The first section of the table describes the attributes related to game data, while the second section details the user attributes collected through the CSPQ questionnaire.

3.7.2 Structure of Data Obtained from the BC

The structure of the biosignals collected using the BC tool is described in Table 3. Each data instance is associated with a TimeStamp attribute. The LevelCS attribute represents the CS level verbally reported by the user during the gameplay activity. Finally, the ECG, EDA, and ACC attributes correspond to the data captured by their respective sensors.

List 1 provides an example of the raw biosignal data recorded during the gameplay activity. Each entry represents a data instance collected at a specific moment throughout the experiment.

3.8 Preprocessing

During the data collection process, it is common to encounter noise and anomalies that may compromise the accuracy of the

Table 2. Attributes of data collected through games. Types: Numeric (N) and Categorical (C).

Game Data		User Data	
Attributes	Type	Attributes	Type
TimeStamp	N	Gender	C
Speed	N	Age	C
Acceleration	N	VR Experience	C
Rotation (x, y, z)	N	Flicker Sensitivity	C
Position (x, y, z)	N	Pre-symptoms	C
Region of Interest	N	Uses Glasses	C
FoV Size	N	Visual Impairments	C
Frame Rate	N	Posture	C
Static Frame	C	Dominant Eye	C
Haptic Feedback	C	Discomfort Level	N
Degree of Control	C		
DoF Simulation	C		
Player Locomotion	C		
Automatic Camera	C		

Table 3. Structure of the collected biosignals. Attributes: Numeric (N) and List (L).

		Biosignal Values		
Timestamp	LevelCS	ECG	EDA	ACC
N	N	L	L	L

```
[
  {
    "TimeStamp": 0,
    "LevelCS": 0,
    "ECG": [0.097, 0.105, 0.108, 0.064, 0.044,
    ↪  0.035, 0.009, -0.009, -0.041, -0.056],
    "EDA": [21.09, 21.09, 21.09, 21.09, 21.09,
    ↪  21.09, 21.09, 21.09, 21.09, 21.09],
    "ACC": [-0.663, -0.654, -0.654, -0.663,
    ↪  -0.654, -0.654, -0.644, -0.663, -0.663,
    ↪  -0.654]
  },
  "..."
  {
    "TimeStamp": 301.497,
    "LevelCS": 2,
    "ECG": [-0.085, -0.1, -0.103, -0.1, -0.103,
    ↪  -0.088, -0.079, -0.062, -0.05, -0.032],
    "EDA": [24.68, 24.68, 24.68, 24.68, 24.71,
    ↪  24.71, 24.71, 24.71, 24.73, 24.73],
    "ACC": [-0.615, -0.625, -0.606, -0.615,
    ↪  -0.606, -0.625, -0.615, -0.606, -0.606,
    ↪  -0.625]
  }
]
```

Listing 1: Example of Raw Data collected in JSON format.

analysis. These interferences can lead to misinterpretations by researchers, affecting the reliability of the results. In this context, preprocessing is essential for analysis [Awal *et al.*, 2014], as it aims to reduce noise and anomalies, enabling a more precise identification of symptoms associated with CS.

In this study, biosignals were originally collected at a rate of 100 Hz. However, to ensure compatibility with the game data used in the experiments [Porcino, 2021] and to meet the research requirements, the sampling rate was reduced to 1 Hz. This adaptation allowed the alignment of biosignal collection frequency with the game's standard, which records one sample per second.

Raw physiological signals underwent a preprocessing stage using algorithms developed in Python specifically for this purpose. These algorithms were designed to facilitate the visual inspection of biosignals, and in some cases, such as ECG, filters including low-pass, high-pass, and median were applied. Additionally, the captured biosignal data was integrated with the game data, including user profiles and information related to in-game interactions.

For ECG data, the reduction in sampling rate from 100 Hz to 1 Hz was achieved by segmenting the data into 10-second blocks (1,000 values) and extracting 10 RR intervals per block. When it was not possible to identify all 10 intervals due to factors such as bradycardia or signal noise, the last available RR interval was repeated to complete the list. The extracted RR intervals were then organized into a global list and linked to the game data, being associated with each second of collection.

For EDA data, integration with game data involved reducing the sampling rate to 1 Hz. This process was performed by calculating the average of values collected at 100 Hz and recording only the mean value per second. This approach ensured the preservation of relevant biosignal information.

The player's body movement data, obtained via an ACC, underwent the same procedure applied to EDA, reducing the sampling frequency from 100 Hz to 1 Hz. The average acceleration values were calculated in 1-second intervals. However, due to a technical limitation of the sensor used in this study, which only records movements along the Z-axis, the average was based on the frontal acceleration values of the participant's movements. These data were then integrated into the dataset used for model training and classification.

4 Approach with Symbolic ML

This study presents an approach that advances beyond the method proposed by Porcino [2021]; Porcino *et al.* [2022] by incorporating biosignal data for analysis. By utilizing RR intervals extracted from ECG, EDA, and body movement recordings obtained from an ACC, it is possible to conduct a more detailed assessment of the causes of CS and establish correlations between reported symptoms and individuals' physiological responses. This methodology significantly enhances the ability to identify symptoms, playing a crucial role in advancing understanding and developing effective strategies to prevent and mitigate CS.

4.1 Proposed Approach

In the proposed approach, illustrated in Figure 6 (a), training begins with the collection of qualitative data (CSPQ and VRSQ), along with quantitative data from the game and biosignals. Next, the data undergoes a preprocessing stage, followed by a newly incorporated phase in this study that allows for model selection (with or without the inclusion of biosignals). The remaining steps of the original reference model were preserved.

The classification process starts with a set of instances from a specific user, followed by data preprocessing. Subsequently, model selection occurs, which may or may not

consider biosignals. After this selection, the data is processed by the trained model. If the result indicates any level of discomfort, the decision path is examined, generating a ranking of the main causative or indicative factors of CS.

To enhance data collection, not only were quantitative questionnaires and game data included, but also the participants' physiological signals. During this stage, biosignal and game data were collected simultaneously and consolidated into a single file per experiment, ensuring an integrated and comprehensive analysis. Based on the reference work, we developed a complete workflow covering biosignal collection to CS cause identification. We applied two distinct ranking methods: the *Potential-Cause Score (PCS)* [Porcino, 2021], presented in the reference study, and the *Potential Discomfort Indicator (PDI)*, proposed in this work, expanding the possibilities for analysis, interpretation, and comparison of results between methods.

4.2 Symbolic Classifiers

Symbolic classifiers are algorithms that use symbols and logical rules to capture and describe patterns in training data. These patterns are represented by a language composed of a set of N_R positional "if-then" rules, described as $\mathbf{h} = \{R_1, R_2, \ldots, R_{N_R}\}$ [Bernardini, 2006]. Each rule is derived directly from the training data and can be adjusted or expanded as the system incorporates new information. These rules consist of disjunctions of conjunctions of attribute tests in the form X_i *op* Value, where X_i represents an attribute, *op* is an operator belonging to the set $\{=, \neq, <, \leq, >, \geq\}$, and *Value* is a valid value associated with the attribute X_i.

Each rule R takes the form **if** B **then** H or symbolically $B \rightarrow H$, where H is the head or conclusion of the rule, and B is the body or condition of the rule. Both H and B are sets of attribute tests with no common attributes. In a classification rule, the head of the rule H takes the form class $= C_i$, where $C_i \in \{C_1, \ldots, C_{N_{Cl}}\}$ [Bernardini, 2006].

In the context of this work, symbolic classifiers based on Decision Tree (DT) were chosen as a suitable approach for data classification due to their ability to provide a transparent and easily interpretable visual representation of the model's decision-making process. Once a DT is constructed, it can be used to classify new examples, allowing for a clear and intuitive understanding of the classification rules employed by the model.

Decision Trees are a ML technique that decomposes complex problems into simpler subproblems. Their structure is recursively defined by leaf nodes, which correspond to classes, or by decision nodes, which contain tests on specific attributes. Each outcome of these tests leads to an edge that connects to a subtree with the same hierarchical structure [Monard and Baranauskas, 2003]. During the construction of a DT, the algorithm seeks the best division of the data at each node, aiming to maximize the homogeneity of the resulting groups.

To evaluate the quality of splits at each node, various metrics are used to measure the degree of impurity or non-homogeneity of a dataset. Among the most common metrics are Entropy and the Gini Index, which play fundamental roles in selecting the most discriminative attributes at each stage of the tree construction. Entropy, based on the axioms established by Shannon [1948], is a measure that quantifies the impurity or disorder of a training dataset. At each hierarchical level of a Decision Tree, the attribute with the highest information gain is selected as the tree node. The mathematical expressions for calculating Entropy and information gain are presented below:

$$\text{Entropy}(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \tag{1}$$

In Equation 1, S represents the dataset or a subset of the data; c is the number of classes in the dataset; and p_i is the probability of class i in the dataset.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \cdot \text{Entropy}(S_v) \tag{2}$$

The calculation of Gain(S, A), presented in Equation 2, evaluates the information gain when splitting the data based on attribute A. In this context, A represents an attribute or feature, while Values(A) denotes the possible values that A can assume. The subset S_v corresponds to the data in S where the attribute A has the value v. Meanwhile, $|S_v|$ and $|S|$ represent the number of instances in the subsets S_v and S, respectively.

The *Gini Index*, adopted in this work, measures the probability that two randomly selected elements from the same subset have different class labels. According to Rokach and Maimon [2005], the *Gini Index* is an impurity-based criterion that evaluates the divergences between the probability distributions of the target attribute values. This metric has been widely used in various studies [Daniya *et al*., 2020; Rokach and Maimon, 2005] and is defined as:

$$Gini(D) = 1 - \sum_{i=1}^{C} (p_i)^2 \tag{3}$$

In Equation 3, D represents the dataset, and C denotes the number of classes present in this dataset. The probability of an item belonging to a specific class or category is indicated by p_i. For each class, the square of the probability (p_i^2) is calculated, and all these squared terms are summed. This value is then subtracted from one (1) to obtain the Gini index.

The Gini index ranges between 0 and 1. A Gini value of 0 indicates a perfectly homogeneous distribution, where all items belong to the same class. Conversely, a Gini value of 1 reflects a completely impure or heterogeneous distribution. This metric is useful for quantifying the impurity or disorder of a dataset based on the distribution of its classes.

4.3 Classifier Evaluation

In ML, an important step is to validate the data and ensure that the trained model can make accurate and useful predictions in new scenarios. Among the available tools for this purpose, Weka (Waikato Environment for Knowledge Analysis) stands out as a powerful data mining platform. Developed in Java by the University of Waikato in New Zealand, Weka was designed to provide researchers with easy access to advanced

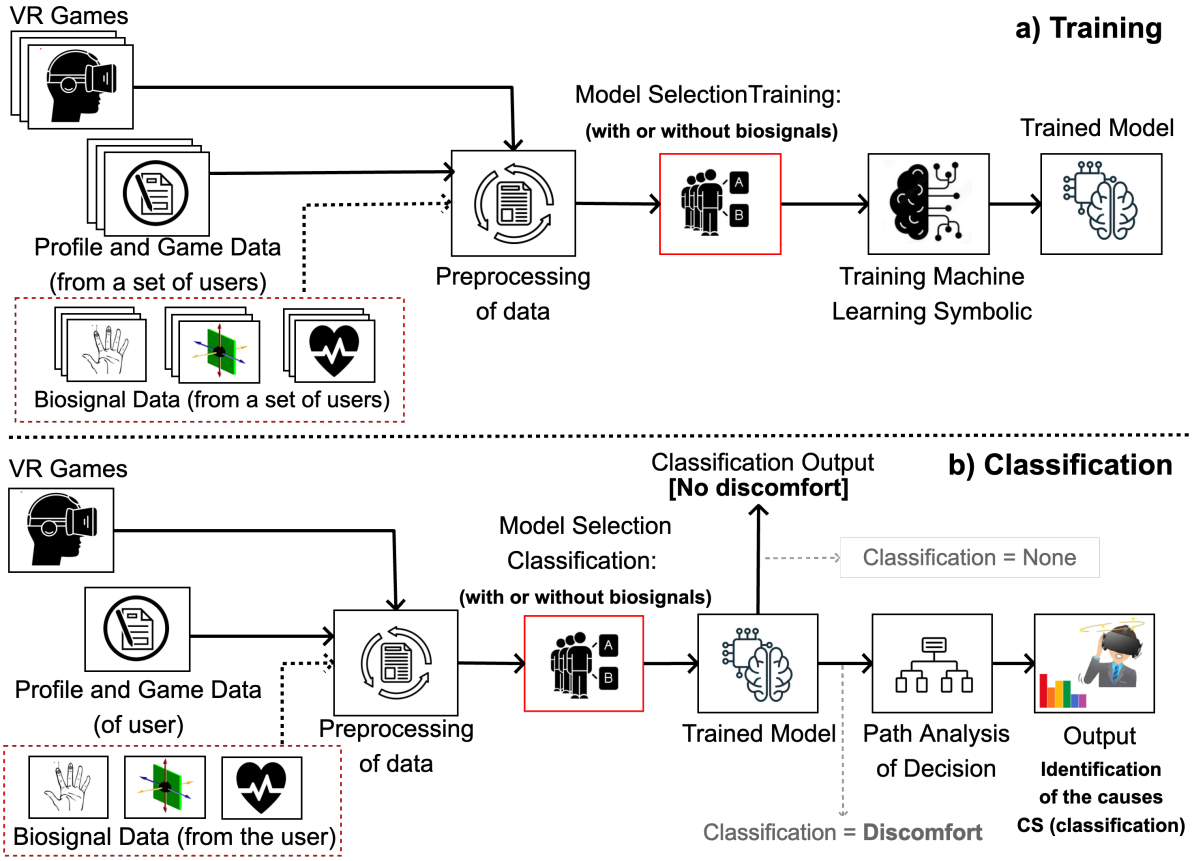


Figure 6. Proposed Approach.

ML techniques. The tool offers a comprehensive and practical solution with complete features for data preprocessing, visualization, and evaluation.

Based on the collected and preprocessed data, we used the Weka tool to apply various ML algorithms for CS classification. Three datasets (Datasets A, B, and C) were organized, containing game information, user profile data, and biosignals, as presented in Tables 2 and 3.

- *Dataset A*: 2309 samples (Car Racing Game).
- *Dataset B*: 2408 samples (Flying Game).
- *Dataset C*: Total of 4717 samples (Combination of Datasets A and B).

The datasets used in this study include discomfort levels assigned by users, classified on a scale from 0 to 3. Although the multiclass format primarily aims to provide a biosignal database that can be leveraged in future research, in this study, we opted for binary classification, which was sufficient for the analyses conducted. In this format, the output classes were reorganized into two categories: 0 = None and 1 = Discomfort, with the latter encompassing Mild, Moderate, and Severe levels.

In this study, we used Weka to evaluate key decision tree (DT)-based algorithms, such as CSForest, DecisionStump, ForestPA, HoeffdingTree, J48, LMT, ExtraTree, RandomForest, RandomTree, and REPTree. To validate the classification models, we applied the k-fold cross-validation technique, configured with 10 splits (Folds=10), for all the symbolic classifiers evaluated. K-fold cross-validation (KCV) is a widely adopted approach by professionals and is essential for both model selection and classifier error estimation. This technique divides the dataset into k subsets and, iteratively, uses a portion for model training while the remaining subsets are used to evaluate its performance [Anguita *et al.*, 2012].

For this analysis, we performed a binary classification by defining two categories: 0 = None and 1 = Discomfort. The RF algorithm stood out in all evaluated configurations, both with and without the inclusion of biosignal attributes, demonstrating superior performance. The detailed results of the binary classification with biosignals are presented in Table 4. Specifically, the RF algorithm achieved accuracy rates of 99.78%, 99.50%, and 99.62% in datasets A, B, and C, respectively. As a result, it was selected as the best classifier.

4.4 Attribute Evaluation

Our main objective is to deepen the understanding of the causes related to discomfort in virtual environments, focusing on predicting the occurrence of CS. To achieve this, it is essential to evaluate the attributes that influence the decision-making process of classification models. After identifying the best classifier, we used the *scikit-learn* library to create a ranking of the most relevant attributes. The chosen evaluation technique was *leave-one-out* cross-validation, which involves training the model multiple times while removing one attribute at a time from the training set and evaluating the model's performance on the test set. This approach is widely used in ML model evaluation, especially for small datasets [Tan *et al.*, 2018]. Moreover, various studies in the literature have already applied this technique to assess RF models [Pashaei and Pashaei, 2019; Yao *et al.*, 2019; Valkonen *et al.*,

Table 4. Binary Classification with Biosignals (Weka).

Classifier	ACC (A)	KPP (A)	ACC (B)	KPP (B)	ACC (C)	KPP (C)
CSForest	99.35%	0.9837	97.80%	0.9533	98.62%	0.9687
DecisionStump	84.54%	0.5982	63.08%	0.0000	69.73%	0.1731
ForestPA	99.61%	0.9901	98.80%	0.9741	99.32%	0.9844
HoeffdingTree	72.76%	0.0000	87.04%	0.7303	78.67%	0.5273
J48	99.70%	0.9923	98.75%	0.9733	99.22%	0.9820
LMT	99.57%	0.9891	98.59%	0.9696	99.13%	0.9801
ExtraTree	98.31%	0.9572	97.76%	0.9519	98.56%	0.9669
RandomForest	**99.78%**	**0.9945**	**99.50%**	**0.9893**	**99.62%**	**0.9912**
RandomTree	98.48%	0.9616	97.34%	0.9427	97.77%	0.9490
REPTree	99.65%	0.9913	98.05%	0.9582	99.17%	0.9810

2017].

The construction of the ranking of the best attributes for the classification model followed a well-defined approach. Initially, the algorithm was executed on all three datasets (A, B, and C), incorporating the biosignal attributes. Subsequently, these attributes were removed from each dataset, and the classification process was repeated to evaluate the impact of their exclusion. We opted to use ten attributes in the model without biosignals and thirteen in the complete model, including ACC, EDA, and ECG. This methodology allowed for a clear analysis of the importance of physiological data in improving the model's performance.

In the analysis of attributes without biosignals, the *TimeStamp* (Exposure Time) attribute stood out as the most relevant for Datasets A and C, as highlighted in previous studies [Porcino, 2021; Porcino *et al*., 2022]. Other attributes widely supported by the literature also demonstrated high relevance, such as *Flicker* [LaViola Jr, 2000], *Rotation* [Fomenko and Kaewpankan, 2022], and *Gender* [Reason and Brand, 1975]. These results emphasize the importance of specific contextual and user-related characteristics in the decision-making process.

During the attribute selection process, we alternated between the top eight attributes identified in Datasets A and B. The frame rate (*FPS*), an attribute highlighted by Porcino [2021]; Porcino *et al*. [2022], was considered relevant both in the games used in their study and in our research, and was therefore incorporated into the model. Additionally, although it did not rank among the top eight attributes, the *Speed* attribute was also included due to its proven importance in triggering CS, as reported in previous studies [Oh and Son, 2022; Porcino, 2021; Porcino *et al*., 2022].

Finally, the selected attributes for composing the models with and without biosignals were organized based on a detailed analysis of the classifier's performance in different scenarios. These attributes, essential for CS prediction, are listed in Table 5, highlighting the relevance of characteristics such as ACC, EDA, ECG, along with other environment- and user-related features. This robust selection reinforces the importance of combining contextual and physiological information to achieve more accurate predictive models.

4.5 Statistical Procedure

The statistical analysis focused exclusively on numerical variables, since nominal categorical variables such as `Flicker`, `Region Of Interest`, and `Gender` are not appropriate for the comparative tests employed. The variable `DiscomfortLevel` was used only as a grouping factor between the classes (with and without *CS* symptoms), and not as a direct variable of analysis.

Table 5. Set of Selected Attributes.

Game Data	User Data	Biosignal Data
TimeStamp	Gender	ACC
Rotation X	Flicker	EDA
Rotation Y	Visual Deficiencies	ECG
Rotation Z		
Region of Interest		
Frame Rate		
Speed		

Initially, the Shapiro–Wilk test [SHAPIRO, 1965] was applied to assess the normality of the numerical variables. As the normality assumption was violated for most variables, a non-parametric approach was adopted. The Kruskal–Wallis test [Kruskal and Wallis, 1952] was used to compare the classes, as it is appropriate for analyzing data that do not follow a normal distribution. In cases of statistical significance ($p < 0.05$), Dunn's *post hoc* test [Dunn, 1964] with Bonferroni correction was performed to identify significant differences between the groups.

The analyses were conducted individually for each participant, considering nine numerical variables: `TimeStamp`, `RotationX`, `RotationY`, `RotationZ`, `Frame Rate`, `Speed`, `ACC`, `EDA`, and `ECG`. Each of these variables was analyzed separately in relation to the *CS* groups. The results presented in the following subsections were consolidated from these individual analyses, providing evidence regarding the consistency and predictive potential of the collected data.

4.5.1 Statistical Analysis

The statistical analysis of both *datasets* revealed relevant patterns in distinguishing users with and without *CS* symptoms. The `TimeStamp` variable was consistent across all participants in both games ($p < 0.05$), confirming its robustness as a temporal marker associated with discomfort. The `ACC` variable also showed high relevance, being significant in all cases in the car game and in several participants in the flight game, reinforcing its potential as a physiological indicator. Variables such as `ECG`, `EDA`, `RotationX`, `RotationY`, and `Speed` demonstrated utility in specific contexts, suggesting a complementary contribution to multivariate models. Conversely, `Frame Rate` and `RotationZ`, although significant only in occasional situations, may add predictive value when

combined with more robust variables. These results highlight the importance of multivariate approaches in developing ML models for CS detection.

4.6 Model Evaluation

We adopted *leave-one-out* cross-validation to evaluate the models. In this method, the algorithm is trained with all participants except one, using the excluded participant to test the model. This procedure is repeated iteratively, alternating the removed participant until all have been used for testing.

Additionally, we conducted an analysis to identify the optimal tree depth in the model built with the RF algorithm. Our goal was to maximize AUC scores while minimizing tree depth, seeking a balance between performance and efficiency in model configuration.

4.6.1 Maximum Tree Depth

The RF algorithm stood out as the best-performing one in the analyses conducted, as previously evidenced. Unlike a simple decision tree, RF combines multiple trees, each trained on random subsets of the training data, providing greater robustness and generalization capability. However, this approach also has a high computational cost due to the need to build and integrate multiple trees.

To optimize the training process in the experiments, we explored different depth configurations in the datasets of the car racing and flying games. Both datasets were split into 70% for training and 30% for testing. The results of the AUC scores as a function of the tested depths are presented in Figure 7. We identified that the optimal depth varied between games: in the Flying Game, the best performance was obtained with a depth of 5, while in the Car Racing Game, a depth of 7 proved to be more efficient.

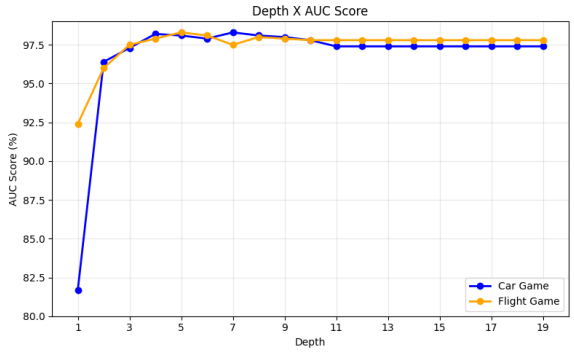


Figure 7. Tree Depth vs. AUC Score (with biosignals).

4.6.2 Model Parameter Configuration

As previously demonstrated, the RF algorithm outperformed the DT algorithm in terms of robustness and generalization capability. However, training RF requires higher computational capacity. During the experiments conducted with the car racing and flying games, we used a 70% training set and a 30% test set. The analysis of AUC scores revealed that the optimal depth was 5 for the Flying Game and 7 for the Car Racing Game.

The RF model parameters were configured based on technical criteria and evidence from the literature:

- **n_estimators = 100**: number of trees in the forest, widely used in previous studies [Wang *et al.*, 2025; De Miras *et al.*, 2023] for balancing predictive performance and processing time. Higher values may increase robustness, but marginal gains often do not offset the additional computational cost.
- **criterion = 'gini'**: Gini index as the impurity criterion, chosen for its computational efficiency and good performance in similar tasks, aligned with the approach of this study.
- **max_depth = 5 or 7**: limits the maximum depth, reducing the risk of overfitting. A value of 5 was more suitable for the Flying Game, while 7 achieved better performance in the Car Racing Game, adapting to the specific characteristics of each task.
- **random_state = 42**: sets the seed for the random number generator to ensure reproducibility, with 42 being a conventional choice without direct impact on performance.

Parameters not specified followed the default values of the `scikit-learn` library.

4.6.3 AUC Scores

AUC is widely recognized as an important evaluation criterion in ML studies [Kotlowski *et al.*, 2011; Nguyen *et al.*, 2023], particularly in binary classification problems. In our research, we analyzed the AUC results obtained by the two decision tree-based algorithms used (DT and RF). Evaluations were conducted on datasets from the car racing and flight simulation games, considering both scenarios without biosignal inclusion and those incorporating different combinations of physiological signals (ACC, EDA, and ECG).

Our focus was to identify the highest AUC scores across all tested combinations. As presented in Table 6, among the five combined scenarios analyzed (Games vs. Algorithms), four demonstrated superior performance when biosignals were included, highlighting the relevance of these data for model improvement.

Table 6. Best AUC Scores.

Attributes	Total	Game	Algorithm	AUC
Game, Profile, ACC, EDA, ECG	**13**	**Flight**	**RF**	**0.95**
Game and Profile	10	Flight	DT	0.85
Game, Profile, EDA, and ECG	12	Car	RF	0.71
Game, Profile, ACC, and EDA	12	Car	DT	0.66
Only ACC, EDA, and ECG	3	Flight	DT	0.66

Table 7 highlights that our model achieved solid performance compared to other studies that also used biosignals for CS prediction. In our approach, the combination of game data, user profile, and biosignals, totaling 13 attributes, resulted in an AUC of 0.95. This performance is comparable to the reference study [Porcino, 2021; Porcino *et al.*, 2022], which achieved the same AUC value using a smaller number of attributes, only 8 to predict CS.

4.7 Cause Identification

The choice of DT-based algorithms is important due to their ability to provide a clear and transparent interpretation of the decision-making process. Decision Trees are widely used in supervised learning, especially for classification tasks [Zhao

Table 7. Comparison of the best model analyzed in our work with previous approaches.

Reference	Task	Attributes	Model	AUC
[Keshavarz *et al.*, 2022]	VIMS	ECG, EDA, EGG, Respiration, Body Temperature, Skin Temperature, and Body Movements	RF	0.75
[Recenti *et al.*, 2021]	MS	EEG, EMG, HR, MSQ Questionnaire	RF	0.80
[Porcino, 2021; Porcino *et al.*, 2022]	CS	Game Data and User Profile	RF	0.95
Our Study	**CS**	**Game Data, User Profile, ACC, EDA, and ECG**	**RF**	**0.95**

and Zhang, 2008]. Their structure is inspired by a common tree, consisting of a root, nodes (splitting points), branches, and leaves.

In a DT, nodes are represented by circles, while branches correspond to the connections between these nodes [Ali *et al.*, 2012]. Each internal node reflects a decision based on a specific attribute, and leaf nodes represent the final outcomes or classes. During the process, the decision flow moves from one node to another based on the evaluated attribute values, repeating until it reaches a leaf node, where the final classification is determined based on the attributes analyzed along the path.

4.8 Ranking of Discomfort Causes

The ranking of the most relevant attributes was created based on the $Gini$ value, calculated from the decision path followed by each instance for a specific user. This value plays a fundamental role in determining the importance of an attribute, as it directly influences node splits during the construction of the decision tree. The implementation was carried out in Python using the *scikit-learn* library. Below, we describe the steps of the process that led to the final list of the most influential attributes, ordered in descending order according to their contribution to user discomfort.

Initially, for each experimental instance of a specific user, we checked whether the discomfort prediction made by the RF model was positive, meaning that the discomfort level was equal to 1 (**DiscomfortLevel = 1**). Since the RF algorithm consists of an ensemble of decision trees, we also iterated over each individual tree (estimator) that composes the model. If the individual tree's prediction matched the final decision of the RF model, the $Gini$ values associated with each attribute were summed [Porcino, 2021]. This process can be better understood through the example illustrated in Figure 8.

With the accumulated $Gini$ values for all user instances (grouped by attribute), we calculated the percentage influence of each attribute relative to the total identified. This overall process is represented by Equation 4, which summarizes how the attributes are ranked based on their relevance in the model.

$$\text{PDI}_u = \downarrow \left(\frac{\sum_i \text{Gini}(\text{Path}_{ij}^u)}{\text{TotalGini}} \right) \times 100 \qquad (4)$$

$\sum_i \text{Gini}(\text{Path}_{ij}^u)$: For each attribute (j) and instance (i) associated with user u, the total sum of *Gini* values is calculated and grouped.

$TotalGini$: Total sum of $Gini$ values for all attributes and all instances.

$\downarrow$: Indicates that the output will be sorted in descending order, starting with the attributes with the highest $Gini$ values.

Let us take as an example two instances represented in Figure 8, which show only the relevant nodes for the decision path of a specific user. The numerator value in Equation 4 for these instances is given by:

$$\sum_i \text{Gini}(\text{Path}_{ij}^u) = (\textbf{\textit{Instance1}}(Gender[0.432] + TimeStamp[0.415] + EDA[0.671] + Age[0.363] + ECG[0.328]) + \textbf{\textit{Instance2}}(TimeStamp[0.822] + Gender[0.482] + Experience[0.440] + Age[0.381] + Speed[0.284]))$$

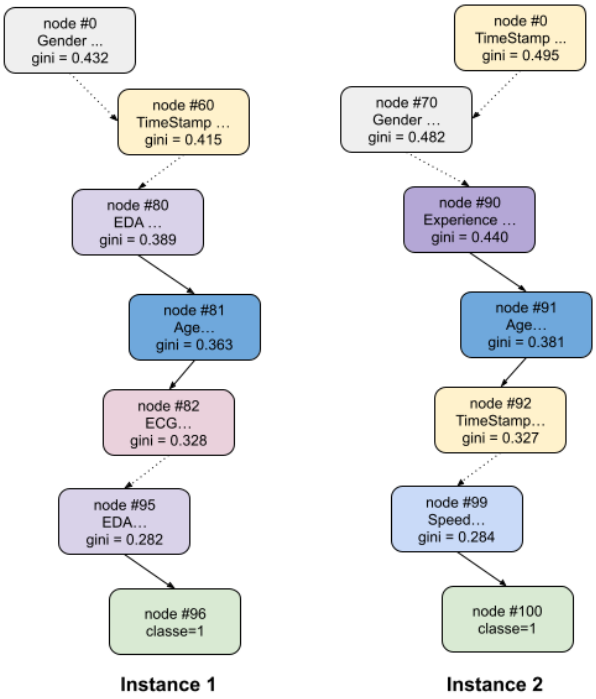


Figure 8. (Example) Gini values for the decision path, for the instances: (1) [0, 60, 80, 81, 82, 95, 96] e (2) [0, 70, 90, 91, 92, 99, 100].

For instance 1, it is observed that the EDA attribute appears twice along the decision path, being grouped and totaling a value of 0.671. Having the highest $Gini$ value indicates that EDA contributed significantly to the overall impurity reduction in the dataset, making it the most important attribute for this decision.

In instance 2, the TimeStamp attribute also appears twice, with its values summed, resulting in a total of 0.822. This result highlights TimeStamp as the most relevant attribute for this instance. The consolidation of grouped values for both instances is presented in Table 8.

After calculating the proportion of each attribute relative to the total sum of all attributes and instances (denominator of Equation 4) and organizing them in descending order, we arrive at the final result: the ranking of the most important attributes, presented in Table 9.

Table 8. Sum of $Gini$ values, grouped by attribute (Example from Figure 8).

Attribute	Gini Value (Total)
Gender	0.914
TimeStamp	1.237
EDA	0.671
Age	0.744
ECG	0.328
Experience	0.440
Speed	0.284

Table 9. Final result of Equation 4, ranking of the most relevant attributes in order of importance (Example from Figure 8).

Attribute	Gini Value (Total)	Percentage
TimeStamp	1.237	26.79%
Gender	0.914	19.79%
Age	0.744	16.11%
EDA	0.671	14.53%
Experience	0.440	9.53%
ECG	0.328	7.10%
Speed	0.284	6.15%

5 Results

In this section, we present a detailed analysis of the results obtained in each evaluated scenario, addressing key aspects of the research. Initially, we explore the VRSQ scores to assess the levels of discomfort reported by participants, considering differences between genders and specific symptom categories, such as Oculomotor and Disorientation. Next, we examine user reports and biosignals collected during the experiments in the Car Racing and Flying games, aiming to identify patterns associated with CS. Finally, we discuss the most relevant factors for the occurrence of CS, with and without the use of biosignals, establishing rankings based on the analyzed attributes.

5.1 VRSQ Scores

The data presented in Table 10 indicate that male users reported greater discomfort compared to female users in the Car Racing Game. On the other hand, female participants obtained higher overall VRSQ scores during the Flying Game, suggesting a higher incidence of CS symptoms.

Table 10. VRSQ Analysis.

Gender	Oculomotor		Disorientation		VRSQ Score	
	Car	Flight	Car	Flight	Car	Flight
F	27.78	47.22	35.56	46.67	31.67	46.94
M	41.67	16.67	33.33	13.34	37.50	15.00

5.2 User Discomfort Levels

We analyzed the levels of discomfort reported by participants during immersion in the virtual environment, emphasizing gender differences and the overall distribution of CS symptoms.

5.2.1 Self-Reports for the Car Racing Game

As presented in Table 11, male participants reported a higher incidence of CS symptoms compared to female participants. These results align with the analysis of mean VRSQ scores by gender from the previous section. Our findings also corroborate those of Porcino [2021], whose study also indicated that male participants reported greater discomfort than female participants during the racing game. Our results contradict the literature, which suggests that women are more susceptible to experiencing CS symptoms compared to men, as they do not possess the same spatial perception [Kolasinski, 1995; LaViola Jr, 2000; Mourant and Thattacherry, 2000].

Table 11. Total and percentage of CS levels reported by gender (Car Racing Game).

Gender	CS Level	Total Reported	%
Female	Mild	171	54.98
	Moderate	76	24.44
	Severe	64	20.58
	Total	**311**	
Male	Mild	218	62.46
	Moderate	102	29.23
	Severe	29	8.31
	Total	**349**	

With a significance level of 0.05, the Fisher's test applied to the total reports (Table 11) yielded a **p-value < 0.0001**, leading to the rejection of the null hypothesis. Thus, we conclude that, in this scenario, there is a significant association between gender and discomfort level.

We analyzed the overall distribution of user self-reports during the Car Racing Game, including the periods in which they reported no symptoms. For 53.2% of the experience time, users did not report discomfort. On the other hand, 27.6% of the time was associated with $mild$ symptoms, while 12.6% and 6.6% corresponded to $moderate$ and $severe$, respectively.

5.2.2 Self-Reports for the Flying Game

We also analyzed the experiments related to the Flying Game. According to Table 12, most participants of both genders reported mild CS symptoms, with no male participants reporting symptoms at the most severe level. However, the analysis was hindered by an insufficient and imbalanced number of participants with valid data per gender, making it impossible to reach a conclusive assessment regarding which gender exhibited a higher incidence of CS. For this reason, a more detailed analysis was conducted only for the Car Racing Game.

Table 12. CS levels reported by gender (Flying Game).

Gender	CS Level	Total Reported	%
Female	Mild	296	65.78
	Moderate	97	21.56
	Severe	57	12.67
Male	Mild	596	89.22
	Moderate	72	10.78
	Severe	0	0.00

We also applied Fisher's exact test to the totals presented in Table 12, considering a significance level of 0.05. The test resulted in a **p-value < 0.0001**, providing strong evidence to reject the null hypothesis. Thus, consistent with the findings from the Car Game, the results for the Flight Game indicate

that gender has a significant influence on the level of discomfort reported by participants.

Finally, we examined the overall proportion of user self-reports for this game, including the periods in which they reported no symptoms. Most users experienced some degree of CS during the experiment. The $Mild$ level was the most common, accounting for 59.1% of cases, followed by $Moderate$ and $Severe$ levels, with 11.2% and 3.8%, respectively. Only 25.9% of the gameplay time was associated with the absence of discomfort.

5.3 Physiological Changes (Rest vs. CS)

Our objective now is to analyze the physiological changes of users during game immersion, comparing them with baseline data. It is important to highlight that the analyses were conducted by comparing resting data with those recorded during the game, particularly in cases where users reported some level of discomfort. For this purpose, the means and standard deviations were calculated, and the *Wilcoxon signed-rank test* was applied to paired samples. This test, widely recognized in the scientific literature, is an essential statistical tool for assessing whether there is a significant difference between two sets of related measurements, without assuming data normality, and is among the most widely used techniques in scientific research [Wilcoxon, 1945].

5.3.1 Physiological Changes in the Car Racing Game

In the case of the Car Racing Game, as indicated in Table 13 and Figure 9, the results showed an average increase in players' body movements (Z-axis) during the game, especially in the presence of CS symptoms, compared to baseline data. This increase can be attributed to the "Postural Instability Theory" [LaViola Jr, 2000], along with the movements made by users to keep the car on track during the game. Additionally, a significant average increase in EDA was observed during the game, corroborating the study by Islam *et al*. [2020], which identifies EDA as one of the main indicators of CS.

Regarding the ECG, specifically RR intervals, there was an average reduction compared to resting data. These intervals represent the time between two consecutive R waves, corresponding to the interval between two heartbeats. Consequently, during CS symptoms, players exhibited a higher heart rate, confirming the findings of the study by Tian and Boulic [2024].

The *Wilcoxon* tests conducted for the Car Game (Table 13) revealed statistically significant differences between the measurements of the two analyzed conditions (baseline and with CS symptoms). This significance is confirmed by the extremely low p-values (< 0.001). Furthermore, considering the 95% confidence interval (CI), none of the intervals for the analyzed biosignals include the value zero, which allows for the rejection of the null hypothesis of no difference between the paired measurements. Thus, the analysis provides consistent evidence of a relationship between the presence of CS symptoms and the observed changes in physiological signals.

5.3.2 Physiological Changes in the Flight Game

The results from the Flying Game, presented in Table 14 and Figure 10, indicate significant changes in all analyzed biosignals. However, the body movement data captured by the ACC show a distinct pattern compared to the same biosignal recorded in the Car Racing Game. In the Flying Game, participants' body movements, especially during the baseline phase, were considerably more restrained, possibly reflecting a more disciplined behavior. This pattern may also be attributed to the calmer and less intense nature of the Flying Game, in contrast to the dynamic and fast-paced nature of the Car Racing Game. Additionally, the considerably high standard deviations, both in the baseline phase and during the experiment, are related to possible abrupt movements made by participants and were therefore not treated as *outliers*, remaining included in the analysis.

Regarding the EDA biosignal, an average increase was also identified during the game activity compared to baseline data, following the same explanatory rationale observed in the Car Racing Game. Similarly, the ECG, particularly the RR intervals, showed an average reduction compared to resting data, indicating an increase in heart rate during the game, with a justification similar to that applied in the previous game.

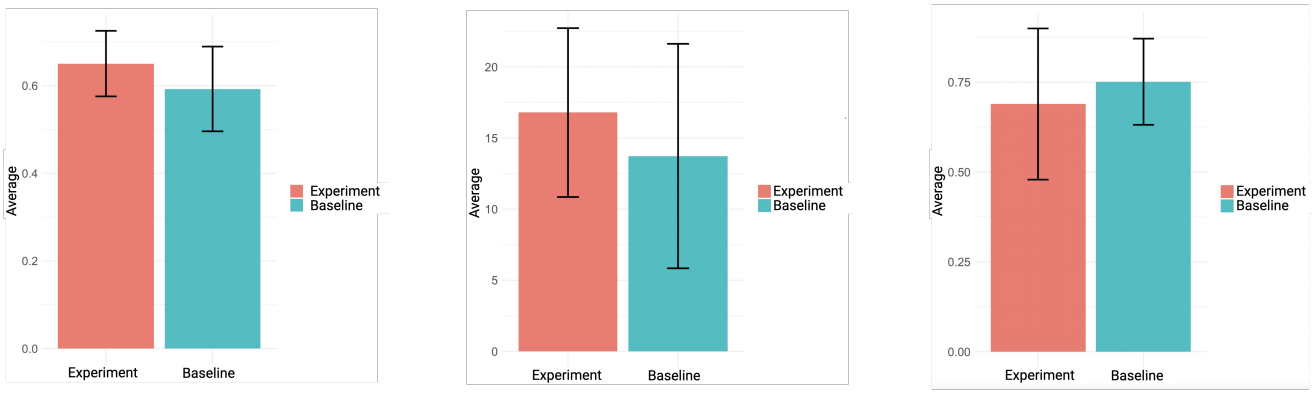
The analysis of the Flight Game data (Table 14), using the *Wilcoxon* test, indicated statistically significant differences between the baseline condition and the condition with CS symptoms. This evidence is supported by the obtained p-values, all of which were very low, some below 0.001. It was also observed that, within the 95% CI, none of the estimated intervals for the analyzed biosignals included the value zero, which supports the rejection of the null hypothesis. Taken together, these results strengthen the association between the occurrence of CS symptoms and the recorded physiological changes, highlighting the relevance of these parameters for understanding and detecting the phenomenon.

5.4 Rankings

CS is influenced by individual factors such as gender, age, and health, as well as software-related elements such as flicker, lag, acceleration, and rotation, making the identification of its causes a complex task. Although previous studies [Garcia-Agundez *et al*., 2019; Krokos and Varshney, 2022; Qu *et al*., 2022; Islam *et al*., 2020] have explored biosignals, they did not use them to classify the most relevant factors of CS. Porcino [2021] investigated eight factors associated with CS symptoms; however, physiological signals were not considered. Below, we present the results of each scenario analyzed in our study.

5.4.1 Results of Causes Without Biosignals

The averages of the main analyzed factors are presented in Table 15, in the **PDI** columns, and in Figure 11, organized by order of importance. In both games used in the experiments, Exposure Time (*TimeStamp*) was identified as the most relevant cause of discomfort, corroborating the findings of Porcino [2021], as highlighted in the **PCSRef** columns of Table 15. Other factors, such as Speed, Z Rotation, Frame Rate, and Gender, were also significant in both studies, although differences exist in the order of relevance and the percentage values assigned to each attribute. According to Fomenko and Kaewpankan [2022], rotations in VR environments can increase the likelihood of sensory conflicts, a finding consistent with other studies in the literature [Fomenko and Kaewpankan, 2022; Bonato *et al*., 2009], reinforcing the importance of these factors in the manifestation of CS symptoms.

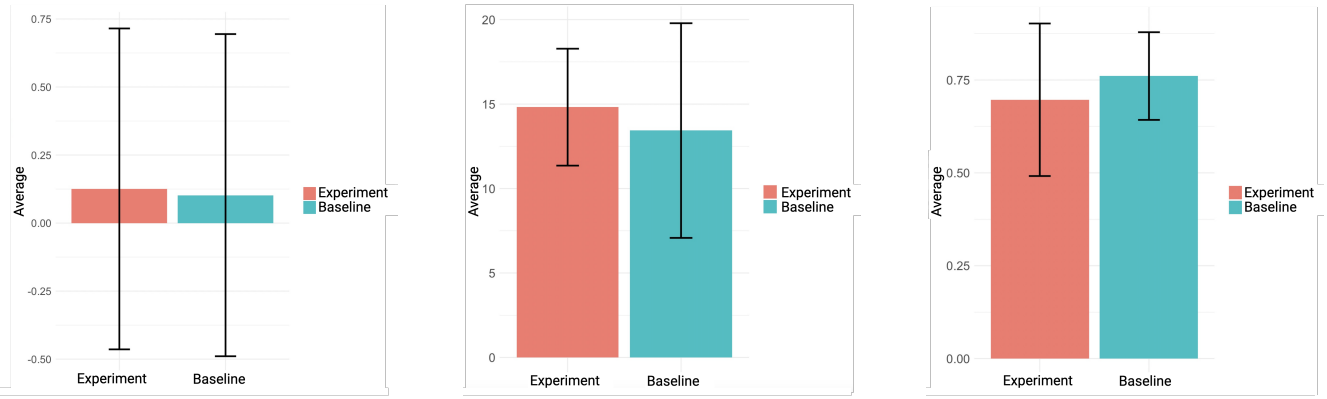


(a) ACC. **(b)** EDA. **(c)** ECG.

Figure 9. Comparison of Means and Standard Deviations for ACC, EDA, and ECG Biosignals in the Car Racing Game.

Table 13. Means, standard deviations, and Wilcoxon test results for the biosignals (ACC, EDA, and ECG) recorded during the Car Game, comparing baseline conditions and the occurrence of CS symptoms.

Biosignal	Means and Standard Deviations		V Statistic	p-value	95% CI	Median Diff.
	Baseline	With CS				
ACC	0.593 (0.097)	0.650 (0.075)	24.936	< 0.001	0.0600 to 0.0750	0.0700
EDA	13.731 (7.876)	16.789 (5.926)	22.190	< 0.001	3.2250 to 4.9100	4.6500
ECG (RR Interval)	0.751 (0.120)	0.689 (0.210)	10.041	< 0.001	-0.0749 to -0.0300	-0.0500



(a) ACC. **(b)** EDA. **(c)** ECG.

Figure 10. Comparison of Means and Standard Deviations for ACC, EDA, and ECG Biosignals in the Flying Game.

Table 14. Means, standard deviations, and Wilcoxon test results for the biosignals (ACC, EDA, and ECG) recorded during the Flight Game, comparing baseline conditions and the occurrence of CS symptoms.

Biosignal	Means and Standard Deviations		V Statistic	p-value	95% CI	Median Diff.
	Baseline	With CS				
ACC	0.103 (0.591)	0.126 (0.589)	52.112	0.031	0.0049 to 0.0399	0.0300
EDA	13.433 (6.351)	14.819 (3.458)	95.886	< 0.001	1.3500 to 2.5950	2.0650
ECG (RR Interval)	0.761 (0.118)	0.697 (0.205)	42.942	< 0.001	-0.0600 to -0.0350	-0.0449

5.4.2 *Results of Causes With Biosignals*

According to our experiments, the averages of the main factors contributing to the identification of CS are presented in Table 16, in the **PDI** columns, with biosignal-related attributes highlighted in bold, and in Figure 12. In the Car Racing Game, player movements (in the Z-axis), captured by an ACC, had the greatest impact on identifying CS. This finding can be explained by the "Postural Instability Theory", described by LaViola LaViola Jr [2000]. Exposure Time (TimeStamp) was identified as the second most relevant attribute in triggering symptoms, corroborating the findings of Porcino [2021]; Porcino *et al.* [2022].

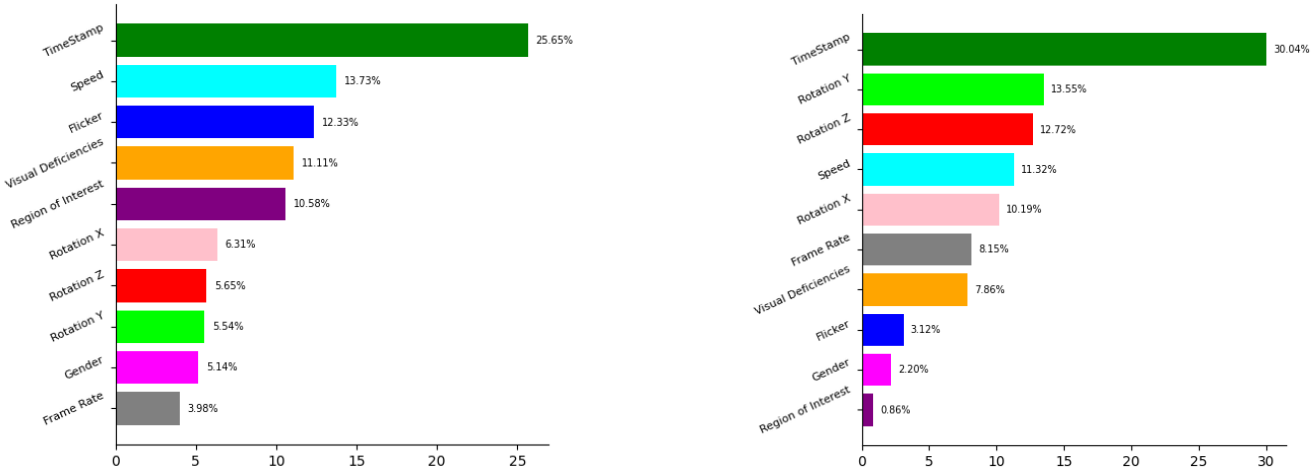
EDA ranked third, aligning with the works of Islam *et al.* [2020] and Jung *et al.* [2021], which also demonstrated a strong correlation with CS. Additionally, RR intervals (extracted from ECG) were also found to be significant factors for CS symptoms. The results obtained using ECG in our study are consistent with other studies described in the literature Islam *et al.* [2020]; Garcia-Agundez *et al.* [2019]; Qu *et al.* [2022].

In the Flying Game, Exposure Time (TimeStamp) once again stood out as the most relevant factor, followed by EDA.

Table 15. Ranking of the average percentages of the main characteristics causing CS (excluding biosignals), compared to the reference study Porcino [2021]. **PDI** - method used in our study, **PCS** - method proposed by Porcino [2021]. Where no value is provided, no comparison was made.

Car Game				**Flight Game**			
Attribute	**PDI ↓**	**PCS**	**PCSRef**	**Attribute**	**PDI ↓**	**PCS**	**PCSRef**
TimeStamp	25.65	28.59	27.32	TimeStamp	30.04	32.49	22.99
Speed	13.73	13.49	12.73	Rotation Y	13.55	13.82	-
Flicker	12.33	12.00	-	Rotation Z	12.72	14.93	18.64
Visual Deficiencies	11.11	12.49	-	Speed	11.32	10.97	12.38
Region of Interest	10.58	11.16	-	Rotation X	10.19	10.34	-
Rotation X	6.31	5.49	-	Frame Rate	8.15	2.80	5.43
Rotation Z	5.65	4.55	13.84	Visual Deficiencies	7.86	7.66	-
Rotation Y	5.54	4.86	-	Flicker	3.12	4.04	-
Gender	5.14	5.30	12.52	Gender	2.20	2.15	6.18
Frame Rate	3.98	2.07	11.92	Region of Interest	0.86	0.80	-
VR Experience	-	-	8.26	*VR Experience*	-	-	4.76
Age	-	-	7.77	*Age*	-	-	5.33
Acceleration	-	-	5.64	*Acceleration*	-	-	11.76

PCSRef - results obtained from Porcino [2021] for comparison purposes.



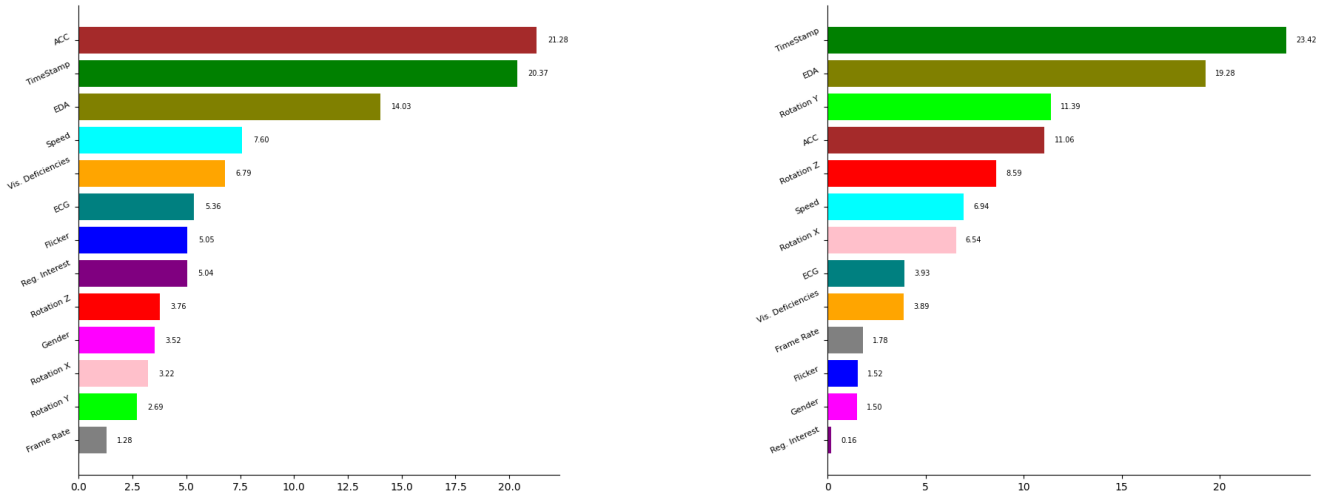
(a) Car Racing Game **(b)** Flying Game

Figure 11. Rankings of the average percentages of the main factors used to identify CS, using the **PDI** method, in the Car Racing (a) and Flying (b) games, **excluding** biosignals.

Table 16. Ranking of the average values of the main CS-indicative factors, **including** biosignals.

Car Game			**Flight Game**		
Attribute	**PDI ↓**	**PCS**	**Attribute**	**PDI ↓**	**PCS**
ACC	21.28	22.53	TimeStamp	23.42	24.83
TimeStamp	20.37	21.46	**EDA**	19.28	16.48
EDA	14.03	12.41	Rotation Y	11.39	11.02
Speed	7.60	7.58	**ACC**	11.06	11.34
Visual Deficiencies	6.79	7.64	Rotation Z	8.59	9.48
ECG	5.36	4.74	Speed	6.94	6.78
Flicker	5.05	5.22	Rotation X	6.54	5.86
Region of Interest	5.04	5.07	**ECG**	3.93	5.06
Rotation Z	3.76	3.14	Visual Deficiencies	3.89	3.77
Gender	3.52	3.86	Frame Rate	1.78	1.75
Rotation X	3.22	2.87	Flicker	1.52	2.17
Rotation Y	2.69	2.34	Gender	1.50	1.27
Frame Rate	1.28	1.15	Region of Interest	0.16	0.19

Y Rotation ranked third, while player movements captured by the ACC appeared in fourth place. Additionally, RR intervals



(a) Car Game **(b)** Flying Game

Figure 12. Rankings of the average percentages of the main factors used to identify CS, using the **PDI** method, in the Car Racing (a) and Flying (b) games, **including** biosignals.

were also identified as a significant factor in predicting CS. The underlying principles explaining and relating these attributes are the same as those supporting the results obtained in the experiment with the Car Racing Game.

5.4.3 Comparison Between PDI and PCS Results

We also conducted the identification of the main characteristics that trigger CS using a combination of the *Random Forest* algorithm with the method proposed by Porcino [2021]. The results presented in Tables 15 and 16, in the **PCS** and **PDI** columns, were evaluated using the Wilcoxon test for both the Car Racing and Flying Game. The obtained *p-values* (**p-value > 0.05**) indicated no statistically significant differences between the approaches, suggesting that the results derived from the *Gini* measures and the method by Porcino [2021] are highly consistent.

Since we aimed to demonstrate that there is no significant difference between the two methods, the null hypothesis became the focus of our interest. It highlights that the results derived from the previously calculated *Gini* values (using the scikit-learn library) are similar to those obtained through the method proposed by Porcino [2021].

6 Limitations

This study presents some limitations, such as a limited number of participants and the exclusion of data due to Bluetooth connectivity issues between the BITalino board and the computer running the OpenSignals software. Additionally, difficulties in electrode placement impacted data collection. Moreover, the sensor used to capture player movements during the experience was limited to measuring acceleration along the Z-axis (forward and backward movements), excluding the analysis of movements in other directions (X and Y).

Furthermore, the need to generate a unified dataset from two distinct sources (game and biosignals) introduced challenges in ensuring consistent data synchronization. In some cases, there was a delay in the start of biosignal capture relative to the player's experience. To mitigate this issue, instances of the game that were not aligned with the biosignals due to time discrepancies between the two datasets were discarded.

Finally, the study still has limitations that include the need to deepen the investigation with other demographic groups (particularly across different age ranges) and to increase the number of participants. It is also essential to explore the potential relationship between individuals who experience Motion Sickness, triggered by real movements, and CS, induced by exposure to VR environments.

7 Conclusion and Future Work

The study of CS has garnered increasing interest, particularly regarding methods for identifying and assessing its intensity in VR environments. Despite this, there is a clear gap in studies that integrate biosignals, user profile data, and game characteristics to identify the factors contributing to CS. Moreover, the use of SML to interpret model decisions remains underexplored. In this context, the present study investigated how biosignals can be utilized to identify potential causes associated with CS in VR games. The results from the VRSQ questionnaire indicated that, in the car racing game, men reported greater discomfort than women, whereas, in the Flying Game, women exhibited higher discomfort scores. The *Wilcoxon* test confirmed significant differences between biosignal data in the resting state and during CS symptoms.

The use of the *Random Forest* algorithm enabled the interpretation of model decisions and the creation of a ranking of the main factors associated with CS. By combining game data, user profiles, and biosignals, the model achieved the best performance, emphasizing the importance of physiological signals in predicting CS. In the car racing game, the most relevant factors were movements detected by the ACC, exposure time, EDA, and RR intervals from the ECG, while in the Flying Game, the key attributes were exposure time, EDA, and Y-axis rotation. These results highlight the relevance of integrating different data sources to identify critical factors related to CS.

Finally, we intend to explore the relationship between Motion Sickness (MS), caused by real-world movement, and CS, induced by VR environments. Additionally, we plan to

investigate how sensors integrated into modern HMD devices, which allow eye and head tracking, can be used to mitigate CS symptoms through dynamic feedback. Identifying the main causative or indicative factors of CS is essential for VR technology developers to implement more effective mitigation strategies, contributing to more enjoyable experiences and positively impacting users' lives.

Declarations

Authors' Contributions

Wedrey, Thiago, Carla, and Ricardo contributed to the conceptualization and methodology of the study. Wedrey Nunes was responsible for data collection and curation, as well as the development and implementation of the model. The writing and revision of the manuscript were carried out collectively by the authors. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare that they have no financial, commercial, or any other conflicts of interest that could influence the results or interpretation of this study.

Availability of data and materials

The datasets collected and analyzed during this study are available at: `https://drive.google.com/drive/folders/1tF0PkSkSJb91aIGsLQSruBxPvyY-8JAL`.

References

Alberdi, A., Aztiria, A., and Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of biomedical informatics*, 59:49–75. DOI: https://doi.org/10.1016/j.jbi.2015.11.007.

Ali, J., Khan, R., Ahmad, N., and Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, 9(5):272.

Andrejašic, M. (2008). Mems accelerometers. In *University of Ljubljana. Faculty for mathematics and physics, Department of physics, Seminar*, volume 49.

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S., *et al*. (2012). The'k'in k-fold cross validation. In *ESANN*, pages 441–446.

Awal, M. A., Mostafa, S. S., Ahmad, M., and Rashid, M. A. (2014). An adaptive level dependent wavelet thresholding for ecg denoising. *Biocybernetics and biomedical engineering*, 34(4):238–249. DOI: https://doi.org/10.1016/j.bbe.2014.03.002.

Bassett, D. R. and John, D. (2010). Use of pedometers and accelerometers in clinical populations: validity and reliability issues. *Physical therapy reviews*, 15(3):135–142. DOI: https://doi.org/10.1179/1743288X10Y.0000000004.

Bernardini, F. C. (2006). *Combinação de classificadores simbólicos utilizando medidas de regras de conhecimento e algoritmos genéticos*. PhD thesis, Universidade de São Paulo.

Bonato, F., Bubka, A., and Palmisano, S. (2009). Combined pitch and roll and cybersickness in a virtual environment. *Aviation, space, and environmental medicine*, 80(11):941–945. DOI: https://doi.org/10.3357/ASEM.2394.2009.

Bouyer, G., Chellali, A., and Lécuyer, A. (2017). Inducing self-motion sensations in driving simulators using force-feedback and haptic motion. In *2017 IEEE Virtual Reality (VR)*, pages 84–90. IEEE. DOI: https://doi.org/10.1109/VR.2017.7892234.

Budhiraja, P., Miller, M. R., Modi, A. K., and Forsyth, D. (2017). Rotation blurring: use of artificial blurring to reduce cybersickness in virtual reality first person shooters. *arXiv preprint arXiv:1710.02599*. DOI: https://doi.org/10.48550/arXiv.1710.02599.

Caserman, P., Garcia-Agundez, A., Gámez Zerban, A., and Göbel, S. (2021). Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. *Virtual Reality*, 25(4):1153–1170. DOI: https://doi.org/10.1007/s10055-021-00513-6.

Daniya, T., Geetha, M., and Kumar, K. S. (2020). Classification and regression trees with gini index. *Advances in Mathematics: Scientific Journal*, 9(10):8237–8247. DOI: https://doi.org/10.37418/amsj.9.10.53.

De Miras, J. R., Ibáñez-Molina, A. J., Soriano, M. F., and Iglesias-Parro, S. (2023). Schizophrenia classification using machine learning on resting state eeg signal. *Biomedical Signal Processing and Control*, 79:104233. DOI: https://doi.org/10.1016/j.bspc.2022.104233.

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252. DOI: https://doi.org/10.1080/00401706.1964.10490181.

Fomenko, I. and Kaewpankan, T. (2022). Thrill vs. cybersickness: A study on camera settings' impact on immersion and cybersickness in vr racing games.

Garcia-Agundez, A., Reuter, C., Caserman, P., Konrad, R., and Göbel, S. (2019). Identifying cybersickness through heart rate variability alterations. *International Journal of Virtual Reality*, 19(1):1–10. DOI: https://doi.org/10.20870/IJVR.2019.19.1.2907.

Islam, R., Lee, Y., Jaloli, M., Muhammad, I., Zhu, D., Rad, P., Huang, Y., and Quarles, J. (2020). Automatic detection and prediction of cybersickness severity using deep neural networks from user's physiological signals. In *2020 IEEE international symposium on mixed and augmented reality (ISMAR)*, pages 400–411. IEEE. DOI: https://doi.org/10.1109/ISMAR50242.2020.00066.

Jeong, D. and Han, K. (2024). Precyse: Predicting cybersickness using transformer for multimodal time-series sensor data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(2):1–24. DOI: https://doi.org/10.1145/3659594.

Jung, S., Li, R., McKee, R., Whitton, M. C., and Lindeman, R. W. (2021). Floor-vibration vr: mitigating cybersickness using whole-body tactile stimuli in highly realistic vehicle driving experiences. *IEEE Transactions on Visualization & Computer Graphics*, 27(05):2669–2680. DOI: https://doi.org/10.1109/TVCG.2021.3067773.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220. DOI: https://doi.org/10.1207/s15327108ijap0303_3.

Kennedy, R. S. and Lilienthal, M. G. (1995). Implications of balance disturbances following exposure to virtual re-

ality systems. In *Proceedings Virtual Reality Annual International Symposium'95*, pages 35–39. IEEE. DOI: https://doi.org/10.1109/VRAIS.1995.512477.

Keshavarz, B., Peck, K., Rezaei, S., and Taati, B. (2022). Detecting and predicting visually induced motion sickness with physiological measures in combination with machine learning techniques. *International Journal of Psychophysiology*, 176:14–26. DOI: https://doi.org/10.1016/j.ijpsycho.2022.03.006.

Kolasinski, E. M. (1995). Simulator sickness in virtual environments.

Kotlowski, W., Dembczynski, K. J., and Huellermeier, E. (2011). Bipartite ranking through minimization of univariate loss. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 1113–1120.

Kourtesis, P., Collina, S., Doumas, L. A., and MacPherson, S. E. (2019). Technological competence is a pre-condition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Frontiers in Human Neuroscience*, 13:342. DOI: https://doi.org/10.3389/fnhum.2019.00342.

Krokos, E. and Varshney, A. (2022). Quantifying vr cybersickness using eeg. *Virtual Reality*, 26(1):77–89. DOI: https://doi.org/10.1007/s10055-021-00517-2.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621. DOI: https://doi.org/10.1080/01621459.1952.10483441.

LaViola Jr, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin*, 32(1):47–56. DOI: https://doi.org/10.1145/333329.333344.

Monard, M. C. and Baranauskas, J. A. (2003). Indução de regras e árvores de decisão. In Rezende, S. O., editor, *Sistemas Inteligentes: Fundamentos e Aplicações*, pages 115–139. Editora Manole, Barueri, SP, Brasil.

Mourant, R. R. and Thattacherry, T. R. (2000). Simulator sickness in a virtual environments driving simulator. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 44, pages 534–537. SAGE Publications Sage CA: Los Angeles, CA. DOI: https://doi.org/10.1177/154193120004400513.

Nguyen, N. D., Tan, W., Du, L., Buntine, W., Beare, R., and Chen, C. (2023). Auc maximization for low-resource named entity recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13389–13399. DOI: https://doi.org/10.1609/aaai.v37i11.26571.

Nunes da Silva, W., Porcino, T. M., Castanho, C. D., and Jacobi, R. P. (2024). Analysis of cybersickness through biosignals: an approach with symbolic machine learning. In *Proceedings of the 26th Symposium on Virtual and Augmented Reality (SVR '24)*, pages 11–20, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3691573.3691582.

Oh, H. and Son, W. (2022). Cybersickness and its severity arising from virtual reality content: A comprehensive study. *Sensors*, 22(4):1314. DOI: https://doi.org/10.3390/s22041314.

Pashaei, E. and Pashaei, E. (2019). Gene selection using intelligent dynamic genetic algorithm and random forest. In *2019 11th international conference on electrical and electronics engineering (ELECO)*, pages 470–474. IEEE. DOI: https://doi.org/10.23919/ELECO47770.2019.8990557.

PLUX Biosignals (2023a). Electrocardiography (ecg) sensor data sheet. `https://support.pluxbiosignals.com/wp-content/uploads/2021/11/revolution-ecg-sensor-datasheet-revb-1.pdf`, Accessed: 24 October 2025.

PLUX Biosignals (2023b). Getting started: Bitalino electrodermal activity (eda) sensor. `https://support.pluxbiosignals.com/knowledge-base/getting-started-bitalino-electrodermal-activity-eda-sensor/`, Accessed: 24 October 2025.

PLUX Biosignals (2024). Bitalino (r)evolution lab guide. `https://support.pluxbiosignals.com/wp-content/uploads/2022/04/HomeGuide2_ECG.pdf`, Accessed: 24 October 2025.

Porcino, T., Rodrigues, E. O., Bernardini, F., Trevisan, D., and Clua, E. (2022). Identifying cybersickness causes in virtual reality games using symbolic machine learning algorithms. *Entertainment Computing*, 41:100473. DOI: https://doi.org/10.1016/j.entcom.2021.100473.

Porcino, T. M. (2021). *Cybersickness Analysis Using Symbolic Machine Learning Algorithms*. PhD thesis, Universidade Federal Fluminense - UFF, Niterói-RJ, Brasil.

Qu, C., Che, X., Ma, S., and Zhu, S. (2022). Biophysiological-signals-based vr cybersickness detection. *CCF Transactions on Pervasive Computing and Interaction*, 4(3):268–284. DOI: https://doi.org/10.1007/s42486-022-00103-8.

Reason, J. T. and Brand, J. J. (1975). *Motion sickness.* Academic press.

Recenti, M., Ricciardi, C., Aubonnet, R., Picone, I., Jacob, D., Svansson, H. Á., Agnarsdóttir, S., Karlsson, G. H., Baeringsdóttir, V., Petersen, H., *et al*. (2021). Toward predicting motion sickness using virtual reality and a moving platform assessing brain, muscles, and heart signals. *Frontiers in Bioengineering and Biotechnology*, 9:635661. DOI: https://doi.org/10.3389/fbioe.2021.635661.

Riccio, G. E. and Stoffregen, T. A. (1991). An ecological theory of motion sickness and postural instability. *Ecological psychology*, 3(3):195–240. DOI: https://doi.org/10.1207/s15326969eco0303_2.

Rokach, L. and Maimon, O. (2005). Decision trees. *Data mining and knowledge discovery handbook*, pages 165–192. DOI: https://doi.org/10.1007/0-387-25465-X_9.

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A., Rizzo, A., and Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. *Frontiers in human neuroscience*, 14:96. DOI: https://doi.org/10.3389/fnhum.2020.00096.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423. DOI: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

SHAPIRO, S. (1965). An analysis of variance test for normality (complete samples) t. *Biometrika*, 52(3/4):591–611. DOI: https://doi.org/10.1093/biomet/52.3-4.591.

Stanney, K. M. and Hash, P. (1998). Locus of user-initiated control in virtual environments: Influences

on cybersickness. *Presence*, 7(5):447–459. DOI: https://doi.org/10.1162/105474698565848.

Tan, J., Liu, H., Li, M., and Wang, J. (2018). A prediction scheme of tropical cyclone frequency based on lasso and random forest. *Theoretical and applied climatology*, 133:973–983. DOI: https://doi.org/10.1007/s00704-017-2233-3.

Tian, N. and Boulic, R. (2024). Who says you are so sick? an investigation on individual susceptibility to cybersickness triggers using eeg, egg and ecg. *IEEE Transactions on Visualization and Computer Graphics*. DOI: https://doi.org/10.1109/TVCG.2024.3372066.

Valkonen, M., Kartasalo, K., Liimatainen, K., Nykter, M., Latonen, L., and Ruusuvuori, P. (2017). Metastasis detection from whole slide images using local features and random forests. *Cytometry Part A*, 91(6):555–565. DOI: https://doi.org/10.1002/cyto.a.23089.

Vince, J. (2004). *Introduction to Virtual Reality*. Springer London, London.

Wang, T., Jia, W., Li, F., Liu, X., Zhang, X., and Hu, F. (2025). Multi-dimensional feature extraction of eeg signal and its application in stroke classification. *Scientific Reports*, 15(1):19589. DOI: https://doi.org/10.1038/s41598-025-04756-0.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83. DOI: https://doi.org/10.2307/3001968.

Yao, D., Zhan, X., and Kwoh, C.-K. (2019). An improved random forest-based computational model for predicting novel mirna-disease associations. *BMC bioinformatics*, 20:1–14. DOI: https://doi.org/10.1186/s12859-019-3290-7.

Young, S. D., Adelstein, B. D., and Ellis, S. R. (2007). Demand characteristics in assessing motion sickness in a virtual environment: Or does taking a motion sickness questionnaire make you sick? *IEEE transactions on visualization and computer graphics*, 13(3):422–428. DOI: https://doi.org/10.1109/TVCG.2007.1029.

Zhao, Y. and Zhang, Y. (2008). Comparison of decision tree methods for finding active objects. *Advances in Space Research*, 41(12):1955–1959. DOI: https://doi.org/10.1016/j.asr.2007.07.020.