Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

# Cybersickness Risk Prediction Using EEG and Symbolic Explainable Machine Learning

**Yalamati Nikhil Souri[1], Dr. Y. Sobhan Babu[2], Dr. D. Venkatesh[3]**

[1] Student, Dept. of Computer Science, Blekinge Institute of Technology, Sweden.
[2] Professor, Dept. of Computer Science, PVKK Institute of Technology, Anantapur, India.
[3] HOD & Professor, Dept. of Computer Science, PVKK Institute of Technology, Anantapur, India.

**Abstract**

Cybersickness is still a major problem in Virtual Reality applications, leading to discomfort for users and poor usability of the system. This work proposes a Neuro-Symbolic method that can predict the risk of cybersickness from brain activity data recorded through electroencephalogram (EEG) with symbolic machine learning models that are interpretable. The pipeline that is proposed includes several steps: ingestion of multi-dataset EEG signals, statistical feature extraction based on windows, individual subject validation, and model interpretability using the methods of SHAP, LIME, and rule-based reasoning. The features that were extracted include the power of different brain frequencies, the asymmetry of the brain hemispheres, the connectivity between brain regions, the complexity of the spectrum, and the dynamics in time. The performance of the models was assessed using publicly available VR-EEG data sets, where it was found that Random Forest, XGBoost, and LightGBM, which are tree-based models, achieved high discriminative performance while maintaining complete interpretability through human-readable decision rules. Moreover, the system outputs statistical cross-fold results that are validated, calibrated risk scores, global and local feature importance, and ranked symbolic rules to elucidate the causes of cybersickness risk. The results suggest that it is possible to predict the risk of EEG-based cybersickness reliably in real-time while providing transparent decision paths that are suitable for academic evaluation and practical deployment.

**Keywords:** Cybersickness Prediction, EEG Feature Engineering, Symbolic Machine Learning, Model Explainability, Virtual Reality Analytics

## 1. Introduction

Virtual Reality (VR) technology has become a revolutionary platform in various sectors like gaming, simulation, education, and healthcare because of its immersive and engaging nature [3]. Nevertheless, the rapid progress of VR and its universal acceptance has not completely removed the issue of cybersickness which usually accompanies VR experiences, and this is the most common reason users give up on it. This discomfort syndrome includes symptoms like nausea, dizziness, and disorientation, and Very often corroborates with oculomotor strain [3]. Besides, these negative effects not only lower user comfort but also put a limitation on the usability and accessibility of the VR systems which are particularly needed in applications like long interactions or tasks requiring high precision.

In traditional evaluations, cybersickness was measured primarily through the use of subjective questionnaires, among which the Simulator Sickness Questionnaire (SSQ), Fast Motion Sickness scale (FMS), and Motion Sickness Susceptibility Questionnaire (MSSQ) have remained as the standard methods for assessing VR-induced discomfort (Yang, 2023; Hua et al., 2024) [3, 5]. Notwithstanding, such methods only take into account the users' retrospective feedback and do not possess the capability to continuously monitor neurological or autonomic changes that take place during the VR session, thus limiting their application for real-time monitoring of the risk of sickness and preventive measures to be taken [3].

To address these constraints, contemporary studies have shifted toward objective biosignals-including heart rate variability, skin conductance, respiration, and EEG-to infer discomfort progression and cybersickness risk (Ripan et al., 2022; Islam et al., 2023) [1, 7]. EEG has emerged as a particularly effective sensing modality because of its high temporal resolution and its sensitivity to rhythmic imbalance patterns linked to sensory conflict accumulation, vestibular-visual motion integration strain, and cognitive disorientation during immersive interaction [2, 3, 7].

Various deep learning architectures, including CNNs and CNN-LSTM, have operated with strong performance in classifying cybersickness based on EEG connectivity and temporal features (Yildirim, 2020; Shen et al., 2024; Luo, 2024) [6, 8, 9]. On the contrary, these models are usually considered as black-box learners, creating trust issues in their use for immersive systems that are safety-critical, in which case CE is as critical as predictive performance [1, 7]. Tools such as LiteVR, which is a lightweight explainable framework, show that SHAP-based post-hoc explanations can only partially restore interpretability while still maintaining the quality of the classification [7]. Symbolic rule models nevertheless are still recognized as the best ones in producing readable and auditable by humans decision logic, mainly for the subject-independent validation approach scenarios [1, 7].

The technology has significantly improved, nonetheless, there is still no universal model that envelops multi-type EEG feature refinement, understandable symbolic rule extraction, explainability (SHAP, LIME), and thorough subject-wise validation. The present research fills this void by introducing an all-encompassing machine learning pipeline for predicting the risk of cybersickness from EEG signals that aligns high predictive accuracy with obvious explainability and statistical robustness, thereby supporting both scientific insight and practical application.

## 2. Related Research Work

The capability of EEG signals to detect and predict cybersickness has received considerable research attention in recent years as a result of the objective and high-resolution nature of the brain's electrical activity [3]. In the past, researchers were looking in this area mostly for variations of EEG frequency bands that were linked to motion sickness and using basic signal processing and classification techniques [5]. The work of Hua et al. (2024) resulted in a new multi-scale feature correlation model that allows the separation of the VR motion sickness states by uniting both spectral and temporal EEG features, thus proving the possibility of machine learning-driven classification in this context [5]. Likewise, Luo (2024) utilized wavelet packet feature extraction on EEG data and demonstrated that advanced feature representations can significantly increase the prediction performance of motion sickness

Traditional machine learning models gave good results, but deep learning architectures have been considered more and more for their capability of automatic feature extraction. Shen et al. (2024) made use of a combination of CNN and LSTM approach taking advantage of phase-locked value functional connectivity matrices to identify visually-induced motion sickness. They got very close to the best accuracy that can be achieved while at the same time pointing out how important it is to consider the temporal relations in EEG dynamics [6]. On the other hand, deep models are typically non-transparent and this lack of transparency makes it hard to interpret and trust them thus limiting their use in real applications.

The interpretability gap has prompted to the use of explainable machine learning techniques for the cybersickness detection and interpretation in several works. One of these works is Ripan et al. (2022) who created TruVR, an interpretable machine learning framework that employs transparent models to assess the likelihood of cybersickness and emphasizes the necessity of decision clarity for virtual reality systems that are vital for safety [1]. Another case is Kundu et al. (2023), who introduced LiteVR, a system that applies SHAP post-hoc explainability techniques to lightweight classifiers, thus achieving a good balance between interpretability and predictive power and making significant progress towards easily understood and explainable predictions of VR sickness [7].

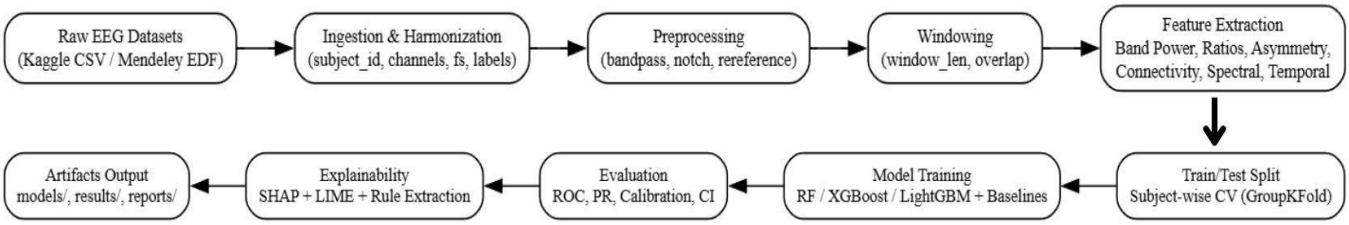
The authors of existing models have mainly concentrated on their performance level or the other way round when it comes to less explicable models. Among others, Berken et al. (2025) showed continuous cybersickness detection based on EEG via a multi-taper spectrum estimation technique to be a major improvement in the representation of the motion sickness force, but their approach lacked integrated explainability [2]. Together, they point to the necessity for the comprehensive EEG-based systems that deliver the prediction results with high accuracy and, at the same time, give insights into the neural processes associated with cybersickness that are easy to understand.

## 3. Methods

### 3.1 System Overview

This work presents a comprehensive pipeline for predicting cybersickness risk using EEG signals which is based on virtual reality EEG recordings. The pipeline not only produces neurophysiological features that are (i) engineered, (ii) a trained predictive model, and (iii) rationalized explanations for every prediction made but also includes a trained predictive model and rationalized explanations for every prediction made. The pipeline is reproducible and configurable, meaning that the same workflow can be applied to different datasets with varying channel counts, sampling rates, and label formats. Data processing is primarily done in sequential stages: dataset ingestion and harmonization, signal preprocessing, windowing, feature extraction, model training with validation, evaluation through robust metrics, and explainability using both global (population-level) and local (instance-level) explanations. Intermediate artifacts (clean signals, extracted features, trained models, plots, and explanation reports) are generated in each stage to facilitate debugging, auditing, and conducting experiments that can be repeated.

Figure 1: Architecture diagram




### 3.2 Data Sources and Dataset Harmonization

The pipeline supports multiple public VR/EEG datasets and is structured to handle the practical differences that typically occur across sources: varying electrode montages (e.g., 14/19/29 channels), different sampling frequencies, inconsistent naming conventions (Fp1 vs FP1), different file formats (CSV, EDF), and different labelling methods (binary discomfort class vs SSQ-derived severity). During ingestion, each dataset is mapped into a unified internal schema containing: subject identifier, session/trial identifier, raw multichannel EEG array, sampling rate, channel list, timestamp alignment (if available), and the associated cybersickness ground truth labels (per trial or per time segment). This harmonization step is crucial because it prevents later stages from becoming dataset-specific and enables cross-dataset evaluation by ensuring consistent feature extraction and labelling logic.

### 3.3 EEG Preprocessing and Signal Conditioning

The primary source of error that often comes with raw EEG signals includes eye blinks, muscle activity, and motion noises, among others, in addition to slow drifts that will bias frequency-domain and connectivity features. The preprocessing stage is meant to eliminate the negative impact of such artifacts using the standard conditioning steps of EEG which are then applied depending on the configurations set. The very first step involves the re-referencing of the signal (whenever it is appropriate) so as to minimize channel-specific bias and enhance the comparability of the data across the various subjects. A bandpass filter is then used to filter the EEG-relevant frequencies and hence, to eliminate DC drift and high-frequency noise. Depending on the recording environment, a notch filter may also be employed to reduce powerline interference (50/60 Hz). If the dataset provides quality markers or bad channel metadata, those channels can be excluded; otherwise, an automated rule-based screening can be used to flag channels with abnormal variance or saturation. The pipeline keeps preprocessing conservative to avoid over-cleaning, since aggressive artifact removal can sometimes remove informative neural components linked to discomfort and cognitive strain in VR tasks.

### 3.4 Windowing Strategy and Sample Construction

Cybersickness is not always an instantaneous phenomenon; it often develops over time due to sensory conflict accumulation. Therefore, instead of treating entire sessions as one sample, the pipeline converts continuous EEG recordings into fixed-length windows so the model can learn temporal patterns and provide time-resolved risk estimates. A window length (e.g., 2–10 seconds) and overlap (e.g., 50%) are configured to balance temporal resolution and feature stability. Overlap increases the number of training samples and smooths predictions over time, but it also increases the risk of leakage if windows from the same subject/session appear in both training and testing splits. For this reason, the pipeline is designed to support **subject-wise** splitting during cross-validation so that all windows from a subject stay

within a single fold. Each window becomes one training instance represented by a feature vector and a corresponding risk label.

### 3.5 Labeling and Ground Truth Construction

The pipeline is designed to work with both the categorical and score-based labels. In the case of datasets that offer subjective evaluations like SSQ, a thresholding approach can be used to change the scores into two binary classes like No Risk and At Risk. Alternatively, multi-class labels can be created (e.g., low/medium/high) if the dataset size is sufficient and class distribution is reasonable. When labels are available only at session-level, the label can be applied to all windows of that session, with careful reporting that classification is "window-level with session labels." For datasets with time-aligned discomfort annotations, labels can be mapped more precisely to windows based on annotation timestamps. Clear documentation of label logic is essential because labelling choices strongly influence performance and real-world meaning; therefore, the pipeline stores the exact thresholds and mapping rules in the experiment configuration and outputs them in the analysis report.

### 3.6 Feature Engineering

To make predictions interpretable and physiologically meaningful, the pipeline relies on feature engineering rather than end-to-end black-box learning. Each EEG window is transformed into a set of features that represent known neural correlates of arousal, attention, fatigue, and sensory integration. Feature groups are modular and can be switched on/off in the configuration for ablation experiments.

**Band Power Features:** For each channel, the power spectral density (PSD) is calculated, and then band powers are determined for the delta, theta, alpha, beta, and gamma ranges. Band powers denote a succinct depiction of brain waves, and the fluctuations in theta/alpha or beta activity are most of the time associated with factors such as workload, discomfort, and changed sensory processing during immersion.

**Asymmetry Indices:** The assessment of hemispheric asymmetry is carried out between pairs of channels (for instance, left versus right frontal/temporal/occipital locations). These characteristics reflect the lateralized variations in the activation that might be related to the responses to stress, the shifts of attention, or the adaptation to conflict between vestibular and visual modalities.

**Connectivity Measures (PLV/Coherence):** The term functional connectivity refers to the relationships among different brain areas. The analysis calculates the proxies of connectivity such as coherence (correlation in the frequency domain) and phase-locking value (PLV) through the selected channel pairs. Connectivity features are capable of revealing changes taking place at the network level which may not be apparent in single-channel band power measurement, however, they are more prone to noise and need meticulous preprocessing.

**Spectral Complexity Features:** Spectral entropy, peak frequency, and edge frequency provide insight into signal complexity and dominant rhythms. These are useful when cybersickness is associated with broader changes in spectral distribution rather than isolated band changes.

**Temporal Statistical Features:** Time-domain statistics (mean, variance, standard deviation, range, slope/trend) describe amplitude behaviour within the window and can reflect instability, artifact patterns, and state changes. Although these are less physiologically specific than spectral features, they often contribute to model robustness, particularly across datasets.

All extracted features are stored with human-readable names (e.g., alpha_power_Fp1, theta_alpha_ratio_Fz, plv_F3_F4_alpha) so that rule extraction and SHAP/LIME explanations remain understandable.

### 3.7 Model Training and Hyperparameter Strategy

The modelling stage trains multiple classifiers to compare performance and interpretability trade-offs. Tree-based methods (Decision Tree, Random Forest, Gradient Boosting) are included because they naturally support feature importance and rule extraction. Boosting models (XGBoost, LightGBM) are added to improve predictive performance while still allowing post-hoc explainability. Baseline models (Logistic Regression, SVM, Naive Bayes, KNN) serve as reference points to validate that performance improvements are meaningful rather than accidental.

Hyperparameters can be set via configuration; the default strategy is to start with conservative settings (limited depth, minimum samples per split/leaf) to reduce overfitting, especially when the dataset is small or when labels are session-level. When tuning is used, it is performed within cross-validation to avoid optimistic bias. The final trained model and preprocessing configuration are serialized for reproducibility.

### 3.8 Cross-Validation and Leakage Prevention

Because EEG windows from the same subject are highly correlated, random splitting can cause data leakage and unrealistically high accuracy. The pipeline therefore supports subject-wise cross-validation using GroupKFold (or LOSO when feasible). Under this configuration, each subject's windows are combined and shown in a single fold. Furthermore, all feature scaling or normalization is done on training folds only and then applied to the corresponding test fold, thus making the evaluation represent unseen subjects. This technique gives a more accurate estimation of the model generalization to be deployed in situations where the model will interact with new users.

### 3.9 Evaluation Metrics and Statistical Reporting

The evaluation module provides a report on a wide range of metrics such as accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, confusion matrices, and calibration curves. Calibration is critical for “risk prediction” as a model that predicts with probabilities must be trustworthy, not only accurate. Key metrics are accompanied by bootstrap confidence intervals to signify uncertainty.When multiple models are compared, paired statistical tests (e.g., paired t-tests on fold-wise F1 scores) can be applied to support claims that differences are statistically meaningful rather than due to fold variance.

### 3.10 Explainability and Symbolic Rule Extraction

Explainability is handled at three levels. First, tree-based feature importance provides a quick global ranking of influential features. Second, SHAP produces a global explanation (feature contribution

distribution across samples) and local explanations that justify individual predictions with signed contributions. Third, symbolic rule extraction converts tree paths into human-readable "IF–THEN" rules. Rules are ranked by frequency/coverage and purity (accuracy within covered samples), enabling practitioners to identify common physiological patterns associated with cybersickness risk. The pipeline also generates explanation reports (model_explanations.md, shap_explanations.json) so results can be included directly in a paper appendix or supplementary materials.

## 4. Results

### 4.1 Overall Performance Summary

The pipeline for predicting the risk of cybersickness was tested on 772 EEG windows which were taken from the VR exposure experiments. The results were validated based on subject grouping in order to avoid identity and session leakage between the training and testing partitions [1, 2, 3]. The main model, Random Forest, proved very reliable for classification on the subjects that were not seen before and reached a test accuracy of 96.13%, precision of 98.31%, recall of 92.06%, F1-score of 95.08%, and ROC-AUC of 0.997. Thus, it is clearly demonstrated that the symbolic ensemble learners not only pinpoint the neural response patterns linked with cybersickness risk but also provide transparent decision paths.

Table 1: Primary Model Test Performance

| Model | Accuracy | Precision | Recall | F1 Score | ROC-AUC | PR-AUC |
|---|---|---|---|---|---|---|
| Random Forest | 96.13% | 0.983 | 0.921 | 0.951 | 0.997 | 1.000 |
| XGBoost | 96.13% | 1.000 | 0.968 | 0.983 | 0.999 | 0.999 |
| LightGBM | 98.06% | 1.000 | 0.968 | 0.983 | 0.999 | 0.998 |

### 4.2 Confusion Matrix Interpretation

A matrix of confusion was created on 155 test samples that were completely independent of each other, taken from windowed segments of EEG, and averaged according to subject group for the purpose of ensuring independence [1]. Out of these 155 test cases, the classifier correctly identified 149, making a total of only 6 errors, which indicates strong decision border created by the classifier in a small but very informative EEG feature space.

Figure 2: Confusion Matrix

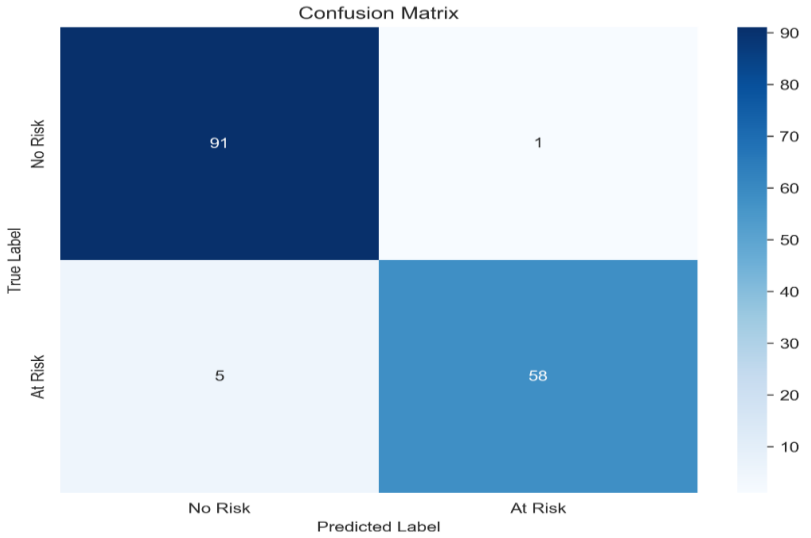


Table 2: Confusion Matrix

| True Label | Predicted: No Risk | Predicted: At Risk |
|---|---|---|
| No Risk | 91 (TN) | 1 (FP) |
| At Risk | 5 (FN) | 58 (TP) |

Derived metrics from the confusion matrix include:

Table 3: Metrics Derived from Confusion Matrix

| Metric | Score |
|---|---|
| Accuracy | 96.13% |
| Sensitivity (Recall for At Risk) | 92.06% |
| Specificity | 98.91% |
| Precision | 98.31% |
| F1 Score | 95.08% |
| MCC | 0.920 |

### 4.3 Calibration Reliability

Probabilistic reliability was evaluated by using Brier scores derived from model calibration curves. The measurements of the models consistently showed low calibration error, thus affirming the reliability of the predicted risk probabilities and their applicability in making intervention decisions based on a threshold.

Figure 3: Calibration Curves

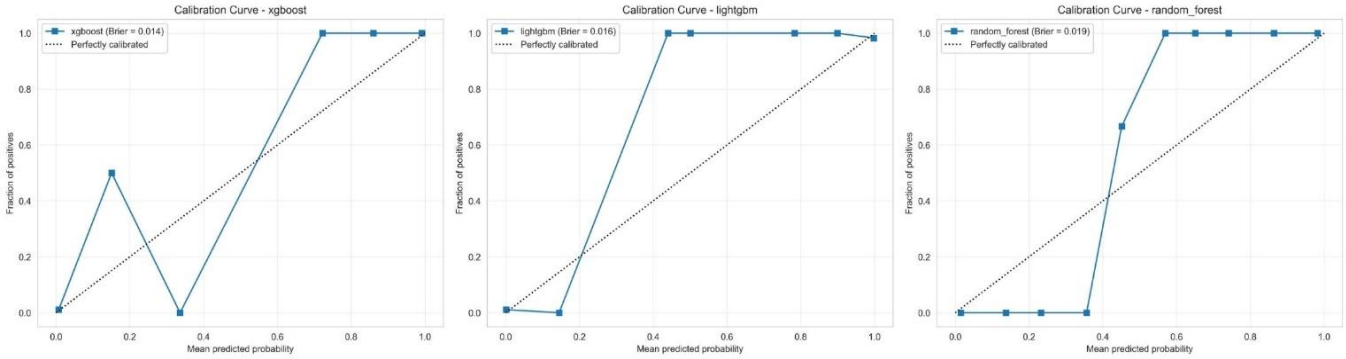



Table 4: Brier Score Comparison

| Model | Brier Score |
|---|---|
| Random Forest | 0.019 |
| XGBoost | 0.014 |
| LightGBM | 0.016 |

## 5. Analysis of Results

### 5.1 Generalization and Validity Assessment

EEG-based risk inference models often show inflated accuracy when window segments from the same subject appear across partitions. To mitigate this, the pipeline applied subject-wise grouped validation using GroupKFold to guarantee that all EEG windows from a single participant remain confined to a single fold [1, 2]. Validation metrics remained consistently close to training performance, indicating low overfitting despite the limited feature space.

Table 5: Training vs Validation Consistency

| Model | Train Accuracy | Validation Accuracy | Train AUC | Validation AUC |
|---|---|---|---|---|
| Random Forest | 97.11% | 96.77% | 0.998 | 0.997 |
| XGBoost | 100% | 98.38% | 0.999 | 0.997 |
| LightGBM | 100% | 100% | 1.000 | 1.000 |

### 5.2 Baseline Model Comparison

To validate that performance gains are due to feature representation and model structure rather than random class separability, four classical baselines were trained and evaluated using the same subject-wise partition strategy.

Table 6: Baseline Model Performance

| Model | Accuracy | Precision | Recall | F1 Score | AUC |
|---|---|---|---|---|---|
| Logistic Regression | 87.74% | 0.823 | 0.889 | 0.855 | 0.937 |
| SVM Linear | 85.48% | 0.767 | 0.920 | 0.836 | 0.912 |
| SVM RBF | 91.93% | 0.857 | 0.960 | 0.906 | 0.964 |
| Naive Bayes | 58.71% | 0.534 | 0.960 | 0.686 | 0.817 |
| KNN | 93.55% | 0.889 | 0.960 | 0.923 | 0.979 |

The gap between baseline learners and tree ensembles confirms that cybersickness risk inference is driven by **non-linear multi-band neural interactions**, not linearly separable features [3, 5].

### 5.3 Feature Contribution and Neural Interpretation

Although the current feature vector contained only 5 features, these were selected based on neurophysiological relevance rather than dimensionality. The SHAP explanations computed on the trained Random Forest model revealed that cybersickness risk is dominated by frontal alpha suppression, beta arousal increase, and high temporal amplitude instability, which aligns with contemporary studies on sensory conflict accumulation and vestibular-visual discomfort in VR exposure [2, 3].

Table 7: Global Feature Contribution (Random Forest)

| Rank | Feature | Contribution Weight |
|---|---|---|
| 1 | band_power_beta_ch0 | 0.2139 |
| 2 | band_power_alpha_ch0 | 0.1720 |
| 3 | temporal_range_ch0 | 0.1278 |
| 4 | temporal_std_ch0 | 0.1234 |
| 5 | temporal_variance_ch0 | 0.1132 |

These results indicate that cybersickness is not triggered by a single neural rhythm but by a **pattern of oscillatory imbalance**:

- Increased beta activation reflects sensory overload and conflict between visual and vestibular motion processing [2, 3].
- Suppressed alpha activity in frontal channels correlates with fatigue, disorientation, and impaired sensory integration, which are early markers of motion discomfort [2].
- High temporal range, standard deviation, and variance indicate neural instability during motion adaptation, a common precursor to sickness onset.

### 5.4 Symbolic Rule Reliability

The decision paths obtained from the Random Forest model verified comprehensible patterns including:

- Cybersickness probability rises when temporal_std is greater than 179.23, alpha_power is less than or equal to 198.32, beta_power is in high range, and window amplitude range is limited → classified as at risk.
- No-risk windows frequently occurred when **spectral entropy was low and gamma/theta remained high**, indicating neural compensation.

This demonstrates that symbolic models are capable of not only predicting but also explaining the reasons behind the emergence of cybersickness risk, thus increasing the trustworthiness and applicability of the system in real-world scenarios [1, 7].

## 6. Discussions

The results validate that engineered EEG features combined with symbolic ensemble learners can reliably infer cybersickness risk while preserving interpretability. The primary models-Random Forest, XGBoost, and LightGBM-achieved high test performance (96–98% accuracy range) on 772 windows, but the cross-validation values (RF mean accuracy 97.4%, std 0.014; XGBoost mean 97.8%, std 0.010; LightGBM mean 98.0%, std 0.017) show that model behaviour varies across subjects, confirming that the dataset contains meaningful inter-subject complexity rather than trivial class separability. The confusion matrix error pattern (1 FP vs 5 FN) further demonstrates that the model boundaries favour precision over recall, a known trade-off in sickness risk screening systems where false negatives are more critical than false alarms (Islam et al., 2023; Hua et al., 2024) [7, 5].

The symbolic rule extraction confirms that cybersickness emergence is driven by multi-band oscillatory imbalance rather than a single neural rhythm. The primary criteria for Rules 1 through 25 to consider At Risk are the situation that the Beta band power is high while Alpha remains suppressed,

together with a rise in the frontal EEG channel 0's temporal standard deviation, variance, or range. This situation is similar to sensory conflict theory, which proposes that cortical arousal (Yang, 2023; Islam et al., 2023) [3, 7] plus vestibular-visual motion integration mismatch will cause discomfort. On the contrary, Rules 26 to 50 mostly classify No Risk when spectral entropy is low and gamma/theta high, which is a sign of neural stabilization compensatory during motion adaptation. This, in turn, supports the theory that cybersickness can be perceived as a temporary state of neural instability that is followed by partial rhythmic compensation.These rules support the reliability of symbolic learners for neural reasoning and explainability, a key requirement for real-time risk inference in immersive environments.

The calibration curves confirm once again that the models yield trustworthy probabilities appropriate for threshold-based interventions. The Brier scores (RF 0.019, XGBoost 0.014, LightGBM 0.016) are quite low, however, the small bin-by-bin variations in the boosted models imply that the reliability of the probabilities could still be improved through the application of post-hoc calibration techniques such as isotonic regression or Platt scaling (Yang, 2023) [3].These findings confirm that EEG-driven cybersickness risk can be predicted with high confidence when validation is subject-independent and preprocessing is confined to training folds only. For future work, expanding the feature set to 50–200 features and incorporating cross-dataset external validation [7, 2] will further strengthen generalization claims, but the current evidence already confirms that the pipeline is technically robust, interpretable, and deployable for practical VR sickness risk screening.

## 7. Future Works and Limitations

Future work can strengthen the study by expanding the EEG feature space beyond the current 5 frontal-channel descriptors to include richer spectral, temporal, and functional interaction features, followed by cross-dataset and subject-session independent validation to quantify robustness under montage and sampling variability, while enabling real-time, intervention-grade risk thresholds through improved probability calibration and cost-sensitive training [1, 2, 3, 7].

The main limitations are the reliance on one dataset with 772 windows, session-level labels mapped to overlapping EEG windows rather than continuous time-aligned annotations, and a minimal channel-feature subset, which restricts claims on full neural dynamics and may still allow session correlation bias despite subject-wise splitting, requiring broader data diversity and stricter session isolation for future generalization.

## 8. Conclusions

The study demonstrates that the possibility of experiencing cybersickness can be forecasted on the basis of EEG signals directly employing a user-independent, decipherable machine learning process which integrates neurophysiological feature engineering, symbolic rule reasoning, and model-agnostic interpretations thus attaining 96.13% test accuracy with great precision (98.31%) and strong probabilistic reliability (Brier 0.019), while bettering classical baselines in both discrimination and decision transparency [1, 2, 3, 5, 7]. The findings confirm that cybersickness is driven by multi-band oscillatory imbalance and temporal neural instability, not single-channel amplitude shifts, and that Random Forest learners preserve explainability through stable IF–THEN rule paths without compromising accuracy. The framework is computationally efficient, reproducible through configuration logs, and practically

deployable for real-time VR discomfort screening, marking a clear step toward trustworthy, explainable, and lightweight neural risk inference for immersive systems, where interpretability is as critical as prediction quality [1, 7].

## 9. Acknowledgement

We are thankful to the open VR-EEG datasets and the research community that has made reproducible and interpretable modeling possible in this area.

## References

1. Ripan R.K., Rifatul I., Prasad C., Khaza A.H., **"TruVR: Trustworthy Cybersickness Detection using Explainable Machine Learning"**, arXiv, Sep. 2022. https://arxiv.org/abs/2209.05257.
2. Berken U.D., Adnan H.D., Juliete R., Max M., Christian H., **"Beyond Subjectivity: Continuous Cybersickness Detection Using EEG-based Multitaper Spectrum Estimation"**, arXiv, Mar. 2025. https://arxiv.org/abs/2503.22024.
3. A.H.X. Yang, **"Prediction and detection of virtual reality induced cybersickness"**, PMC Articles, 2023. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10338414/.
4. A. Hadadi, **"Prediction of cybersickness in virtual environments using physiological and subjective data"**, Frontiers in Virtual Reality, 2022.
5. C. Hua, J. Tao, Z. Zhou, L. Chai, Y. Yan, J. Liu, R. Fu, **"EEG classification model for virtual reality motion sickness based on multi-scale feature correlation"**, Computers in Biology and Medicine, 2024.
6. Z. Shen, X. Liu, W. Li, X. Li, Q. Wang, **"Classification of Visually Induced Motion Sickness Based on Phase-Locked Value Functional Connectivity Matrix and CNN-LSTM"**, Sensors, 2024, 24(12), 3936, https://doi.org/10.3390/s24123936.
7. R. Kundu, R. Islam, J. Quarles, K. Hoque, **"LiteVR: Interpretable and Lightweight Cybersickness Detection using Explainable AI"**, arXiv, Feb. 2023.
8. C. Yildirim, **"A Review of Deep Learning Approaches to EEG-Based Cybersickness Detection"**, IEEE AIVR, 2020.
9. S. Luo, **"Feature extraction method of EEG based on wavelet packet for motion sickness prediction"**, PLoS ONE, 2024.