International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 04(1) (April),2026
Received: 04-03-2026 Accepted: 11-04-2026 Published: 17-04-2026

EEG-Based Early Prediction of Cybersickness Using Machine Learning and Explainable AI

Y Nikhil Souri[1], G Pradeep Reddy[2], H Grace Lilly[3], G Poojitha[4], S. Mohammed Adnan[5], B. Ajay Kumar Reddy[6]

[1]*Student, CSE, JNTUA College Of Engineering, Anantapur, AP, India,* mymedia.yns@gmail.com.
[2]*Asst. Professor (Adhoc), CSE, JNTUA College Of Engineering, Anantapur, Andhra Pradesh, India*
[3]*Student, CSE, JNTUA College Of Engineering, Anantapur, Andhra Pradesh, India*
[4] *Student, CSE, JNTUA College Of Engineering, Anantapur, Andhra Pradesh, India*
[5] *Student, CSE, JNTUA College Of Engineering, Anantapur, Andhra Pradesh, India*
[6] *Student, CSE, JNTUA College Of Engineering, Anantapur, Andhra Pradesh, India*

***Abstract*—** Prediction of cybersickness, which refers to discomfort or sickness caused by virtual reality (VR), is an essential issue in developing VR environments. Proactive identification and mitigation of users' cybersickness can be achieved through early prediction of cybersickness based on physiological signals, such as EEG data. This paper proposes a machine learning framework for predicting users' cybersickness risk in advance using their EEG signals. For this task, we leverage spectral and temporal features extracted from EEG signals for training and validating several classification models. Experimental results obtained on the public VR EEG dataset show that ensemble learning algorithms, i.e., Random Forest, XGBoost, and LightGBM, yield superior performance, where the best-performing model achieves high accuracy and AUC score. In addition, XAI approaches, specifically, feature importance and SHAP method, are adopted to interpret classification models trained on EEG features to reveal important physiological predictors of cybersickness. Our experimental findings suggest that beta and alpha band power, along with temporal variability, contribute to cybersickness risk prediction.

***Keywords*—** Cybersickness, EEG Signal Processing, Virtual Reality, Machine Learning, Random Forest, XGBoost, Explainable AI, SHAP, Biomedical Signal Analysis

INTRODUCTION

Virtual Reality (VR) technologies have gained significant popularity; however, users experience cybersickness as a major hindrance to their virtual reality experience because the condition causes them to feel dizziness, nausea, and disorientation during virtual reality use. Users experience reduced comfort through cybersickness which also restricts their ability to use and accept virtual reality systems that create immersive environments. [1], [2].

The conventional methods used to assess cybersickness depend on subjective questionnaires which include the Simulator Sickness Questionnaire (SSQ) that researchers administer after exposure to test individuals because these assessments do not provide real-time detection capabilities. The current situation demands development of objective tracking systems which detect cybersickness symptoms during their initial stages. Scientists consider physiological signals especially electroencephalography (EEG) as a strong solution because these signals show brain activity changes when people experience motion perception together with sensory conflicts [3].

EEG-based analysis has found widespread application in multiple biomedical fields and brain-computer interface applications which include cognitive state monitoring and emotion recognition

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 04(1) (April),2026
Received: 04-03-2026 Accepted: 11-04-2026 Published: 17-04-2026

and neurological disorder detection [4][8]. In virtual reality environments, EEG signals demonstrate changes in brain activity through different frequency bands which include alpha, beta, and gamma bands that scientists associate with sensory processing and discomfort. Researchers can create predictive models through machine learning methods which use these signals to forecast cybersickness risk before users develop severe symptoms.

Machine learning algorithms, such as Random Forest, XGBoost, and LightGBM, have proven highly efficient when working on classification problems with complex and high-dimensional input data [5] – [7]. Such ensemble methods can effectively deal with nonlinear relations and interaction among features within the EEG dataset. However, one of the crucial issues of machine learning in medicine is the lack of interpretability, which is crucial for understanding the physiological background of the prediction results and gaining trust to apply these approaches in practice.

As an innovative solution to overcome this obstacle, Explainable Artificial Intelligence (XAI) methods have been developed. Shapley Additive explanations (SHAP) is a new technique for evaluating the impact of individual features on the model's decision-making process [9], [10]. Implementing XAI algorithms in EEG-based classification systems makes it possible not only to achieve reliable performance but also to gain insight into the neural signatures responsible for cybersickness.

This paper presents a novel machine learning-based algorithm for early prediction of cybersickness risk based on features extracted from EEG data. The proposed methodology combines the feature engineering, ensemble machine learning algorithms, and XAI approaches to ensure accurate prediction and explainability of results.

METHODOLOGY

The proposed system offers a full pipeline for the early prediction of the risk of cybersickness with EEG signals. This methodology combines signal preprocessing, feature engineering, classification based on machine learning, and XAI to ensure high performance coupled with interpretability.

A. *Data Acquisition and Preprocessing*

The EEG data used in this study is acquired from a public virtual reality (VR)-based EEG dataset, which involves exposing participants to immersive virtual environments. Before use, the raw EEG signals are segmented into fixed-length windows. Labels for each segment are assigned based on experimentally defined risk levels.

To improve the signal quality, preprocessing procedures like removal of noises, normalization, and filtering are performed. The EEG signals are broken down into the standard bands of frequency that are the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma above 30 Hz, which is considered to represent cognizance and sensory processing [4], [8].

B. *Feature Extraction*

The system extracts a complete collection of manually built features from each EEG window to capture its spectral and temporal aspects. The extracted features include:

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 04(1) (April),2026
Received: 04-03-2026 Accepted: 11-04-2026 Published: 17-04-2026

- Band Power Features: Energy distribution across frequency bands
- Spectral Features: Spectral entropy, peak frequency, and spectral edge frequency
- Temporal Features: Mean, variance, standard deviation, and signal range
- Derived Features: Band ratios and statistical descriptors

These features not only provide a compact but also an informative representation of EEG signals for effective learning by machine learning models.

C. *Machine Learning Models*

For classifying the risk of cybersickness, a number of supervised learning methods have been considered. The ensemble learning algorithms have mostly been preferred because of their robustness:

- Random Forest (RF): An ensemble of decision trees that improves generalization through bagging [7]
- XGBoost: A gradient boosting framework optimized for speed and performance [5]
- LightGBM: A highly efficient boosting algorithm designed for large-scale data [6]

These models are trained using extracted EEG features and evaluated using standard classification metrics. Their ability to capture nonlinear relationships makes them suitable for EEG-based prediction tasks.

D. *Cross-Validation and Statistical Evaluation*

To ensure generalizability and robustness, a stratified k-fold cross-validation technique is implemented. In the current study, a 10-fold cross-validation scheme is applied, which involves splitting the data set into training and validation sets several times.

Furthermore, statistical significance tests, such as paired t-tests are performed to evaluate differences in the models' performances among different folds. This will eliminate any chance that performance discrepancies are the result of randomness.

E. *Explainable Artificial Intelligence (XAI)*

The interpretability property is a fundamental requirement in biomedical use cases. To this end, SHAP (Shapley Additive Explanations) is used to formally assign contribution values to each feature in the model predictions [9].

SHAP values provide both:

- Global explanations: Overall feature importance across the dataset
- Local explanations: Feature contributions for individual predictions

The system identifies essential EEG characteristics which affect cybersickness risk assessment thus building trust in the model through enhanced transparency. The use of XAI aligns with recent trends in interpretable machine learning for healthcare applications [10].

F. *System Architecture*

The complete design of the suggested cybersickness prediction system appears in Figure 1. The system processes its operations through a structured pipeline which starts with raw EEG signal acquisition and continues with preprocessing and segmentation into temporal windows. The system

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 04(1) (April),2026
Received: 04-03-2026 Accepted: 11-04-2026 Published: 17-04-2026

requires feature extraction to obtain both spectral and temporal characteristics from EEG signals. The extracted features serve as input data to train machine learning models which will perform classification tasks. The system employs explainability techniques to analyse the model prediction results.

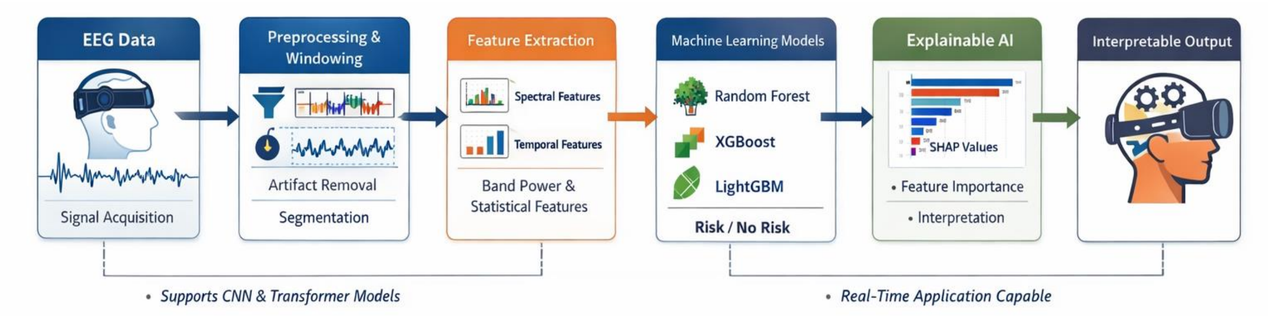

Fig. 1 Architecture Diagram

The architecture is modular and scalable; therefore, it can be implemented for conventional machine learning techniques as well as sophisticated deep learning techniques. This makes experimentation, replication, and even deployment in real-time feasible in virtual reality settings.

RESULTS

The proposed system for prediction of cybersickness based on an EEG signal was tested on a publicly available EEG VR database that contains 772 samples. The EEG dataset comprises two classes: “No risk” (59.5%) and “At risk” (40.5%), thus giving rise to a fairly balanced classification task.

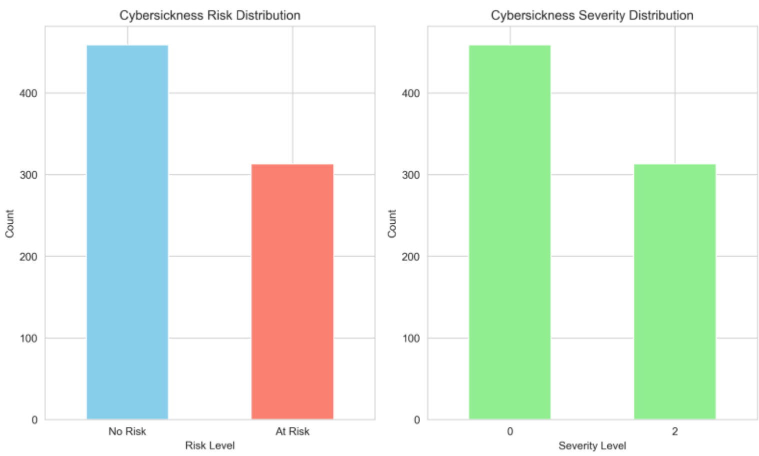

Fig. 2 Class Distribution Graph

A. Model Performance

The classification quality of the proposed system was measured by means of traditional metrics: accuracy, precision, recall, F1-score and AUC. The best-performing model achieved the following results:

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 04(1) (April),2026
Received: 04-03-2026 Accepted: 11-04-2026 Published: 17-04-2026

- Accuracy: 0.961
- Precision: 0.983
- Recall: 0.921
- F1-score: 0.951
- AUC: 0.997

These results indicate strong predictive capability across all evaluation metrics.

B. ROC and Precision–Recall Analysis

The Receiver Operating Characteristic (ROC) curves for the evaluated models are shown in Fig. 3. The AUC values for all models are close to 1.0 which demonstrates their ability to distinguish between different classes. The Precision–Recall (PR) curves shown in Fig. 4 demonstrate that the classification models maintain their ability to accurately identify both classes through all threshold settings of the different test conditions.

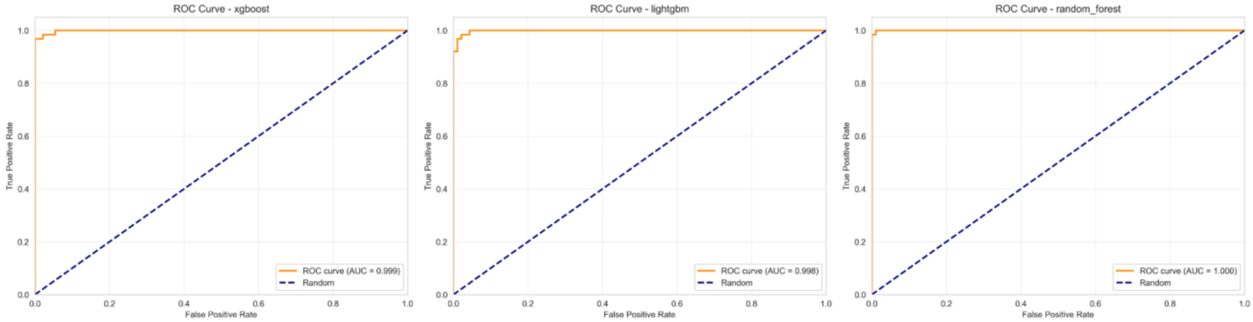

Fig. 3 ROC Curves of all models

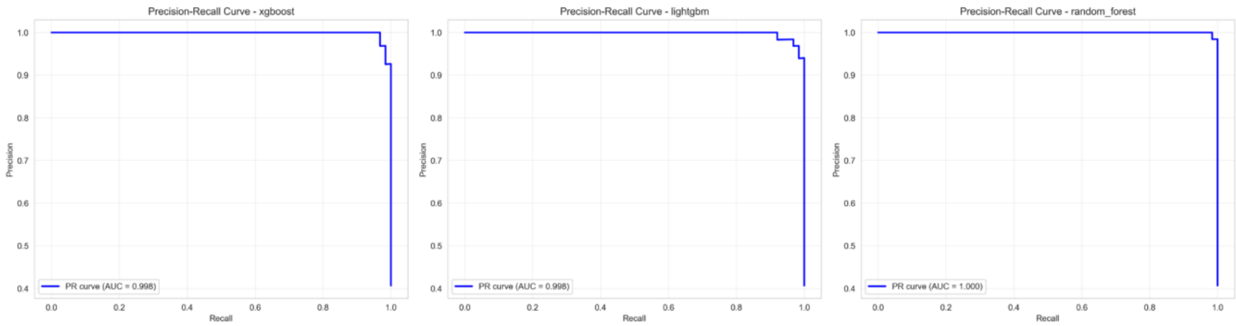

Fig. 4 Precision Recall of all models

C. Calibration Analysis

Fig. 5 shows the calibration plots used to measure the accuracy of probability estimates from the trained models. From the plots, it is clear that there is a high degree of correlation between probability predictions and actual outcomes.

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 04(1) (April),2026
Received: 04-03-2026 Accepted: 11-04-2026 Published: 17-04-2026

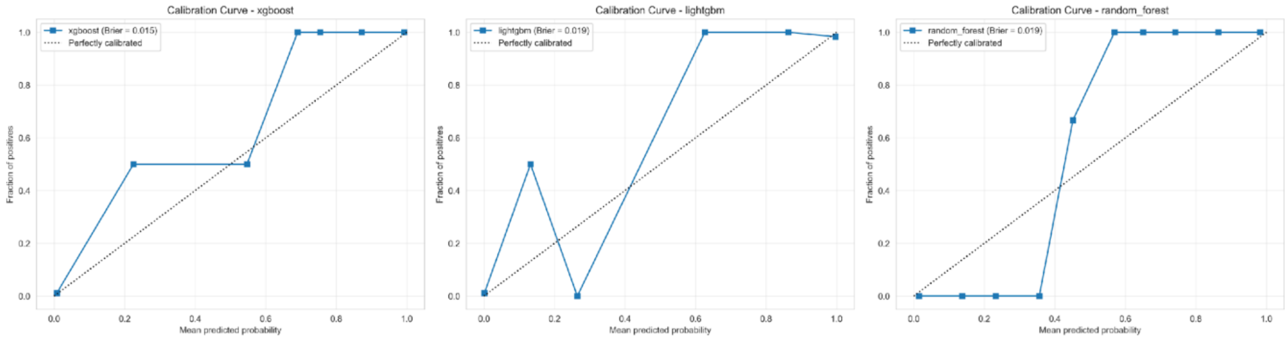

Fig. 5 Calibration Curves of all models

D. *SHAP-Based Feature Importance*

Table I presents the top EEG features ranked by SHAP importance values which are derived from the explainability analysis.

TABLE I
Top EEG Features Ranked by SHAP Importance

FEATURE NAME	SHAP IMPORTANCE
band_power_beta_ch0	0.1179
band_power_alpha_ch0	0.0810
temporal_std_ch0	0.0714
band_power_gamma_ch0	0.0554
temporal_variance_ch0	0.0548
temporal_range_ch0	0.0412
spectral_edge_frequency_ch0	0.0272
spectral_entropy_ch0	0.0205
band_power_delta_ch0	0.0194
spectral_peak_frequency_ch0	0.0188
band_power_theta_ch0	0.0071

E. *Summary of Results*

The obtained experimental results show that the proposed methodology is efficient regarding classification accuracy, which is evidenced by several evaluation measures. The additional assessment using ROC, PR, and calibration plots confirms the reliability and robustness of the obtained models. The SHAP-based feature importance analysis helps understand the contribution of EEG-based features.

CONCLUSIONS

In this paper, a full machine learning-based framework for the early prediction of cybersickness risk using EEG signals is presented, where the proposed approach combines EEG signal preprocessing, feature extraction, ensemble-based classification models, and explainable artificial intelligence (AI) techniques for accurate and interpreted predictions.
Experimental evaluation on a VR-based EEG dataset shows that the proposed system performs well against several metrics, such as accuracy, precision, recall, F1-score, and AUC. It is shown that ensemble learning methods such as Random Forest, XGBoost, and LightGBM are effective

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

Received: 04-03-2026 Accepted: 11-04-2026 Published: 17-04-2026

for grasping the complex patterns of EEG signals. Also, the addition of explainability has enabled the identification of most influential features that help in prediction thus increasing reliability and transparency of the model.

Besides, the proposed framework is a scalable and efficient cybersickness prediction framework and can be extended to real-time applications of the proposed classifiers in adaptive virtual reality systems. The system proposed in this framework can be used to improve upon user experience, safety, and usability in an immersive environment by predicting and therefore detecting any discomfort the user experiences early enough.

Future work can be directed toward incorporating deep learning-based EEG representations, multi-dataset validation, and real-time deployment to strengthen robustness and generalization.

Acknowledgment

The authors would like to sincerely thank Mr. G. Pradeep Reddy, M.Tech, Ph.D, Assistant Professor (Adhoc), Department of Computer Science and Engineering, JNTUA College of Engineering, Anantapur, for his invaluable advice, unwavering support, and perceptive recommendations during this work.

References

[1] S. Weech, S. Kenny, and M. Barnett-Cowan, “Presence and cybersickness in virtual reality are negatively related: A review,” Frontiers in Psychology, vol. 10, p. 158, 2019.

[2] K. Stanney, R. Kennedy, and J. Drexler, “Cybersickness is not simulator sickness,” Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 41, no. 2, pp. 1138–1142, 1997.

[3] M. A. Alizadeh and S. M. M. Fard, “EEG-based detection of motion sickness in virtual reality using machine learning,” Biomedical Signal Processing and Control, vol. 68, p. 102598, 2021.

[4] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, “Automated EEG analysis of epilepsy: A review,” Knowledge-Based Systems, vol. 45, pp. 147–165, 2013.

[5] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD, 2016, pp. 785–794.

[6] G. Ke et al., “LightGBM: A highly efficient gradient boosting decision tree,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017.

[7] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[8] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, “A review of classification algorithms for EEG-based brain–computer interfaces,” Journal of Neural Engineering, vol. 4, no. 2, 2007.

[9] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017.

[10] Z. J. Wang, O. Turaga, and J. Yuan, “Explainable AI for EEG-based medical diagnosis: A review,” IEEE Reviews in Biomedical Engineering, vol. 15, pp. 264–279, 2022.

2250-3676 ijesatj@gmail.com https://www.ijesat.com/