Investigating Personalization Techniques for Improved Cybersickness Prediction in Virtual Reality Environments

Umama Tasnim*, Rifatul Islam*, Kevin Desai, and John Quarles

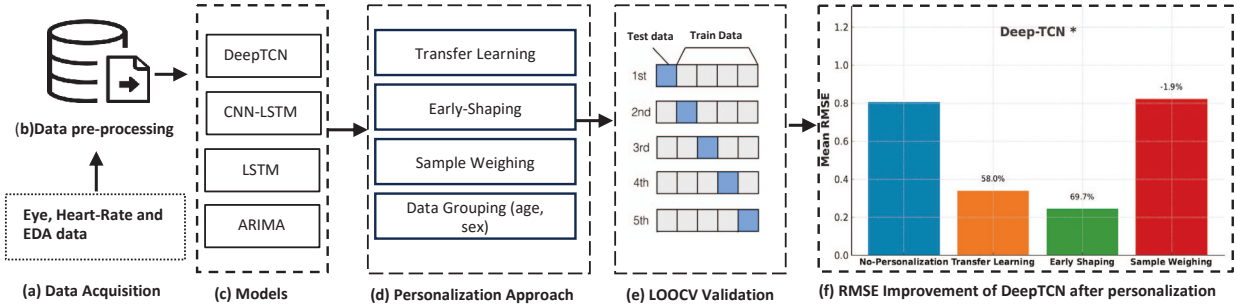

Figure 1: Overview of the Proposed Approach: (a) Simulation 21 data acquisition which includes eye-tracking, heart rate, and EDA data; (b) Preprocessing of the data; (c) Cybersickness Prediction Models(d) Outline of the personalized approach; (e) Leave-One-Out Cross-Validation (LOOCV) method; (f) Summary of DeepTCN model results demonstrating the significant enhancements achieved by our proposed models over non-personalized models.

Abstract—In recent cybersickness research, there has been a growing interest in predicting cybersickness using real-time physiological data such as heart rate, galvanic skin response, eye tracking, postural sway, and electroencephalogram. However, the impact of individual factors such as age and gender, which are pivotal in determining cybersickness susceptibility, remains unknown in predictive models. Our research seeks to address this gap, underscoring the necessity for a more personalized approach to cybersickness prediction to ensure a better, more inclusive virtual reality experience. We hypothesize that a personalized cybersickness prediction model would outperform non-personalized models in predicting cybersickness. Evaluating this, we explored four personalization techniques: 1) data grouping, 2) transfer learning, 3) early shaping, and 4) sample weighing using an open-source cybersickness dataset. Our empirical results indicate that personalized models significantly improve prediction accuracy. For instance, with early shaping, the Deep Temporal Convolutional Neural Network (DeepTCN) model achieved a 69.7% reduction in RMSE compared to its non-personalized version. Our study provides evidence of personalization techniques' benefits in improving cybersickness prediction. These findings have implications for developing personalized cybersickness prediction models tailored to individual differences, which can be used to develop personalized cybersickness reduction techniques in the future.

Index Terms—Cybersickness, Cybersickness Personalization, Cybersickness Prediction, Transfer Learning, Early Shaping, Deep Learning, Machine Learning

1 Introduction

Despite the vast potential of virtual reality(VR) in gaming, education, healthcare, and entertainment, the comfortable use of VR is often impeded by cybersickness—a discomfort akin to motion sickness experienced by some users when engaging with VR. These discomforts may include but are not limited to dizziness, headaches, sweating, and nausea [22]. Moreover, individual factors, including age, gender, and prior VR experience, often influence the severity of these symptoms [4] referred to as individual susceptibility to cybersickness [48]. For instance, previous results have suggested that younger individuals are generally less susceptible than older ones, women more susceptible than men, and novice users in VR are likely more prone to cybersickness [37]. Despite the inconsistency in frequency and severity of cybersickness observed across prior research, attributed to individual differences [61], integrating these differences into the automatic prediction and mitigation of cybersickness remains a largely underexplored research area.

A range of subjective questionnaires has been proposed by researchers as potential instruments for the assessment and prediction of cybersickness [9, 14, 22]. For example, the Simulator Sickness Questionnaire (SSQ) [22] and the Virtual Reality Motion Sickness Questionnaire (VRSQ) [26] aim to evaluate cybersickness severity following VR immersion. However, these post-immersion questionnaires are unable to provide a granular understanding of cybersickness during immersion and fail to integrate individual differences. In contrast, the cybersickness Susceptibility Questionnaire (CSSQ) [9] and the Motion Sickness Susceptibility Questionnaire (MSSQ) [14] are designed to determine an individual's susceptibility to cybersickness before VR immersion. Despite this, their current iterations have not been thoroughly validated for predictive performance in estimating cybersickness and do not adequately account for the integration of individual differences [29,42]. If users with high cybersickness susceptibility can be accurately identified, VR experiences could be tailored with slower movements, reduced visual effects, and shorter durations to minimize discomfort. Personalized cybersickness predictions could help users gradually build up their VR tolerance, allowing them to enjoy more

- *Umama Tasnim and Rifatul Islam are with Kennesaw State University and have contributed equally. E-mail: utasnim@kennesaw.edu (Umama Tasnim), rislam11@kennesaw.edu (Rifatul Islam)*
- *Kevin Desai and John Quarles are with the University of Texas at San Antonio. E-mail: kevin.desai@utsa.edu (Kevin Desai), john.quarles@utsa.edu (John Quarles)*

Manuscript received 4 October 2023; revised 17 January 2024; accepted 24 January 2024. Date of publication 4 March 2024; date of current version 15 April 2024. Digital Object Identifier no. 10.1109/TVCG.2024.3372122

complex and immersive experiences without discomfort over time.

Recently, deep learning approaches have emerged as promising methods for predicting cybersickness. Various deep learning techniques, such as those utilized by Islam et al. [16, 17], have achieved high predictive accuracy through the use of eye-tracking and head-tracking data. Additionally, hybrid multimodal deep fusion neural networks, including Long Short-Term Memory networks, N-BEATS, and Deep Temporal Convolutional Networks (DeepTCN) have been employed to anticipate cybersickness onset up to 60 seconds in advance. Other researchers have unveiled significant associations between cybersickness intensity and physiological parameters, such as brain activity and individual factors like age and sensitivity [19, 34, 45]. However, the impact of individual differences on cybersickness prediction remains understudied [62].

To address these inconsistencies, this study focuses on personalizing cybersickness prediction methods, considering deep-learning approaches like data grouping based on age and gender, early-shaping, and transfer learning [54](Figure). The results indicate that age and gender significantly impact cybersickness prediction, especially concerning eye-tracking data. Furthermore, the personalization methods providing the best outcomes were transfer learning and early-shaping, emphasizing the potential of personalized models to enhance the performance of cybersickness prediction [57, 61].

Our vision is that each user should have a model that is personalized to their data. Based on our current approach, this would require users to interact with VR for a few minutes. Data collected from this short interaction could be used to fine-tune an existing aggregate model with their specific data, either through early shaping, or transfer learning. Gender-specific models would already be pre-trained and only require that the user select their gender, which would select the appropriate model.

Overall, this study offers critical insights into the significance of personalization in cybersickness prediction, emphasizing that the application of personalization techniques can considerably enhance the performance of these prediction models. The findings presented herein underscore the potential benefits of incorporating personalized models in the prediction and mitigation of cybersickness, thereby contributing to improved user experiences in virtual environments. Future investigations in this domain should continue exploring and refining cybersickness reduction strategies, delving further into developing more sophisticated personalized models.

2 Background

In the subsequent section, we briefly discuss several key aspects pertinent to our work: 1) cybersickness , 2) cybersickness measurement, 3) the current state of the art in cybersickness prediction, 4) the influence of individual factors in cybersickness, and 5) personalization of deep learning. These components are intended to provide a holistic comprehension of the existing challenges and underpin the motivation for our research endeavor, underscoring the need for advancing personalized cybersickness prediction methods.

2.1 Cybersickness

Cybersickness is a collection of motion sickness-like symptoms that can happen both during and after the VR experience, including cold sweats, nausea, dizziness, and disorientation [23]. Vestibular (balance and spatial orientation senses) and visual information processing are the two main senses affected by cybersickness [11]. Popular explanations for cybersickness include the postural instability theory and the sensory conflict theory [7, 32, 58]. The most widely accepted of these theories is the sensory conflict theory, which contends that people who experience cybersickness do so because they mistakenly perceive that they are moving when, in fact, they are still. Postural instability occurs when a person's perception and familiar surroundings are inconsistent [24].

2.2 Cybersickness Measurement

Most previous work has assessed cybersickness through self-reported questionnaires. The simulator sickness questionnaire (SSQ) is the most widely used tool for assessing the signs of cybersickness [22]. The SSQ consists of 16 questions broken down into three categories (nausea, oculomotor, and disorientation) to assess the severity of each potential cybersickness symptom. In addition, Hyun et al. introduced the Virtual Reality Sickness Questionnaire (VRSQ) [26] which reduced the previous SSQ items from 16 to 9. This also resulted in limiting the assessment to two factors: oculomotor and disorientation. Even though the VRSQ was created specifically for VR, the SSQ is still widely used as the validity of the VRSQ has not yet been universally accepted.The Fast Motion Sickness Scale (FMS) was established by Keshavarz et al. in response to the limitations of subjective measurements following VR immersion [25]. High correlations were found between the SSQ and the FMS scoring system, which asks for short self-assessments on a 0-to-20 scale. However, to gather the subjective measurements of cybersickness during VR immersion, the dataset we used included an FMS scale from 0-10.

While subjective measures of cybersickness are useful, researchers have also conducted objective measures to better understand and quantify the severity of cybersickness [12, 28, 59]. Objective measures include physiological measures such as heart rate (HR), heart-rate variability (HRV), electrodermal activity (EDA), stomach tachyarrhythmia, eye-blink rate, and electroencephalography (EEG) [13, 44, 51]. These measures can be used to assess when and how cybersickness is affecting an individual and help compare the severity of cybersickness across different individuals and virtual environments. Recent research has focused on the correlation between other objective measures and cybersickness, including gaze, pupillometry, and postural stability [1, 5, 15, 43]. Nam et al. [43] investigated cybersickness by using physiological measures such as the center gaze ratio and scanpath length. According to Reddy et al. the change in pupil diameter is a physiological measure of cybersickness [49]. Litleskare et al. described in their study how postural stability can be used to assess cybersickness among a group of people in terms of both predictors and objective measures [35]. Kourtesis et al. introduce The CSQ-VR, which makes it possible to evaluate cybersickness while engaging in VR, based on pupil size, a biomarker of cybersickness [30]. Compared to SSQ and VRSQ, CSQ-VR showed noticeably improved internal consistency.

2.3 Cybersickness Prediction

There has been much research on predicting cybersickness more precisely. Monteiro et al. investigated that compression rate can be used as a potential marker or indicator for cybersickness [41]. The compression rate refers to the amount of data being transmitted from the VR system to the user, which can be affected by various factors such as the complexity of the virtual environment, the frame rate of the system, and the latency of the tracking sensors. They focused on how users' cybersickness levels were recorded while playing two VR games and found correlations between cybersickness and variations in the compression rate of movement data. Martin et al. used machine learning models to predict cybersickness using physiological signals (heart rate and electrodermal activity) that were recorded during VR game sessions with an accuracy of up to 91 percent and explained variance up to 75 percent (in a regression approach) [38]. Kundu et al. [31] used three xML (Logistic Regression, Decision Tree, and Explainable Boosting Machine (LR)) models to predict cybersickness using publicly available physiological and gaming datasets. Islam et al. [17] proposed multimodal deep fusion neural networks in order to forecast cybersickness 30–90 seconds in advance. Although there has been increasing research on automatic cybersickness prediction, very few researchers have focused on improving cybersickness prediction through personalization. Kim et al. described that it is feasible to automatically anticipate the cybersickness level that represents brain activity [27]. They captured brain activity using 8 channels of EEG data as more than 200 individuals viewed 44 different VR contents. After extensive preparation, they demonstrate that, without EEG data, the proposed approach reliably assesses cognitive states.

2.4 Individual differences in cybersickness susceptibility

Cybersickness has become one of the major research questions in virtual reality environments. More precisely, some research shows that cybersickness varies from person to person according to race, age, gender, etc [10, 39, 40, 47]. A recent study by Pöhlmann et al. highlights the importance of gender for cybersickness studies. Martingalo et al. [39] investigate if the susceptibility to cybersickness varies among racial groupings. Six independent, racially varied samples (n = 931) were used to gather self-reported cybersickness ratings. Three out of six studies revealed significant racial disparities in cybersickness. In comparison to white individuals, black participants experienced cybersickness on average a third of a standard deviation lower (d = -0.31, p ¡.001). Melo et al. [40] study focused on examining the role type and gender in a VE and their effects on the perception of presence and cybersickness. They conclude no disparities between the genders were discovered using immersive virtual reality (VR) setups. Luong et al. [36] investigated the relationship between cybersickness and demographic, user experience, and behavioral parameters; they conducted a large lab-in-the-field study with (n = 837). The experiment concludes that, compared to male participants, female participants experienced significantly higher degrees of cybersickness, and no significant effect was found on age. This leads us to investigate more about the personalization of cybersickness [10].

2.5 Deep Learning Models for Cybersickness Predictions

In the context of our study, we briefly discuss the deep learning models that we have used for cybersickness prediction and personalization.

The Deep Temporal Convolutional Network (DeepTCN): The DeepTCN model has emerged as a robust model for predicting cybersickness [17]. This model is characterized by its deep learning architecture, which leverages temporal convolutions to effectively capture time-dependent features in data [6, 33]. The model's strength lies in its ability to handle long sequences of data, making it particularly suitable for analyzing time-series data associated with cybersickness symptoms. Unlike traditional convolutional neural networks, DeepTCN employs dilated convolutions, enabling the model to cover a broader range of input data without losing resolution or clarity. Its layered structure, consisting of convolutional layers followed by pooling layers, allows for the extraction of complex patterns and dependencies in the data. In the context of cybersickness prediction, DeepTCN's architecture facilitates the analysis of physiological signals over time, offering nuanced insights into how symptoms develop and change in response to various stimuli [17].

Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM): The CNN-LSTM model, a hybrid architecture that combines Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, offers a unique approach to cybersickness prediction [16, 18, 20]. The CNN component extracts spatial features from input data, such as patterns in physiological signals, while the LSTM part efficiently processes time-series data, capturing temporal dependencies. This synergy makes the CNN-LSTM model adept at handling the complex nature of cybersickness, where both spatial and temporal features play a crucial role. Separately, the LSTM model, with its ability to remember long-term dependencies, is also well-suited for cybersickness prediction. It effectively analyzes time-series data from physiological sensors, providing insights into the progression of cybersickness symptoms over time.

Autoregressive Integrated Moving Average (ARIMA): The ARIMA model, traditionally used in econometrics and weather forecasting, has been adapted for cybersickness prediction [60]. This model excels due to its straightforward yet efficient approach to analyzing time-series data, especially when dealing with non-stationary datasets. In the context of cybersickness, ARIMA can be used to analyze trends and patterns in physiological data over time, providing forecasts of cybersickness occurrence based on the data. In our research, we have employed ARIMA to establish baseline predictions against which more complex models are compared. While it may not capture the intricate relationships between various physiological indicators as effectively as deep learning models, ARIMA's predictions offer valuable benchmarks and have been instrumental in validating the more sophisticated approaches taken by DeepTCN, CNN-LSTM, and LSTM models in cybersickness prediction.

2.6 Personalizing Prediction Models

There has been some work on how to personalize models, e.g., deep learning models, but not with respect to cybersickness prediction. We investigated the personalization of cybersickness prediction models with the following methods.

2.6.1 Grouping

Data grouping is one of the first methods we tried to investigate in our model. Grouping with common characteristics can help to improve prediction for other data in that group. For example, we might consider demographic information (such as age, gender, and race). In a similarity-based approach, data points that are similar to each other are grouped together [8]. In our work, we grouped the data based on participant age and gender using a similarity-based approach.

2.6.2 Transfer Learning

Transfer learning refers to the process of enhancing learning in a new task by leveraging knowledge gained from a related task that has already been mastered [63]. Typically, a pre-trained model has already learned to recognize features from a large dataset. By using a pre-trained model as a starting point, the new model can learn more efficiently, as it does not need to start from scratch and learn all the features from the data. Instead, it can build on the features learned by the pre-trained model and fine-tune them for the specific task. This approach has been shown to be effective in many computer vision and natural language processing tasks, where large amounts of labeled data are often required for training deep learning models. According to Schneider et al. [54] Transfer learning is one of the proven personalization methods.

2.6.3 Early Shaping

Early shaping (i.e., Curriculum learning) refers to a machine learning technique where the training data is presented to the model in a pre-defined order or curriculum, with easier examples presented first and more complex examples gradually introduced, to facilitate better and faster learning [3]. This is done to help the model learn more efficiently and generalize better to new, unseen data. According to Schneider et al. [54], early shaping is another personalization method that is similar to transfer learning but performed in the opposite order. For early shaping, we trained the model based on one individual's data first and then finished the training with the rest of the data from the training dataset.

2.6.4 Sample Weighing

Sample weighing is a process of assigning different weights to individual data points or observations in a dataset to account for their varying levels of importance or representations in the analysis. This technique is commonly used in statistics and machine learning to address imbalanced data distribution, where some classes or categories have significantly fewer examples than others. The idea behind sample weighing is to give more weight to samples that are more important or difficult to learn, and less weight to samples that are less important or easier to learn [54]. By doing this, the model can learn more effectively from the data and improve its performance. The weights can then be used during training to adjust the contribution of each sample to the overall loss function that the model is trying to optimize.

3 Proposed Contribution

Existing studies have established that individual characteristics, including race, age, gender, eye movement, and physiological responses, significantly influence the onset of cybersickness [57, 61]. In light of these findings, our primary research question (**RQ**) is formulated as follows: RQ_1: ***"How does the performance of a model, trained for either individual users or small groups sharing common individual factors such as age and gender, impact cybersickness prediction?"***.

We hypothesize that, ***"A personalized cybersickness prediction model will outperform non-personalized models when trained based on individual factors"***. To test and validate the hypothesis and identify the most effective method of personalization for cybersickness prediction, we developed and evaluated four distinct personalization strategies. The key contributions of this paper are as follows:

- **Development of Personalized Cybersickness Prediction Models**: We adapted several established deep learning personalization techniques, such as Transfer Learning, Early Shaping, and Sample Weighting [54], specifically for cybersickness prediction.
- **Incorporation of Individual Factors**: Our approach involved accounting for individual differences, such as age and gender, in cybersickness. We utilized data grouping strategies in both the training and testing phases of our cybersickness prediction models.
- **Application to Various Deep/Machine Learning Models**: In line with previous studies [17], we applied personalization techniques to several models previously used for cybersickness prediction, including Deep-TCN, ARIMA, LSTM, and CNN-LSTM.
- **Analysis of Training Data Impact**: We conducted an extensive analysis to understand the influence of different types of training data (e.g., eye tracking, heart rate, and electrodermal activity) on the personalized model's performance.
- **Rigorous Model Evaluation**: For evaluating our models, we employed the leave-one-out cross-validation (LOOCV) method. In this method, the model is trained on all data points except one, which serves as the test set. This comprehensive approach allows our models to be tested against all possible scenarios within the dataset, offering a more rigorous evaluation compared to a random train/test split.

The methodologies and findings presented in this paper significantly advance the field of personalized cybersickness prediction, offering a new paradigm for understanding and mitigating the impact of cybersickness across diverse user groups.

4 METHODS

In this research study, we employed several methods for personalizing cybersickness prediction models, including data grouping, early shaping, sample weighing, and transfer learning (Figure 2). To assess the efficacy of these personalization approaches, we employed previously established models for predicting cybersickness, including DeepTCN, CNN-LSTM, LSTM, and ARIMA [5, 17, 38]. The models are tested and validated using data sources by Islam et al. [17].

4.1 Dataset

In this study, we employed the 'Simulations 2021' public dataset[1] and extended the open-source code base developed by Islam et al. [17]. The 'Simulations 2021' dataset includes data from 30 participants, characterized by a mean age of 30.04 years (standard deviation: 4.12 years) and diverse ethnic backgrounds, including Asian, Black, Caucasian, Hispanic, and Pacific Islander. This dataset encompasses a range of physiological metrics — heart rate (HR), eye tracking, head tracking, and electrodermal activity (EDA) — along with individual demographic factors (age, sex, ethnicity). Additionally, it features the Fast Motion Sickness Scale (FMS) scores, ranging from 0 to 10, which serve as the ground truth. These scores represent the participants' discomfort levels compared to their baseline comfort state at rest (FMS = 0). Although the dataset contains head-tracking data, for this study, we did not consider head-tracking data as they were not reported not significant in terms of cybersickness prediction by Islam et al. All the data were time-synchronized for each individual. The details of the dataset are presented in Table 1.

[1] Simulation 2021 Dataset: https://sites.google.com/view/savelab/research

Table 1. Simulation 2021 Dataset Description by Islam et. al [17]. Head-tracking data is not used in this study as they were reported as not significant.

Data Type	Data
Participant Information	Number of Participants: 30 Mean Age: 30.04 years Standard Deviation of Age: 4.12 years Male: 15, Female: 15 Individual Factors: Age, Sex FMS Score Range: 0-10 (Baseline FMS at rest = 0)
Eye-Tracking Data	Left pupil diameter (mm) Right pupil diameter (mm) Left normalized gaze direction Right normalized gaze direction Convergence distance (mm)
Head-Tracking Data	Head Quaternion Rotation (i.e., x, y, z and w)
Physiological Data	Heart rate (HR) Electrodermal activity (EDA)

4.2 Data Pre-processing

The initial phase of our data preprocessing involved outlier detection. We employed a z-score analysis, excluding any data points that deviated more than three standard deviations from the mean [53]. Subsequently, to mitigate sensitivity to abrupt fluctuations, we applied a rolling average method with a window size of 5 [16]. This step effectively smoothed sudden noise, such as abrupt peaks and valleys in the dataset. Given the varied sampling rates across the data, a standardization to a 1Hz sampling rate was necessary. In doing so, we adopted an undersampling approach, ensuring no significant loss of information by computing the mean of the respective data samples within each period. The final preprocessing step entailed data normalization, which was achieved using the following formula:

$$I_n = \frac{I_n - \mu}{\sigma} \tag{1}$$

In this equation, I_n represents an individual data point (for n ranging from 1 to 30), where μ is the mean and σ signifies the standard deviation of the respective data. Note that all the data for each individual is time-synchronized, thus preserving the time dimension of the dataset.

4.3 Personalization of cybersickness Prediction

This section provides an elaboration of the four personalization strategies utilized in the study, as depicted in Figure 2. The succeeding subsections detail each of the four personalization methods employed in our methodology.

4.3.1 Cybersickness Personalization using Data Grouping

For the personalization of cybersickness prediction, there are many potential factors to group the data, such as age, gender, ethnicity, and such [2, 56]. In our model, we grouped our dataset by participants' age and gender, as these factors are already been integrated with the Simulation 2021 dataset.

Age: Previous research suggests that age is a significant factor in susceptibility to cybersickness [2,46]. In our study, we analyzed a dataset of 30 subjects stratified into two age groups based on the median age of 27 years. The **Group 1** included individuals aged 18–27 years (based on median split), with a mean age of 23.43 years (SD = 2.68), while **Group 2** comprised participants aged 28–60 years, averaging 36.23 years (SD = 10.43).

Gender: In order to evaluate the impact of gender on cybersickness prediction, we evenly split between 15 male and 15 female participants

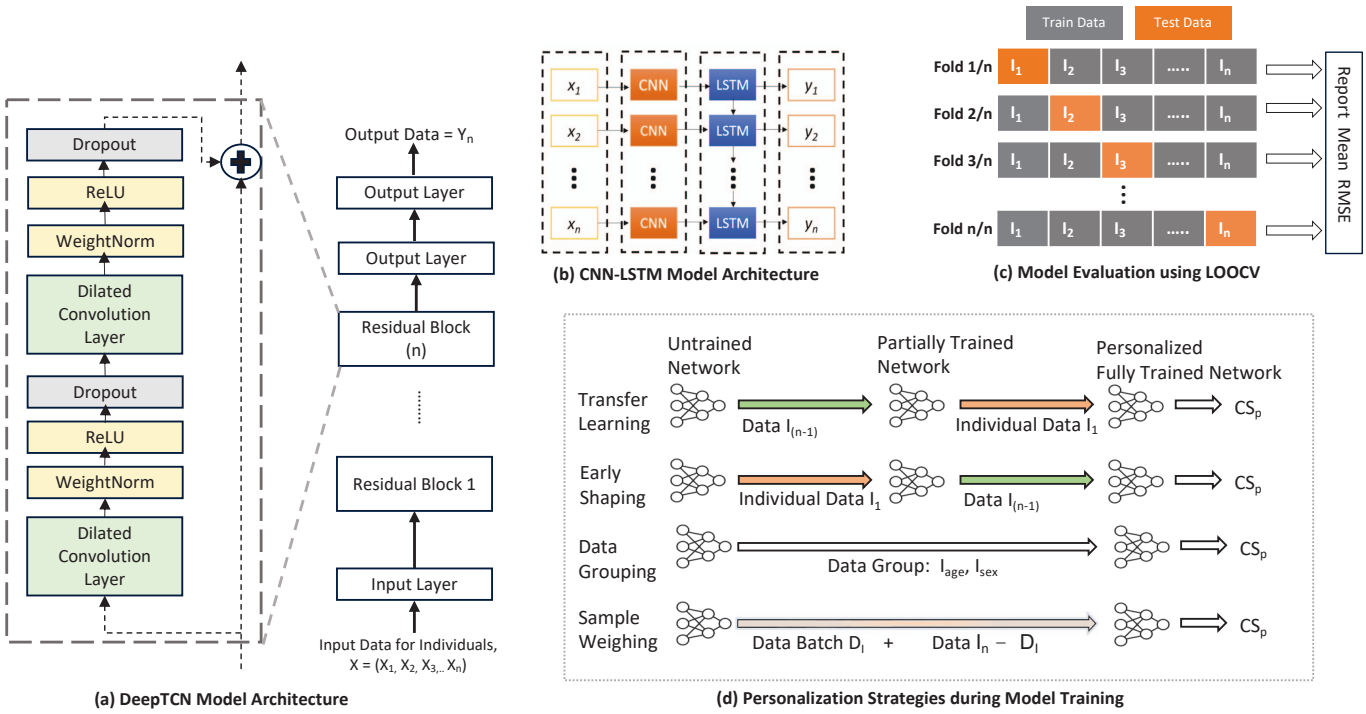

Fig. 2. This figure illustrates our approach to personalizing cybersickness prediction and evaluation using LOOCV (Leave-One-Out Cross-Validation). **(a)** Illustrates the DeepTCN Model Architecture, employing a Dilated Convolution Neural Network. **(b)** Depicts the CNN-LSTM model. **(c)** Explicates the model's evaluation and validation via the LOOCV approach, detailing the division of training and test data in each fold. **(d)** Highlights our personalization strategy, which incorporates Transfer Learning, Early Shaping, Data Grouping, and Sample Weighting, with individual data denoted by I_n.

[37, 56]. Additionally, we employed leave-one-out cross-validation to evaluate the predictive performance of our models.

4.3.2 Cybersickness Personalization using Transfer Learning

We utilized transfer learning to personalize cybersickness prediction. First we pre-trained the models using the dataset from Islam et al. [17] for $(n-1)$ individuals, denoted as I_{n-1}. This pre-training phase allowed the models to retain key features from the dataset applicable to $(n-1)$ individuals, as illustrated in Figure 2.d.

Subsequently, we modified the pre-trained models by removing their last "fully connected dense" layer and introducing a new, untrained layer of the same type. This modification was followed by re-training solely on this newly added layer, tailoring it to the data of the remaining individual not included in the pre-training phase (i.e., the residual one individual). Through this approach, the model leverages the comprehensive knowledge acquired during pre-training while focusing the re-training process on the unique characteristics of the residual individual, thus personalizing the features for a single individual.

We repeated this process for each individual in the dataset using a LOOCV strategy, thereby creating a personalized model for every participant. The effectiveness of our approach was quantitatively assessed by averaging the RMSE values obtained from each model, as depicted in Figure 2.c, which outlines our LOOCV validation methodology.

4.3.3 Cybersickness Personalization using Early-Shaping

We employed early shaping, a form of curriculum learning, to personalize the prediction of cybersickness as indicated in the dataset [17]. Early shaping, in the context of deep learning, involves progressively training models on tasks of increasing complexity. In our implementation, this technique began with the partial training of the model on data from a single individual, denoted as I_1 (See Figure 2.d). This step allowed the model to initially adapt to specific individual patterns, laying a foundational personalization of the cybersickness factors. Subsequently, we extended the training to incorporate the residual data from the remaining (n-1) participants, denoted as I_{n-1}. This step was crucial for the model to generalize its learning across different individual factors while retaining the nuanced understanding gained from the initial individual-specific training. By employing LOOCV validation techniques throughout this process, we ensured a robust evaluation of the model's performance, culminating in the reporting of the mean RMSE value (Figure 2.c).

4.3.4 Cybersickness Personalization using Sample Weighing

Finally, we utilized a sample weighting personalization technique. Initially, we created a training batch of size three, designated as D_I, optimized through hyper-parameter tuning. The model training involved data from both the $(I_n - D_I)$ group and the individual D_I dataset. This procedure increased the representation of the D_I individual's data within the training sample. This personalized training method was systematically applied to each subject using the LOOCV validating approach (Figure 2.c) and reported the mean RMSE values.

4.4 Model Architecture for Cybersickness Prediction

We have employed several models, such as DeepTCN LSTM, CNN-LSTM, and ARIMA, drawing inspiration from previous studies [17] for the cybersickness prediction task. These models were evaluated based on their mean RMSE values during the LOOCV validation process. The following sections will detail the architecture of each model used in our approach.

- **DeepTCN:** The Temporal Convolutional Network (TCN) is a type of deep learning model specifically designed for time-series data, as described in previous works [6, 33]. Our proposed Deep TCN model consists of five layers that apply filters across a sliding window of input data (X_n). Each TCN layer, featuring exponentially increasing dilations, is adept at capturing long-term

dependencies in the data. This model is enhanced with residual and skip connections and is tasked with predicting FMS values. In the construction of our DeepTCN, we utilized five residual blocks, adopted ReLU as the activation function, and implemented Weighted Batch Normalization. A dropout rate of 0.4 was chosen to prevent overfitting. The input and output of the model are denoted as X_n (i.e., HR, EDA, EYE) and Y_n(i.e. FMS), respectively.

- **LSTM:** Long Short-Term Memory (LSTM), a variant of Recurrent Neural Networks (RNN), is commonly utilized in cybersickness prediction, as evidenced in prior studies [34, 55]. In our model, we incorporated an input layer followed by two LSTM layers containing 60 and 120 neurons, respectively. The second LSTM layer includes a recurrent dropout rate of 0.2 to enhance model generalization. This is succeeded by a dense layer with 256 units and a subsequent dropout layer with a rate of 0.2 to prevent overfitting. Finally, we added a dense layer with ReLU as an activation function.

- **CNN-LSTM:** The CNN-LSTM model combines the strengths of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs) networks, forming a hybrid architecture as detailed in [18, 20]. Our proposed model starts with an input layer, followed by a 1D convolution layer equipped with 60 filters and a kernel size of 5. Subsequent to this, a max pooling layer with a pool size of 2 is applied. The architecture then includes two LSTM layers, each with 200 neural units, and a dropout layer with a rate of 0.15 to minimize overfitting. Additionally, a dense layer of 200 neural units with ReLU activation is included. The architecture concludes with a dense output layer containing ten units, also utilizing ReLU activation.

- **ARIMA:** The Autoregressive Integrated Moving Average (ARIMA) model, a renowned method for time-series forecasting, is extensively discussed in prior research [55, 64]. ARIMA is composed of three integral components: Autoregressive (AR), Integrated (I), and Moving Average (MA). The AR component establishes the relationship between a current observation and its preceding values or lags. The MA component, conversely, models the connection between an observation and the residual errors from applying a moving average model to past observations. In our implementation of the ARIMA model, we configured the AR parameter (p) as 30, representing the number of lag observations. The integrated parameter (d) was set to 1, indicating the level of differencing applied to achieve stationarity. Finally, the MA parameter (q) was set to 3, defining the size of the moving average window (adapted from Islam et. al. [17]).

4.5 Model Setup and Evaluation

We used TensorFlow-2.4 on an NVidia DGX-1 server with Ubuntu 18.0 to train and evaluate our deep-learning models. For the gradient decent optimization and learning, we used the Adam optimizer for 200 training epochs and batch sizes of 128. All the hyperparameters of the models during the training steps are fine-tuned with the Optuna framework. To train and validate our models, we used the Leave-One-Out Cross-Validation (LOOCV) strategy [50], which involves generating numerous training and testing sets by methodically excluding one observation from the dataset each time, training the model on the remaining data, and then testing it against the excluded observation. This approach allows for a thorough and rigorous evaluation compared to a random train/test split of the model's performance across all data points (Figure 2.c). To evaluate the prediction performance of our models, we employed the RMSE metric, as defined in Equation 2. The RMSE values were calculated using the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(FMS_t^i - \hat{FMS}_t^i)^2}{N}} \quad (2)$$

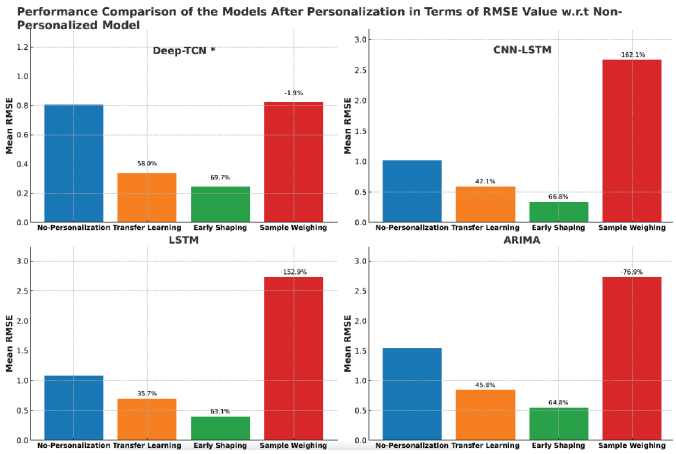

Fig. 3. Comparative Analysis of Deep Learning Models for Cybersickness Personalization: Evaluating Various Personalization Techniques Against Non-Personalized Baselines.

In this equation, FMS_t^i denotes the actual fast motion scale (FMS) values at time t for individual i, and $\hat{FMS}_t^i$ the predicted FMS values by the models.

5 Results

This section details the mean RMSE values obtained from the LOOCV evaluation, as shown in Tables 2 to 6. We present the RMSE values for various data fusion methods, such as eye-tracking (EYE), electrodermal activity (EDA), and heart rate (HR), in Tables 3 to 6. Notably, we excluded head-tracking data from our analysis, following the findings of Islam et al. [17], which indicated negligible improvement in results upon its inclusion. Initially, we compared the performance of DeepTCN, CNN-LSTM, LSTM, and ARIMA models against a non-personalized baseline (refer to Table 2 and Figure 3). Given that the DeepTCN model demonstrated superior performance compared to its counterparts, it was selected for all subsequent experiments. The following subsections provide an in-depth analysis of these results.

5.1 Model Performance with Transfer Learning, Early Shaping, and Sample Weighing

In this comprehensive analysis, we incorporated various data modalities, including eye-tracking (EYE), electrodermal activity (EDA), and heart rate (HR). We evaluated the performance of DeepTCN, CNN-LSTM, LSTM, and ARIMA models under different personalization strategies: transfer learning, early shaping, and sample weighing. The mean RMSE values from the leave-one-out cross-validation (LOOCV) evaluation are presented in Table 2 and Figure 3. The results indicated that the DeepTCN model significantly outperformed the others, with early shaping and transfer learning reducing the RMSE by 69.7% and 58.0%, respectively, when compared to the baseline non-personalized approach. Further details and insights on these RMSE improvements are provided in Figure 3.

The study further demonstrated that transfer learning and early shaping considerably enhanced cybersickness prediction across all tested models. Specifically, the DeepTCN model recorded mean RMSE values of 0.3383 with transfer learning and 0.2438 with early shaping. This contrasts with the higher RMSE values of 0.8205 (sample weighing) and 0.8054 (non-personalized) for the same model. For the CNN-LSTM models, non-personalized RMSE was 1.0167, which decreased to 0.5891 with transfer learning and 0.3371 with early shaping, representing improvements of 35.7% and 63.1%, respectively. However, the sample weighing approach did not demonstrate significant enhancement in prediction accuracy. Table 2 and Figure 3 detail this analysis and confirm the superior performance of the DeepTCN model. It was noted that, with the exception of sample weighing, personalization strategies generally led to improvements in RMSE values

Table 2. Comparison of mean RMSE between three personalization approaches with different deep models(i.e., Transfer Learning, Early Shaping and Sample Weighing)

Models	Without Personalization	Transfer Learning	Early Shaping	Sample Weighing
Deep-TCN *	**0.8054**	**0.3383**	**0.2438**	**0.8205**
CNN-LSTM	1.0167	0.5891	0.3371	2.6648
LSTM	1.0792	0.6936	0.3983	2.7295
ARIMA	1.5436	0.8362	0.5437	2.7309

across all evaluated models.

5.2 Model Performance on Data Grouping

In our data grouping analysis, we exclusively utilized the DeepTCN model, owing to its demonstrated superior performance relative to the CNN-LSTM, LSTM, and ARIMA models. We employed various fusion strategies for data grouping across different data modalities, including heart rate (HR), electrodermal activity (EDA), and eye tracking (EYE). The criteria for data grouping were based on median age splits and gender categories (male and female). The subsequent subsection provides a detailed account of the outcomes derived from this data grouping approach.

5.2.1 Grouping based on Age

Tables 3 and 4 detail the results of cybersickness prediction for different age groups, determined through a median age split (the median age being 27 years). We present the mean RMSE values obtained during the LOOCV validation process for various fusion modalities (namely EYE, EDA, HR) both with and without personalization across these age groups (Group 1 and Group 2).

Table 3 demonstrates that, for Group 1 (age 18-27), utilizing EYE tracking data in data grouping led to a decrease in the RMSE value to 0.6236, a reduction of 49.97% compared to the RMSE prior to grouping. However, the use of other fusion modalities resulted in increased RMSE values. When all modalities were fused, the personalized model exhibited a 5.51% improvement in RMSE over the model without age-based grouping.

Conversely, Table 4 indicates that for Group 2 (age 28-60), employing EYE tracking data in data grouping reduced the RMSE to 0.5883, a 52.80% decrease from the value before grouping. The fusion of all modalities in the personalized model yielded an 11.81% improvement in RMSE compared to models without age-based grouping. The performance improvement for age Group 2 was better compared to age Group 1.

5.2.2 Grouping based on Gender

Tables 5 and 6 present the cybersickness prediction results (i.e., FMS) before and after implementing gender-based grouping (male and female categories). The mean RMSE values obtained from the LOOCV validation process for each data modality (EYE, EDA, HR), both individually and in combination, are reported in these tables, highlighting the impact of gender-based data grouping.

For the male group, as shown in Table 5, the combination of all modalities post-grouping resulted in an RMSE of 0.4049. This represents a substantial decrease of 69.73% compared to the model without gender-based grouping. Furthermore, each individual modality also demonstrated a significant reduction in RMSE values following the grouping, with the EYE tracking data achieving an RMSE of 0.378.

In contrast, the female group, detailed in Table 6, exhibited significant improvements. Post-gender grouping, the combined modalities achieved an RMSE of 0.213, a significant reduction of 84.05% from the pre-grouping value. Specifically, the EYE tracking data reported an RMSE of 0.2422, reflecting an 80.57% decrease compared to its pre-grouping figure. Notably, the overall performance enhancements in the female group surpassed those in the male group across all modalities.

6 Discussion

The primary aim of our study was to investigate the effectiveness of personalization in enhancing cybersickness prediction performance. To achieve this, we implemented various personalization training strategies, including transfer learning, early shaping, and sample weighing. We also employed a data grouping method, organizing data based on participants' age and gender, and applied different data fusion techniques as outlined in Tables 3-6. Our analysis involved deep learning models such as DeepTCN, CNN-LSTM, LSTM, and ARIMA. These models underwent training and testing via the Leave-One-Out Cross-Validation (LOOCV) method, a more rigorous approach than the conventional random train/test split. We used mean RMSE values as our primary evaluation metric. Our findings indicate a significant improvement in cybersickness prediction when using the proposed personalization approaches, particularly with the DeepTCN model. This model demonstrated significant performance improvement in terms of RMSE values, specifically in the early-shaping (an improvement of 69.7%) and transfer learning (an improvement of 58.0%) personalization approaches. The following subsection discusses our key findings and potential limitations of our study.

6.1 Personalization with Transfer Learning, Early Shaping, Sample Weighing

Our findings, as detailed in Table 2 and Figure 3, demonstrate that both transfer learning and early shaping methods significantly enhanced cybersickness prediction across all models, outperforming the non-personalized counterparts. Notably, in the case of the DeepTCN model, the application of transfer learning resulted in a substantial 58% decrease in the RMSE value. This trend of improvement was also observed in other models, including CNN-LSTM, LSTM, and ARIMA, when employing early shaping and transfer learning techniques.

The superior performance of early shaping and transfer learning can be attributed to their inherent characteristics. Early shaping, by adjusting the model in its initial stages based on specific features of the dataset (I_1), enables a more tailored and effective learning process. This approach ensures that the model is better equipped to handle the nuances of the dataset(I_{n-1}) from the onset [3, 54]. Transfer learning, on the other hand, leverages pre-existing knowledge from related tasks or domains (pre-trained with I_{n-1}). This method allows models to bypass the initial learning phase, leading to faster convergence and improved prediction performance for a specific individual (I_1) [54,63]. These methods, thus, contribute to a more refined and accurate model training process, as evidenced by the enhanced RMSE values in our study. We hypothesize that the gradual introduction of participants to the training sample during early shaping may have contributed to the significantly better performance of the cybersickness prediction models as compared to other personalization approaches.

However, our analysis indicated that the sample-weighing approach failed to produce significant improvements in cybersickness prediction across all tested models. For instance, the DeepTCN model exhibited an increase of 1.9% in RMSE error compared to the non-personalized model. More notably, the CNN-LSTM model demonstrated a substantial 162.1% increase in RMSE error relative to its non-personalized counterpart. Several factors could explain this underperformance in RMSE with sample weighing. Primarily, sample weighing hinges on the premise of assigning differential weights to each data point (Data batch D_I) (Figure 2.d), presupposing the varying importance of each

Table 3. Mean RMSE Results for Age Group 1 (Ages 18-27) Using Median Split Approach. This table presents the calculated mean RMSE values derived from the leave-one-out cross-validation process.

Data Source	RMSE Without Grouping	RMSE with Grouping (personalized)	Improvement
EYE *	**1.2465**	**0.6236**	**49.97**%
HR	1.3098	1.2645	3.46%
EDA	1.6239	1.2520	22.90%
EYE + HR	1.3765	1.0665	22.52%
EYE + EDA	1.5792	1.0188	35.49%
HR + EDA	1.6256	1.2080	25.69%
EYE + HR + EDA	1.3378	1.2641	5.51%

Table 4. Mean RMSE Results for Age Group 2 (Ages 28-60) Using Median Split Approach. This table presents the calculated mean RMSE values derived from the leave-one-out cross-validation process.

Data Source	RMSE Without Grouping	RMSE with Grouping (personalized)	Improvement
EYE *	**1.2465**	**0.5883**	**52.80**%
HR	1.3098	1.0811	17.46%
EDA	1.6239	1.0793	33.54%
EYE + HR	1.3765	1.0196	25.93%
EYE + EDA	1.5792	1.0703	32.23%
HR + EDA	1.6256	1.2827	21.09%
EYE + HR + EDA	1.3378	1.1798	11.81%

observation. This approach might not be suitable for datasets where every observation contributes equally to model training. Misalignment in weight assignment can lead to issues such as model overfitting or underfitting, consequently degrading RMSE performance. Moreover, in situations where data is inherently balanced or embodies complex interrelationships, the sample weighing technique may not effectively enhance prediction accuracy, as suggested by [54]. The data used in our study likely did not benefit from this approach, leading towards the need for a more thorough analysis of data batch size and its impact on model performance when employing sample weighing.

6.2 Data grouping based on Age and Gender

Previous research, such as the study by Petri et al. [46], has highlighted age as a potential factor influencing the severity of cybersickness. In an effort to understand this relationship further, Arns et al. [2] conducted a study to examine whether the established correlation between age and motion sickness could be extrapolated to cybersickness. Contrary to the traditional view, their findings suggested that older individuals might experience more severe cybersickness compared to younger ones, thus challenging the conventional paradigm of age and motion sickness in the context of cybersickness.

Expanding upon these findings, our study sought to delve into the role of age in cybersickness severity, particularly by grouping participants based on age and analyzing the effectiveness of HR, EDA, and EYE data in predicting cybersickness. Our analysis revealed that incorporating EYE data notably enhanced the prediction performance for both age groups (Table 3 and 4). However, the addition of HR and EDA modalities did not yield a similar improvement, as indicated by the increased RMSE values. This underscores the need for further investigation into how various modalities influence cybersickness prediction and their interplay with age-related factors.

Regarding gender differences in cybersickness, existing literature, including works by MacArthur et al. [37], Stanney et al. [56], and Kelly et al. [21], has not established significant differences in FMS values between genders. Our research supports this finding, as we did not observe notable variations in FMS values across genders. However, we did find that gender-based data grouping significantly enhanced the accuracy of cybersickness prediction models. Our results indicated a significant reduction in RMSE values post-grouping in both male and female groups. Notably, in the female group, the application of all data modalities collectively led to an RMSE value of 0.2133, an 84.04% improvement over the non-personalized model, as shown in Table 6. These findings contribute to the evolving discourse on the role of gender in cybersickness and underscore the necessity of incorporating gender considerations in predictive models. Further investigation is warranted to better understand the underlying mechanisms of gender-based differences in cybersickness and to develop effective, personalized prediction models for different gender groups.

6.3 Limitations

While our study demonstrated significant performance improvements using early shaping and transfer learning across various machine learning and deep learning models for cybersickness prediction, it is important to acknowledge certain limitations. One key limitation is the range of models and data groupings explored. While we have investigated specific models and groupings, numerous other models and potential data groupings remain unexplored. Given our findings that certain groupings yield better results than others, further exploration into groupings based on participants' demographic backgrounds could be fruitful. However, this endeavor would require the collection of additional, diverse data. We remain optimistic that expanding our dataset will enhance the personalization and effectiveness of our models.

Additionally, our study's age distribution did not encompass individuals aged 70-90 years. The inclusion of a broader age range could potentially alter the results. For a more comprehensive evaluation of our proposed personalization techniques, utilizing more diverse datasets, encompassing a wider range of ages, is essential. Future research should aim to include participants from varied demographic backgrounds, including different gender identities (including non-binary), racial profiles, and a broader age spectrum. Such diversity in the participant pool is crucial to increase the generalizability and applicability of our findings in cybersickness prediction.

Although data was fed into the models sequentially based on time, we did not include time specifically as a feature in our models. Previous work has shown that cybersickness severity often increases over time but repeated exposure may lead to habituation [52]. Thus, including time as a feature in our models could affect results. However, several previous works in cybersickness prediction have also not used time as a feature [16–18]. In the present work, because we were focused on investigating personalization approaches compared to prior approaches without personalization, we chose to use the same feature set as prior work. In the future, we plan to investigate how including

Table 5. Mean RMSE Values for Male Participants Group. This table presents the results obtained from the Leave-One-Out Cross-Validation (LOOCV) method, specifically reporting the mean RMSE for the group of male participants.

Data Source	RMSE Without Grouping	RMSE with Grouping (personalized)	Improvement
EYE	1.2465	0.3785	69.63%
HR	1.3098	0.4617	64.75%
EDA	1.6239	0.6336	60.98%
EYE + HR	1.3765	0.5891	57.20%
EYE + EDA	1.5792	0.6230	60.55%
HR + EDA	1.6256	0.6535	59.80%
EYE + HR + EDA	1.3378	0.4049	**69.73**%

Table 6. Mean RMSE Values for Female Participants Group. This table presents the results obtained from the Leave-One-Out Cross-Validation (LOOCV) method, specifically reporting the mean RMSE for the group of Female participants.

Data Source	RMSE Without Grouping	RMSE with Grouping (personalized)	Improvement
EYE *	1.2465	0.2422	80.57%
HR	1.3098	0.5697	56.50%
EDA	1.6239	0.5888	63.74%
EYE + HR	1.3765	0.3722	72.96%
EYE + EDA	1.5792	0.2675	83.06%
HR + EDA	1.6256	0.4231	73.97%
EYE + HR + EDA	1.3378	0.2133	**84.05**%

time as a feature can affect cybersickness prediction results

7 Conclusion and Future Work

This research primarily sought to assess the potential of enhancing cybersickness prediction through personalized modeling. To realize this objective, we employed and evaluated a variety of grouping and training strategies typically used for model personalization. The majority of these methods yielded improved model performance, exemplified by the early shaping approach, which resulted in a 69.7% reduction in the RMSE value of the DeepTCN model, and transfer learning with DeepTCN, which diminished the RMSE by 58% compared to non-personalized models. Moreover, the significance of eye data for age-based and gender-based improvements was highlighted through our data grouping approach. Conclusively, our study identified early shaping and transfer learning as the most potent methods for personalizing cybersickness predictions. The overarching aim is to provide each individual with a model tailored to their specific data. In practice, this would necessitate users to engage with VR for a brief period, allowing for the fine-tuning of an existing aggregate model with their unique data via early shaping or transfer learning. Pre-trained gender-specific models would only need the user to specify their gender to select the appropriate model. Future research will explore more extensive data grouping, considering factors like cybersickness level and race, and examine the applicability of personalization strategies for cybersickness classification. Additionally, the effectiveness of other personalization techniques will be evaluated. The projected direction is to utilize these techniques to develop adaptive cybersickness mitigation systems, which hold critical implications for VR developers by potentially enabling more adaptable environments to reduce cybersickness incidences.

Acknowledgments

This work was funded through a generous gift from Intel Corporation and grants from the National Science Foundation (IIS-2007041, IIS-2211785).

References

[1] B. Arcioni, S. Palmisano, D. Apthorp, and J. Kim. Postural stability predicts the likelihood of cybersickness in active hmd-based virtual reality. *Displays*, 58:3–11, 2019. 2

[2] L. L. Arns and M. M. Cerney. The relationship between age and incidence of cybersickness among immersive environment users. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 267–268. IEEE, 2005. 4, 8

[3] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009. 3, 7

[4] P. Bockelman and D. Lingum. Factors of cybersickness. In *HCI International 2017–Posters' Extended Abstracts: 19th International Conference, HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part II 19*, pp. 3–8. Springer, 2017. 1

[5] E. Chang, H. T. Kim, and B. Yoo. Predicting cybersickness based on user's gaze behaviors in hmd-based virtual reality. *Journal of Computational Design and Engineering*, 8(2):728–739, 2021. 2, 4

[6] Y. Chen, Y. Kang, Y. Chen, and Z. Wang. Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, 399:491–501, 2020. 3, 5

[7] S. V. Cobb, S. Nichols, A. Ramsey, and J. R. Wilson. Virtual reality-induced symptoms and effects (vrise). *Presence: Teleoperators & Virtual Environments*, 8(2):169–186, 1999. 2

[8] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano. Deep learning and model personalization in sensor-based human activity recognition. *Journal of Reliable Intelligent Environments*, pp. 1–13, 2022. 3

[9] J. P. Freiwald, Y. Göbel, F. Mostajeran, and F. Steinicke. The cybersickness susceptibility questionnaire: predicting virtual reality tolerance. In *Proceedings of Mensch und Computer 2020*, pp. 115–118. 2020. 1

[10] J. M. Fulvio, M. Ji, and B. Rokers. Variations in visual sensitivity predict motion sickness in virtual reality. *Entertainment Computing*, 38:100423, 2021. 3

[11] M. Gallagher and E. R. Ferrè. Cybersickness: a multisensory integration perspective. *Multisensory research*, 31(7):645–674, 2018. 2

[12] A. Garcia-Agundez, C. Reuter, P. Caserman, R. Konrad, and S. Göbel. Identifying cybersickness through heart rate variability alterations. *International Journal of Virtual Reality*, 19(1):1–10, 2019. 2

[13] A. M. Gavgani, K. V. Nesbitt, K. L. Blackmore, and E. Nalivaiko. Profiling subjective symptoms and autonomic changes associated with cybersickness. *Autonomic Neuroscience*, 203:41–50, 2017. 2

[14] J. F. Golding. Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain research bulletin*, 47(5):507–516, 1998. 1

[15] R. Islam, S. Ang, and J. Quarles. Cybersense: A closed-loop framework to detect cybersickness severity and adaptively apply reduction techniques. In *2021 IEEE Conference on Virtual Reality and 3D User*

Interfaces Abstracts and Workshops (VRW), pp. 148–155. IEEE, 2021. 2

[16] R. Islam, K. Desai, and J. Quarles. Cybersickness prediction from integrated hmd's sensors: A multimodal deep fusion approach using eye-tracking and head-tracking data. In *2021 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 31–40. IEEE, 2021. 2, 3, 4, 8

[17] R. Islam, K. Desai, and J. Quarles. Towards forecasting the onset of cybersickness by fusing physiological, head-tracking and eye-tracking with multimodal deep fusion network. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 121–130. IEEE, 2022. 2, 3, 4, 5, 6, 8

[18] R. Islam, Y. Lee, M. Jaloli, I. Muhammad, D. Zhu, P. Rad, Y. Huang, and J. Quarles. 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)*, pp. 400–411. IEEE, 2020. 3, 6, 8

[19] D. Jeong, S. Yoo, and J. Yun. Cybersickness analysis with eeg using deep learning algorithms. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 827–835. IEEE, 2019. 2

[20] X. Jin, X. Yu, X. Wang, Y. Bai, T. Su, and J. Kong. Prediction for time series with cnn and lstm. In R. Wang, Z. Chen, W. Zhang, and Q. Zhu, eds., *Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019)*, pp. 631–641. Springer Singapore, Singapore, 2020. 3, 6

[21] J. W. Kelly, S. B. Gilbert, M. C. Dorneich, and K. A. Costabile. Gender differences in cybersickness: Clarifying confusion and identifying paths forward. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 283–288. IEEE, 2023. 8

[22] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993. doi: 10.1207/s15327108ijap0303
3 1, 2

[23] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993. 2

[24] B. Keshavarz and J. F. Golding. Motion sickness: current concepts and management. *Current Opinion in Neurology*, 35(1):107–112, 2022. 2

[25] B. Keshavarz and H. Hecht. Validating an efficient method to quantify motion sickness. *Human factors*, 53(4):415–426, 2011. 2

[26] H. K. Kim, J. Park, Y. Choi, and M. Choe. Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied ergonomics*, 69:66–73, 2018. 1, 2

[27] J. Kim, W. Kim, H. Oh, S. Lee, and S. Lee. A deep cybersickness predictor based on brain signal analysis for virtual reality contents. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10580–10589, 2019. 2

[28] Y. Y. Kim, H. J. Kim, E. N. Kim, H. D. Ko, and H. T. Kim. Characteristic changes in the physiological components of cybersickness. *Psychophysiology*, 42(5):616–625, 2005. 2

[29] S. Klosterhalfen, S. Kellermann, F. Pan, U. STOcKHORST, G. Hall, and P. ENcK. Effects of ethnicity and gender on motion sickness susceptibility. *Aviation, space, and environmental medicine*, 76(11):1051–1057, 2005. 1

[30] P. Kourtesis, R. Amir, J. Linnell, F. Argelaguet, and S. E. MacPherson. Cybersickness, cognition, & motor skills: The effects of music, gender, and gaming experience. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2326–2336, 2023. 2

[31] R. K. Kundu, R. Islam, P. Calyam, and K. A. Hoque. Truvr: Trustworthy cybersickness detection using explainable machine learning. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 777–786. IEEE, 2022. 2

[32] J. J. LaViola Jr. A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin*, 32(1):47–56, 2000. 2

[33] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 156–165, 2017. 3, 5

[34] T. M. Lee, J.-C. Yoon, and I.-K. Lee. Motion sickness prediction in stereoscopic videos using 3d convolutional neural networks. *IEEE transactions on visualization and computer graphics*, 25(5):1919–1927, 2019. 2, 6

[35] S. Litleskare. The relationship between postural stability and cybersickness: it's complicated–an experimental trial assessing practical implications of cybersickness etiology. *Physiology & Behavior*, 236:113422, 2021. 2

[36] T. Luong, A. Pléchata, M. Möbus, M. Atchapero, R. Böhm, G. Makransky, and C. Holz. Demographic and behavioral correlates of cybersickness: A large lab-in-the-field study of 837 participants. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 307–316, 2022. doi: 10.1109/ISMAR55827.2022.00046 3

[37] C. MacArthur, A. Grinberg, D. Harley, and M. Hancock. You're making me sick: A systematic review of how virtual reality research considers gender & cybersickness. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2021. 1, 5, 8

[38] N. Martin, N. Mathieu, N. Pallamin, M. Ragot, and J.-M. Diverrez. Virtual reality sickness detection: an approach based on physiological signals and machine learning. In *2020 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 387–399. IEEE, 2020. 2, 4

[39] A. J. Martingano, E. Brown, S. H. Telaak, A. P. Dolwick, and S. Persky. Cybersickness variability by race: Findings from 6 studies and a mini meta-analysis. *Journal of Medical Internet Research*, 24(6):e36843, 2022. 3

[40] M. Melo, G. Gonçalves, D. Narciso, and M. Bessa. Impact of different role types and gender on presence and cybersickness in immersive virtual reality setups. In *2021 international conference on graphics and interaction (ICGI)*, pp. 1–8. IEEE, 2021. 3

[41] D. Monteiro, H.-N. Liang, X. Tang, and P. Irani. Using trajectory compression rate to predict changes in cybersickness in virtual reality games. In *2021 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 138–146. IEEE, 2021. 2

[42] J. Munafo, M. Diedrick, and T. A. Stoffregen. The virtual reality head-mounted display oculus rift induces motion sickness and is sexist in its effects. *Experimental brain research*, 235:889–901, 2017. 1

[43] Y. Nam, U. Hong, H. Chung, and S. R. Noh. Eye movement patterns reflecting cybersickness: evidence from different experience modes of a virtual reality game. *Cyberpsychology, Behavior, and Social Networking*, 25(2):135–139, 2022. 2

[44] S. A. A. Naqvi, N. Badruddin, M. A. Jatoi, A. S. Malik, W. Hazabbah, and B. Abdullah. Eeg based time and frequency dynamics analysis of visually induced motion sickness (vims). *Australasian physical & engineering sciences in medicine*, 38:721–729, 2015. 2

[45] H. Oh and W. Son. Cybersickness and its severity arising from virtual reality content: A comprehensive study. *Sensors*, 22(4), 2022. doi: 10.3390/s22041314 2

[46] K. Petri, K. Feuerstein, S. Folster, F. Bariszlovich, and K. Witte. Effects of age, gender, familiarity with the content, and exposure time on cybersickness in immersive head-mounted display based virtual reality. *American Journal of Biomedical Sciences*, 12(2), 2020. 4, 8

[47] K. M. Pöhlmann, G. Li, M. McGill, F. Pollick, and S. Brewster. Can gender and motion sickness susceptibility predict cybersickness in vr? In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 277–282. IEEE, 2023. 3

[48] L. Rebenitsch and C. Owen. Individual variation in susceptibility to cybersickness. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 309–317, 2014. 1

[49] G. R. Reddy, C. A. Spencer, K. Durkee, B. Cox, O. Fox Cotton, S. Galbreath, S. Meyer, M. Natali, T. Seech, G. Severe-Valsaint, et al. Estimating cognitive load and cybersickness of pilots in vr simulations via unobtrusive physiological sensors. In *Virtual, Augmented and Mixed Reality: Applications in Education, Aviation and Industry: 14th International Conference, VAMR 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part II*, pp. 251–269. Springer, 2022. 2

[50] J. D. Rennie. On the value of leave-one-out cross-validation bounds. *Computers in Biology and Medicine*, pp. 123–129, 2003. 6

[51] J. J. Reyes-Lagos, J. C. Echeverría-Arjonilla, M. Á. Peña-Castillo, M. T. García-González, M. del Rocio Ortiz-Pedroza, G. Pacheco-Lopez, C. Vargas-Garcia, S. Camal-Ugarte, and R. Gonzalez-Camarena. A comparison of heart rate variability in women at the third trimester of pregnancy and during low-risk labour. *Physiology & behavior*, 149:255–261, 2015. 2

[52] D. Risi and S. Palmisano. Effects of postural stability, active control, exposure duration and repeated exposures on hmd induced cybersickness. *Displays*, 60:9–17, 2019. 8

[53] S. Saleem, M. Aslam, and M. R. Shaukat. A review and empirical comparison of univariate outlier detection methods. *Pakistan Journal of Statistics*, 37(4), 2021. 4

[54] J. Schneider and M. Vlachos. Personalization of deep learning. In *Data Science–Analytics and Applications: Proceedings of the 3rd International Data Science Conference–iDSC2020*, pp. 89–96. Springer, 2021. 2, 3, 4, 7, 8

[55] S. Siami-Namini, N. Tavakoli, and A. S. Namin. A comparison of arima and lstm in forecasting time series. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pp. 1394–1401. IEEE, 2018. 6

[56] K. Stanney, C. Fidopiastis, and L. Foster. Virtual reality is sexist: but it does not have to be. *Frontiers in Robotics and AI*, 7:4, 2020. 4, 5, 8

[57] K. Stanney, B. D. Lawson, B. Rokers, M. Dennison, C. Fidopiastis, T. Stoffregen, S. Weech, and J. M. Fulvio. Identifying causes of and solutions for cybersickness in immersive technology: reformulation of a research and development agenda. *International Journal of Human–Computer Interaction*, 36(19):1783–1803, 2020. 2, 3

[58] K. M. Stanney, R. S. Kennedy, and J. M. Drexler. Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, vol. 41, pp. 1138–1142. SAGE Publications Sage CA: Los Angeles, CA, 1997. 2

[59] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Latency and cybersickness: Impact, causes, and measures. a review. *Frontiers in Virtual Reality*, 1:582204, 2020. 2

[60] M. Tektaş. Weather forecasting using anfis and arima models. *Environmental Research, Engineering and Management*, 51(1):5–10, 2010. 3

[61] N. Tian, P. Lopes, and R. Boulic. A review of cybersickness in head-mounted displays: raising attention to individual susceptibility. *Virtual Reality*, 26(4):1409–1441, 2022. 1, 2, 3

[62] Y. Wang, J.-R. Chardonnet, F. Merienne, and J. Ovtcharova. Using fuzzy logic to involve individual differences for predicting cybersickness during vr navigation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 373–381, 2021. doi: 10.1109/VR50410.2021.00060 2

[63] K. Weiss, T. M. Khoshgoftaar, and D. Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016. 3, 7

[64] G. P. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003. 6