LiteVR：借助可解释人工智能实现可解释、轻量级的晕动症检测

Ripan Kumar Kundu * Rifatul Islam † John Quarles ‡ Khaza Anuarul Hoque §密苏里大学 肯尼索州立大学 圣安东尼奥德克萨斯大学 密苏里大学

arXiv:2302.03037v1 [cs.HC] 5 Feb 2023

摘要

网络晕动症是虚拟现实（VR）用户体验中常见的不适症状。目前已有多种基于机器学习（ML）和深度学习（DL）的自动化方法用于检测网络晕动症。然而，人们普遍认为，这些晕动症检测方法大多计算开销大，且属于黑箱方法。因此，这些技术既不可信，也不适合部署在能耗受限的独立式 VR 头戴式设备（HMD）上。本文提出一种基于可解释人工智能（XAI）的晕动症检测框架 *LiteVR*，可解释模型的输出结果，并减少特征维度和总体计算开销。首先，我们基于长短期记忆网络（LSTM）、门控循环单元（GRU）和多层感知机（MLP）开发了三种网络晕动症深度学习模型。随后，我们采用 SHapley Additive Explanations（SHAP）等事后解释方法解释结果，并提取网络晕动症最具主导性的特征。最后，我们使用数量减少后的特征重新训练深度学习模型。结果表明，眼动追踪特征是晕动症检测中最具主导性的特征。此外，基于 XAI 的特征排序和降维，我们将模型大小最多减少了 4.3×，训练时间最多缩短了 5.6×，推理时间最多缩短了 3.8×，同时提高了晕动症检测准确率，并降低了回归误差（即在快速运动量表（FMS）上的误差）。我们提出的轻量级 LSTM 模型在晕动症分类和回归（即 FMS 1–10）任务中准确率达到 94%，均方根误差（RMSE）为 0.30，性能优于最先进水平。我们提出的 *LiteVR* 框架可帮助研究人员和从业者分析、检测晕动症，并将其基于深度学习的检测模型部署到独立式 VR HMD 中。

关键词： 虚拟现实、晕动症检测、可解释人工智能、深度学习、模型简化

索引词： 以人为中心的计算——人机交互（HCI）——交互范式——虚拟现实；以人为中心的计算——人机交互（HCI）——HCI 设计与评估方法。

1 引言

虚拟现实（VR）的应用十分广泛，并且在医疗保健 [19,56]，教育 [2]，军事训练 [82]，外科手术训练 [48]，以及灾害管理 / 公共安全 [16,51]等领域迅速增长。然而，VR 用户经常会出现虚拟现实晕动症，也称网络晕动症，从而影响其沉浸式体验。因此，网络晕动症已成为 [49]妨碍 VR 得到更广泛认可的重要障碍。网络晕动症可定义为一系列令人不适的症状，例如眼疲劳、头痛，恶心、方向感障碍，甚至呕吐 [59]。检测网络晕动症的一种常用方法，是在沉浸体验结束后使用主观问卷，例如模拟器晕动症问卷（SSQ）和虚拟现实晕动症问卷（VRSQ） [66]。相比之下，快速晕动症量表（FMS） [32] 可在沉浸过程中评估晕动症严重程度。请注意，FMS 依赖用户反馈，也就是说，在 VR 沉浸过程中需要人工介入 [47]。为克服这些局限，深度学习（DL）和机器学习（ML）近年来已成为检测网络晕动症的热门方法 [1,21,22,24,27,30,58]。

最先进水平的深度学习模型能够以较高准确率检测网络晕动症。例如，Islam 等人 [24] 提出了一种利用深度融合网络进行网络晕动症严重程度检测的方法，通过用户的眼动追踪和头部追踪数据，准确率达到 87.7%。其他研究人员则使用脑电图（EEG） [58,72,83]，立体视频 [24,52] 和生理信号 [21,58] 数据，以更高的准确率检测网络晕动症。尽管深度学习模型在检测网络晕动症方面具有良好前景，但这些方法仍存在若干局限性：

- 大多数基于深度学习的网络晕动症检测模型依赖黑箱模型，因此缺乏可解释性。深度学习模型的可解释性能够显著提升对模型的理解，并深入揭示深度学习模型为何以及如何得出特定决策。识别并理解导致网络晕动症的重要特征，有助于设计人员开发更有效的网络晕动症检测模型。

- 头戴式显示器的集成传感器以及外部生理传感器（脑电图、皮肤电反应、心率）生成的多模态数据，会形成复杂的数据集，并需要耗能较大的深度学习模型来进行晕动症检测。因此，将这些模型部署在独立式能源受限头戴式显示器（例如 Meta Quest Pro）上往往颇具挑战，甚至难以实现。通过识别主导特征，可以设计和开发轻量级深度学习模型，从而降低计算成本、缩短推理时间并降低模型复杂度。

为说明所提出的可解释人工智能方法的必要性，可以设想一位虚拟现实游戏开发者想要基于使用 Meta Quest Pro 头戴式显示器的 Alpha/Beta 测试玩家，为其游戏开发晕动症预测模型。如果开发者在游戏中使用的是缺乏可解释性的黑箱晕动症预测模型，就很难确定模型的哪些特征（例如瞳孔直径）对晕动症预测产生了影响。因此，他们只能通过反复试错来缩小模型大小，以尽量减少对资源本就有限的 Meta Quest Pro 的资源消耗。

针对这些研究空白，相关研究仍然有限。近期，Kundu 等人 [38] 使用具有内在可解释性的机器学习模型，结合生理 [25] 和游戏过程数据集 [57]进行晕动症检测与解释。然而，这项工作采用了二元分类器，只能检测是否存在网络晕动症。这种用于网络晕动症检测的二元分类方法不太适用于现实的虚拟现实应用。

*电子邮件：rkcgc@umsystem.edu

†电子邮件：rislam11@kennesaw.edu

‡电子邮件：john.quarles@utsa.edu §电子邮件：hoquek@umsystem.edu

此外，具有内在可解释性的模型通常依赖于数据属性，因此会受到维度问题的影响 [64]。例如，基于决策树的内在可解释模型可能会因树结构过于复杂而出现过拟合，无法很好地泛化数据，最终可能导致其在网络晕动症分类中的表现不佳 [54,58]。

其他一些研究采用特征选择或降维技术（如主成分分析（PCA）） [34,36,45,63,70] ，以解决深度学习模型用于晕动症检测时的高维问题。例如，Mawalid 等人在 [46] 中采用时域特征提取方法，提取脑电图的统计特征以对网络晕动症进行分类。类似地，Lin 等人 [41]将机器学习模型与主成分分析结合，提取晕动症相关特征，以预测网络晕动症程度。此外，Kottaimalai 等人在 [37] 中使用主成分分析降低脑电信号的维度、复杂度和计算时间，从而检测网络晕动症。需要注意的是，主成分分析旨在尽可能覆盖数据集中各特征空间的方差。如果未能谨慎选择数据集中的主成分数量，与原始特征集相比，就可能遗漏某些信息 [3]。因此，采用基于主成分分析的降维方法可能会丢失对准确晕动症检测至关重要的特征。

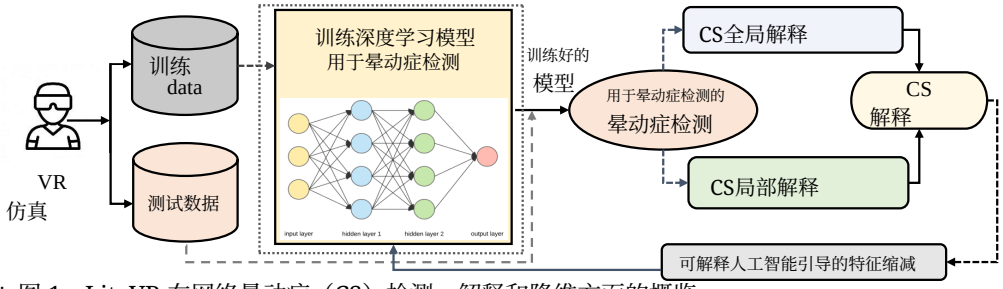
为应对上述挑战，我们提出了一种新颖的方法 LiteVR——一种用于晕动症检测的基于可解释人工智能的框架，用于解释和缩减特征规模（也称为降维）。具体而言，我们首先基于长短期记忆网络（LSTM）、门控循环单元（GRU）和多层感知机（MLP），开发了三种用于晕动症检测的深度学习模型。随后，我们采用事后解释技术 SHapley Additive Explanations (SHAP) [44], ，提供全局和局部解释，以分析、识别并排序导致网络晕动症的主导特征。然后，我们利用已识别的主导特征重新训练模型（即使用较少的特征进行训练，而非使用全部特征）。由此得到的晕动症检测模型易于使用，且可训练参数数量大幅减少。在保持基准准确率的同时，这使训练和推理速度更快。例如，我们的结果表明，使用全部特征的 LSTM 模型可将网络晕动症严重程度分为 4 个等级（无、低、中和高），准确率为 92%；并可对当前网络晕动症进行回归预测（FMS 1–10），均方根误差（RMSE）为 0.36。然而，对同一 LSTM 模型进行基于可解释人工智能的特征缩减（即仅使用基准模型三分之一的特征）后，我们分别以 94% 的准确率对网络晕动症严重程度进行分类，并以 0.30 的 RMSE 预测当前网络晕动症。此外，我们基于可解释人工智能的模型简化方法使模型大小最多缩减 4.3× （LSTM），训练和推理时间分别最多提升 5.5× 和 3.8× 。因此，我们认为，所提出的方法可帮助未来研究人员理解、分析和设计适用于独立式头戴显示器实时晕动症检测的模型。据我们所知，这是首项将可解释人工智能用于解释基于深度学习的晕动症检测模型并缩减其特征规模的研究。

2 相关工作

解释网络晕动症成因的主流理论是感觉冲突理论 [39]。该理论认为，当视觉感觉系统感知到运动，而前庭系统却没有感知到时，就会发生网络晕动症。不过，其他理论，例如中毒理论和姿势不稳定理论，甚至网络安全，也被认为是导致网络晕动症的原因 [17,39,74,75]。此外，用户的年龄、性别和过往虚拟现实体验等因素也会影响网络晕动症的程度 [12,14,20]。该晕动症检测领域的最先进研究大致可分为两类，分别采用主观测量和客观测量。近年来，采用先进机器学习和深度学习方法进行自动化晕动症检测越来越受欢迎，因为这类方法无需人工干预。本节将重点介绍晕动症检测领域的最先进研究。

为了测量网络晕动症，研究人员提出了多种主观测量方法，例如模拟器晕动症问卷（SSQ） [6,7,9,13,50,66,78,84,85],、FMS [32],和晕动症易感性问卷（MSSQ） [31]。相比之下，一些研究人员还提出了用于网络晕动症 [21,25,57] 检测的客观测量方法（即生理信号）。既有研究表明，当发生网络晕动症时，客观测量指标（例如心率、皮肤电反应和脑电图（EEG）信号）会发生显著变化 [11,25,42,42,60,62]。例如，研究发现，HR和脑电图δ波与网络晕动症呈正相关，而脑电图β波与网络晕动症呈负相关 [42]。相反，另一项研究报告称，与其他能够检测网络晕动症的客观测量指标相比，GSR与网络晕动症呈现更强的正相关性 [25,71]。

为了从头戴式显示器中的多种客观测量（生理信号）数据、主观测量（FMSQ、SSQ）数据以及集成传感器测量（眼动追踪、头部追踪、运动流等）数据中自动检测网络晕动症，近期已有大量基于机器学习和深度学习的方法被提出 [1,18,21,24–26,33,53,58,61,62,77,80,81]。例如，在 [1], 中，采用机器–深度–集成学习方法对网络晕动症进行分类。相比之下，Porcino 等人 [57] 采用基于符号机器学习的方法识别网络晕动症的程度。近年来，深度学习日益受到网络晕动症研究人员的关注。例如，Qu 等人 [58] 采用基于长短期记忆网络的注意力网络，利用生物生理信号检测网络晕动症。另一方面，Jin 等人 [30] 使用了三种深度学习/机器学习算法：卷积神经网络（CNN）、长短期记忆网络和支持向量机（SVM）来检测网络晕动症。同样，Padmanaban 等人 [52] 利用虚拟现实视频数据中的深度和光流特征预测网络晕动症。相比之下， [40] 中的作者采用三维卷积神经网络和多模态深度融合方法，结合光流、视差和显著性特征；与 Padmanaban 等人 [52]相比，他们报告了更高的晕动症检测准确率。因此， [25], 中的作者使用长短期记忆网络模型，根据用户的生理信号（如心率、皮肤电反应等）对晕动症严重程度进行分类。此外， [24]提出了一种深度融合方法，用于对晕动症严重程度进行分类，从头戴式显示器的集成传感器数据中预测网络晕动症。另一方面，Jeong 等人 [26] 采用基于注意力机制的深度学习模型，从多模态传感器数据中检测网络晕动症。尽管深度学习模型在晕动症检测方面取得了巨大成功，但将可解释人工智能应用于网络晕动症模型解释仍存在显著的研究空白。事实上，了解网络晕动症模型中特征的重要性，并解释样本为何被正确或错误地标记为网络晕动症，是采取适当缓解措施的重要一步。此外，在网络晕动症模型中应用可解释人工智能，可以显著提高模型的可信度，并深入揭示深度学习模型作出特定决策的原因及方式。在本文中，“可信”是指通过一系列机制（如全局和局部解释以及可解释层）运用可解释人工智能来解释网络晕动症，从而使模型透明、易于理解，并因此获得用户信任 [73]。最近，Kundu 等人 [38] 使用了本质上可解释的、基于机器学习的模型来解释网络晕动症。他们使用可解释提升机（EBM）、决策树（DT）和逻辑回归（LR），根据用户的生理信号和主观测量信号检测并解释网络晕动症。但遗憾的是，他们提出的方法无法解释基于深度学习的晕动症检测模型。



Fig图 1：LiteVR 在网络晕动症（CS）检测、解释和降维方面的概览。

相比之下，我们在本研究中采用基于 SHAP 的事后解释方法，解释基于深度学习的晕动症检测模型。 [38] 中的另一项局限是，他们的机器学习模型仅限于二分类（网络晕动症与无网络晕动症）。相比之下，我们研究的是网络晕动症检测中的多分类问题（无、低、中、高）。

目前已有多项关于网络晕动症模型降维的研究。一些研究人员使用基于主成分分析的方法提取重要特征并缩小模型大小 [34,36,45,63,68,70] [41] 。例如，文献中的作者使用主成分分析从脑电信号中提取网络晕动症相关特征。随后，他们使用 3 种机器学习/深度学习模型：线性回归（LR）、支持向量机和自组织神经模糊推理网络（SOFIN），来预测用户的网络晕动症程度。同样，Kottaimalai 等人在 [37]中也使用主成分分析从脑电信号中发现模式，并使用神经网络（NN）基于科罗拉多大学脑电信号数据集对认知任务进行分类。相比之下，Singla 等人在 [68] 中使用主成分分析缩减了 SSQ 仿真中的问题集。另一方面，Mawalid 等人在 [46] 中使用基于统计特征（例如均值、变化量、标准差、峰值数量）和功率百分比频带的时域特征提取方法，以了解网络晕动症特征，随后应用 K 近邻和朴素贝叶斯分类器对网络晕动症进行分类。然而，基于主成分分析的降维并非总是可靠。这是因为主成分分析通过投影将高维数据映射到低维空间，这往往会导致原始数据中的信息丢失 [29]。与此不同，我们采用一种可靠的方法，即基于可解释人工智能的方法，对晕动症检测模型进行降维。

3 提出的 LiteVR 框架

图 1 展示了所提出的 LiteVR 框架概览，该框架用于虚拟现实网络晕动症的检测、解释和模型简化。首先，使用训练数据 [24] 训练深度学习模型（即长短期记忆网络、门控循环单元和多层感知机），用于分类和回归任务 [1]。接下来，训练好的深度学习模型通过将测试数据集划分为四个类别——无、低、中和高，来检测网络晕动症。训练好的深度学习模型还可以根据 $0-10$ 的 FMS 评分范围，预测当前网络晕动症的下一个值。随后，我们使用最先进的工具 SHAP 来解释网络晕动症特征，从而提供局部和全局解释。全局解释用于识别对整体预测至关重要的特征，而局部解释则用于识别主导单个样本预测的特征。在解释网络晕动症特征后，我们使用全局解释来识别并排列起主导作用的特征。最后，在模型简化阶段，我们仅使用最具主导性的特征（在本研究中为排名前 30% 的特征）重新训练深度学习模型，以简化网络晕动症模型。由此得到的模型显著更小，可训练参数更少，因此训练和推理时间更短。

[1]数据集 [24]: https://tinyurl.com/2p92p45h

3.1 晕动症分类

基于长短期记忆网络、门控循环单元和多层感知机的网络晕动症分类模型，其网络架构分别如表 5.3、表 2 和表 3 所示。长短期记忆网络和门控循环单元模型均有 6 层，而多层感知机模型有 8 层。我们选择长短期记忆网络和门控循环单元模型，是因为它们能够捕捉数据集的时间特征和空间特征。对于长短期记忆网络和门控循环单元模型，我们都设置了 60 个时间步，并使用 43 个特征。第一层长短期记忆网络的循环丢弃率为 20%，之后接入一个丢弃率同为 20% 的层，以应对过拟合。在下一个长短期记忆网络层中，我们将丢弃率最多降低至 5%，因为较高的丢弃率会减慢模型的收敛速度，并且常常损害最终性能 [79]。另一方面，在门控循环单元模型中，所有门控循环单元层都使用 30% 的循环丢弃率以减少过拟合。对于长短期记忆网络和门控循环单元模型，我们在第五个全连接层中使用 ReLU 作为激活函数。最后，我们在最后一个全连接层中使用 Softmax 作为激活函数，该层包含四个输出，对应四个网络晕动症类别。由于多层感知机的 8 层均为全连接层，因此它不像长短期记忆网络和门控循环单元模型那样使用时间步。我们在每个全连接层中使用 ReLU 作为激活函数，并设置 25% 的递归丢弃率。与门控循环单元和长短期记忆网络模型类似，多层感知机的最后一个全连接层使用 Softmax 激活函数，以识别四种网络晕动症严重程度类别。如公式 1 所示，我们在深度学习分类模型中使用分类交叉熵作为损失函数。

$$\mathscr{L}(y,\hat{y}) = -\sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij}\log(\hat{y}_{ij})) \tag{1}$$

其中，$\mathscr{L}$表示损失函数，$\hat{}yij$ 表示预测的晕动症严重程度类别标签，(yij)表示实际的晕动症严重程度类别标签。此外，p 表示晕动症类别的数量，q 表示分类器的训练样本总数。

3.2 晕动症回归

晕动症回归任务可定义如下：给定历史观测到的虚拟现实数据（例如眼动追踪数据、头部追踪数据等）以及先前时间步$t-1$的FMS评分，预测下一个时间步t的FMS评分。我们将时间t的晕动症FMS评分记为CS_t。例如，如果我们预测$t=10$秒时的晕动症FMS评分，那么$CS_t \rightarrow [S_{t-9}, S_{t-8}, S_{t-7}, S_{t-6}\ldots, S_t]$，其中$S$表示用户的认知状态。对于晕动症回归任务，我们也使用与分类任务相同的三种深度学习模型，但损失函数除外（回归任务使用RMSE），即预测取值范围为0到10的晕动症FMS评分。对于回归模型，我们使用均方根误差（RMSE）作为损失函数，定义见公式2。

$$RMSE = \sqrt{\frac{1}{|N|}\sum_{y\in S}\sum_{t=1}^{N}(y_t-\hat{y}_t)^2} \tag{2}$$

其中，y_t 为输入，例如用户在t时口头报告的 FMS 评分；$\hat{y}_t$ 为模型在该时刻预测的 FMS 评分。因此，N 和 S 分别表示样本总数和时间步数。

算法 1： 基于 SHAP 的网络晕动症特征缩减

输入 ：网络晕动症深度学习模型 θ，训练数据集X_{Train}，测试数据集X_{Test}

输出 ：$F_{rank} = \{\}$

1: **对每个** 特征 $i \in X_{train}$ **执行**
2: Ex_{val} = 模型解释器(θ, X_{Train})
3: F^{imp} = SHAP值(Ex_{val},X_{Test})
4: $F_{rank} = F_{rank} \bigcup F^{imp}$
5: </style结束于
6: **返回**F_{rank}

3.3 使用 SHAP 解释深度学习模型

在本研究中，我们采用基于SHAP的事后解释技术来解释晕动症结果。SHAP是一种特征重要性解释方法，通过为每个预测分配一个特征显著性值来发挥作用。它基于合作博弈论中的沙普利值这一数学基础 [44]。对于给定的一组输入样本（例如眼动追踪数据、头部追踪数据等）和深度学习模型（例如长短期记忆网络、门控循环单元和多层感知机），SHAP旨在通过计算每个特征对预测的贡献，解释输入样本的预测结果。换言之，SHAP值可以解释给定的特征值如何提高或降低模型预测。例如，对于一个预测样本是否患有晕动症的晕动症预测模型，SHAP解释可以让我们了解某个特征在多大程度上促成了特定预测。此类解释可以是全局的，也可以是局部的。在SHAP中，可以用条形图展示晕动症总体特征重要性排序（全局解释）。相反，在局部解释中，从包含全部特征的测试数据集中随机选取每个样本，并确定哪些特征会提高或降低发生晕动症的可能性。我们使用*Deep SHAP* 方法 [67] 计算基于深度学习的晕动症模型的SHAP值。随后，利用这些SHAP值计算晕动症严重程度预测过程中的特征重要性。

3.4 通过基于 SHAP 的解释进行特征缩减

算法 1 展示了使用 SHAP 进行特征缩减的整体方法。该算法以晕动症检测模型、训练和测试数据集为输入，并返回特征排序。对于数据集中的每个特征，给定模型 θ，算法使用 ModelExplainer() 函数获取期望值 Ex_{val} （第 1–2 行）。ModelExplainer() 函数使用 X_{Train} 个样本 [44]近似计算 SHAP 值的条件期望。接下来，算法使用 SHAPvalue() 函数计算每个特征的特征重要性分数 F^{imp} （第 3 行），然后将其追加到集合 F_{rank} 中（第 4 行）。SHAPvalue() 函数通过计算每个特征的平均值来计算 F^{imp} 值。最后，在对所有特征完成这一过程后，算法返回集合 F_{rank} （第 6 行）。根据排序后的特征 F_{rank}，我们按照重要性对它们进行排序（降序排列，即排名第一的特征最重要），并仅选取部分重要特征/排名靠前的特征来重新训练模型。例如，在我们的实验中，我们仅选取排序列表中排名前 1/3 的特征，重新训练晕动症深度学习模型。这样得到的模型更轻量，可训练参数显著减少，因此训练和推理时间也更短。

4 数据集与实验设置

本节介绍用于验证所提出的 LiteVR框架的实验设置和数据。我们使用 Scikit-Learn [55] 和 TensorFlow-2.4 [65] 训练并评估深度学习模型。为解释深度学习模型，我们使用了 SHAP [44] 库。深度学习模型在配备 32GB RAM 和 NVIDIA GeForce RTX 3080 Ti GPU 的英特尔酷睿 i9 处理器上进行训练。

4.1 超参数

我们使用 Adam优化器 对深度学习模型进行优化，训练轮数为 300，批次大小为 256。训练深度学习模型时，我们使用的学习率为 0。001。为了防止模型过拟合，我们在训练深度学习模型时采用了提前停止策略，耐心值设为 30。与 [4] 类似，我们采用十折交叉验证来训练深度学习模型并测试其性能，其中数据集被划分为 k 组（即在我们的研究中 k = 10）。其中一组用于测试模型，其余各组用于训练。该方法重复k 次，每次选择一个新的分组作为测试集，其余（$k-1$）组作为训练数据集，以消除偏差。

4.2 数据集

我们使用集成传感器数据集来验证所提出的深度学习模型的有效性。该集成传感器数据集 [24]包含 30 名参与者沉浸于 5 种不同虚拟现实仿真中的眼动追踪、头部追踪和生理信号数据：Beach City、Road Side、过山车、SeaVoyage 和家具店。眼动追踪、头部追踪和生理数据由不同的子类别构成。例如，眼动追踪数据的子类别包括瞳孔直径（左眼）、瞳孔位置（x、y、z）、注视方向（x、y、z）、辐辏距离和眼睛张开度百分比；头部追踪数据则包括 X、Y、Z 和 W 轴的四元数旋转。类似地，生理信号的子类别包括皮肤电活动（电皮肤活动）和心率测量。该数据集共包含 20104 个样本，记录时间最长为 7 分钟的虚拟现实仿真。此外，如前所述，该数据集包含四种不同的网络晕动症严重程度类别：无、低、中和高；FMS评分范围为 0 至 10，用于回归分析。我们使用该数据集的 70% 样本训练深度学习模型，其余 30% 样本用于测试。

4.3 性能指标

使用准确率、召回率、精确率和 F1分数等标准质量指标，评估深度学习模型（如 LSTM、门控循环单元和门控循环单元）对网络晕动症严重程度进行分类的性能 [15]。同样，使用均方误差 (MSE)、均方根误差 (RMSE)、平均绝对误差 (MAE) 和 R^2 分数等标准损失函数，评估网络晕动症回归模型的性能 [10]。例如，如果 y_t 和 $\hat{y}_t$ 分别表示某候选人在时间t的实际和预测网络晕动症程度，那么 MAE 可定义为：

$$MAE = \frac{1}{|N|}\sum_{y\in S}\sum_{t=1}^{N}(y_t - \hat{y}_t)$$

其中，N 和S 分别表示所有轨迹的样本总数和时间步数。MAE 越小，回归模型越好。如果 $\sum(y_t - \hat{y}_t)^2$ 和 $\sum(y_t - \bar{y})^2$ 分别表示回归平方和（SSR）与总平方和（SST），那么R^2 分数可表示为：

$$R^2 = 1 - \frac{\sum(y_t - \hat{y}_t)^2}{\sum(y_t - \bar{y})^2}$$

表 1：LSTM 模型架构（包含全部特征）

层	Type	输出形状	参数数量	丢弃	激活函数
1	LSTM	128	66560	0.2	-
2	丢弃	128	0	0.2	-
3	LSTM	64	49408	0.15	-
4	丢弃	64	0	0.15	-
5	全连接层	16	3136	-	ReLU
6	全连接层	4	68	-	Softmax
参数总数：			**119, 172**		

表 2：门控循环单元模型架构（包含全部特征）

层	Type	输出形状	参数数量	丢弃	激活函数
1	GRU	32	3360	0.3	-
2	丢弃	32	0	0.3	-
3	GRU	64	18816	0.3	-
4	丢弃	64	0	0.3	-
5	全连接层	16	74496	-	ReLU
6	全连接层	4	516	-	Softmax
参数总数：			**97,188**		

因此，较低的R^2 值表明该回归模型无法充分捕捉输出方差。

5 结果

本节介绍网络晕动症检测、解释和简化的结果。

5.1 使用全部特征时的晕动症分类与回归性能

在介绍我们提出的 LiteVR 方法在网络晕动症解释和模型简化方面的结果之前，本节首先详细介绍模型开发情况及其重要统计数据，以及晕动症分类和回归准确率。这将有助于我们在论文的下一小节中比较该方法的有效性。

5.1.1 晕动症检测模型开发 ent

表 5.3、2、3 展示了基于长短期记忆网络、门控循环单元和多层感知机的晕动症检测模型的网络架构及可训练参数数量。我们观察到，机器学习、长短期记忆网络和门控循环单元模型的可训练参数总数分别为 17,028、119,172 和 97,188。与长短期记忆网络和门控循环单元模型相比，多层感知机模型的可训练参数数量分别少 7 倍和 5.7 倍。这些模型的训练和推理时间见表 4。使用全部特征训练多层感知机、长短期记忆网络和门控循环单元模型，分别需要 51.25、2076.21 和 3236.21 秒。事实上，多层感知机模型的训练速度比门控循环单元模型快 232.33 倍，比长短期记忆网络模型快 63.14 倍。同样，多层感知机、长短期记忆网络和门控循环单元模型所需的推理时间分别为 0.15、4.05 和 4.94 秒。事实上，多层感知机模型的推理速度分别比长短期记忆网络和门控循环单元模型快 27 倍和 33 倍。

5.1.2 使用全部特征进行晕动症分类和回归

表5总结了使用LSTM、GRU和MLP模型进行晕动症严重程度分类的准确率、精确率、召回率和F1分数。使用LSTM和GRU模型进行晕动症分类时，准确率分别为92%和90%，而MLP的准确率仅为79%。事实上，在晕动症严重程度分类的精确率、召回率和F1分数方面，LSTM模型的整体表现略优于GRU模型。此外，表6总结了使用LSTM、GRU和MLP模型进行晕动症回归的MSE、RMSE、R^2, 和MAE。与晕动症分类类似，使用LSTM进行回归的表现优于GRU和MLP模型。与分类任务一样，MLP在回归任务中的表现也不佳。

表 3：MLP模型架构（使用全部特征）

层	Type	输出形状	参数数量	丢弃	激活函数
1	全连接层	128	5504	0.25	ReLU
2	丢弃	128	0	0.25	
3	全连接层	64	8256	0.25	ReLU
4	丢弃	64		0.25	-
5	全连接层	32	2080	0.25	ReLU
6	丢弃	32	0	0.25	-
7	全连接层	16	1056	-	ReLU
8	全连接层	4	132	-	Softmax
参数总数：			**17,028**		

表 4：未降维深度学习模型（全部特征）与降维深度学习模型（精简特征）的训练和推理时间

模型	可解释人工智能之前		可解释人工智能之后	
	训练（秒）	推理（秒）	训练（秒）	推理（秒）
LSTM	2076.21	4.05	376.37	1.07
GRU	3236.21	4.94	774.29	1.32
MLP	51.25	0.15	21.55	0.09

5.2 基于 SHAP 的网络晕动症解释进行特征降维

在本节中，我们采用基于SHAP的事后解释方法来解释深度学习模型的输出。随后，我们利用这些解释，特别是全局解释，来识别起主导作用的特征，以帮助精简网络晕动症深度学习模型。

5.2.1 晕动症严重程度全局解释

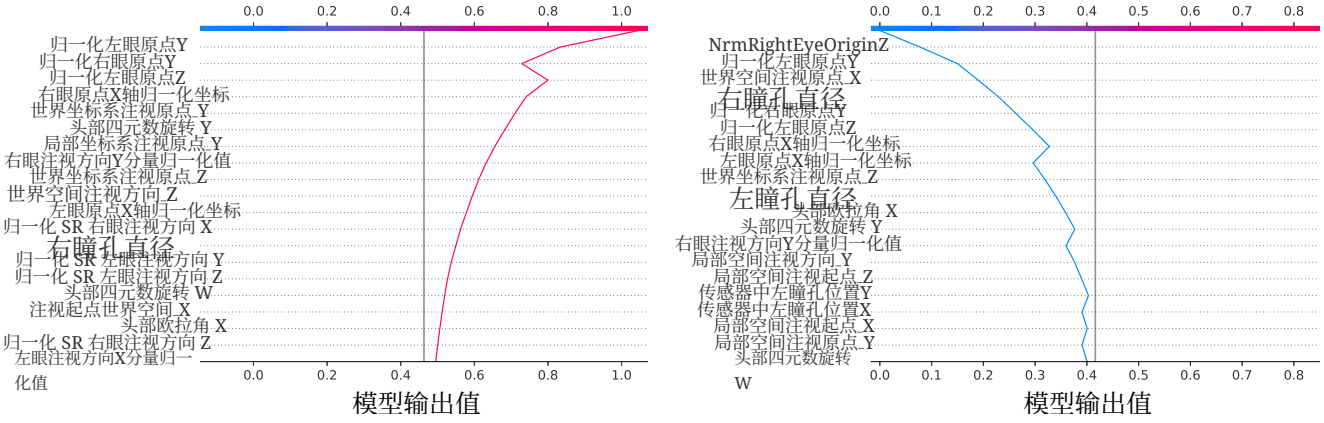
图 5 展示了使用全部特征的 LSTM、GRU 和 MLP 模型对晕动症严重程度进行分类时的整体特征重要性。沙普利值用于计算各特征对晕动症严重程度分类的贡献排名，重要特征位于顶部，最不重要的特征位于底部。绝对沙普利值较大的特征被视为重要特征，因为它们对模型输出的平均影响更大。从图 5a 和 b 可以看出，对于 LSTM 和 GRU 模型，*NrmRightEyeOriginZ*、归一化左眼原点*Y*、*Nrm- RightEyeOriginY*、归一化左眼原点*Z* 、*GazeOriginWrldSpc Y* 等特征是晕动症严重程度分类中影响最大的特征，这与眼动追踪特征相对应。有趣的是，对于 LSTM 和 GRU 模型，眼动追踪特征对晕动症严重程度分类的影响远大于头部追踪特征。这是因为眼动测量数据包含丰富的信息，例如用户的注视行为、瞳孔位置和眨眼类型，可用于追踪用户活动 [24]。因此，这些眼动追踪特征会影响网络晕动症的程度。另一方面，对于 MLP 模型，*HeadQRota- tionY*、*GazeOriginWrldSpc˙Y*、*GazeOriginWrldSpc Z*、头部欧拉角 *X*、电皮肤活动等特征似乎是晕动症分类中预测能力最强的特征（图 5c）。MLP 模型在晕动症严重程度分类中产生了许多假阳性和假阴性。然而，如第 5.1.2 节所述，与 LSTM 和 GRU 模型相比，MLP 模型的晕动症严重程度分类准确率相当低。因此，图 5c 中将*HeadQRotationY*排在特征列表首位等情况，反映了特征排名不准确。先前的研究 [8, 24, 43, 76] 指出，与头部追踪特征相比，眼动追踪特征对晕动症严重程度分类的影响最大。这种针对误分类的网络晕动症解释，能为分类结果提供重要见解，并增强对模型预测的信任。

5.2.2 晕动症严重程度局部解释 ion

随机从模型中抽取单个样本，以解释其网络晕动症局部解释的结果。结果

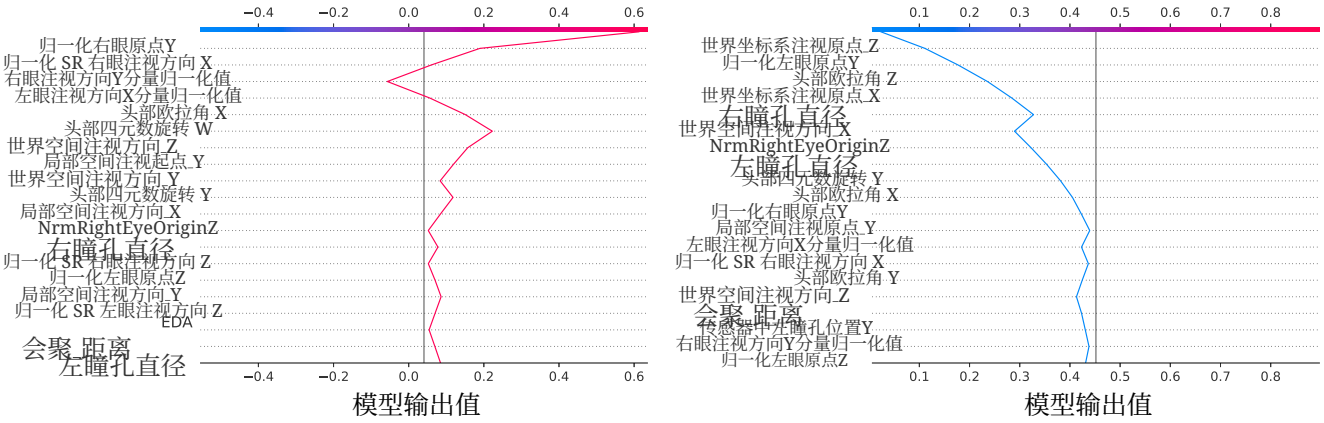
表5：未降维深度学习模型在晕动症严重程度分类上的10折交叉验证性能

模型	准确率%	精确率%				召回率%				F1分数%			
		None	Low	中	High	None	Low	中	High	None	Low	中	High
LSTM	92	96	86	87	90	92	95	86	93	95	92	83	90
GRU	90	95	89	81	83	94	90	79	87	95	90	81	86
MLP	79	83	79	80	75	81	85	72	78	82	77	70	75



(a) (b)

图2：基于长短期记忆网络的晕动症分类局部解释：（**a**）对高网络晕动症严重程度的解释；（**b**）对无网络晕动症严重程度的解释。



(a) (b)

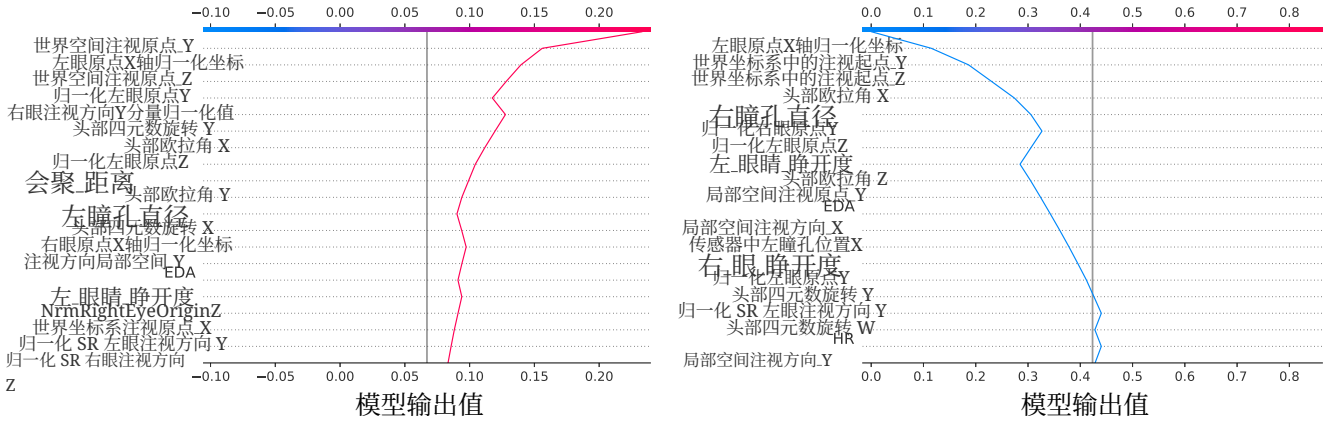
图3：基于门控循环单元的晕动症分类局部解释：（**a**）对高网络晕动症严重程度的解释；（**b**）对无网络晕动症严重程度的解释。

表6：使用未降维深度学习模型进行晕动症回归（使用全部特征）

回归模型	MSE	RMSE	R^2	MAE
LSTM	0.24	0.36	90	0.21
GRU	0.29	0.41	86	0.24
MLP	0.42	0.57	73	0.40

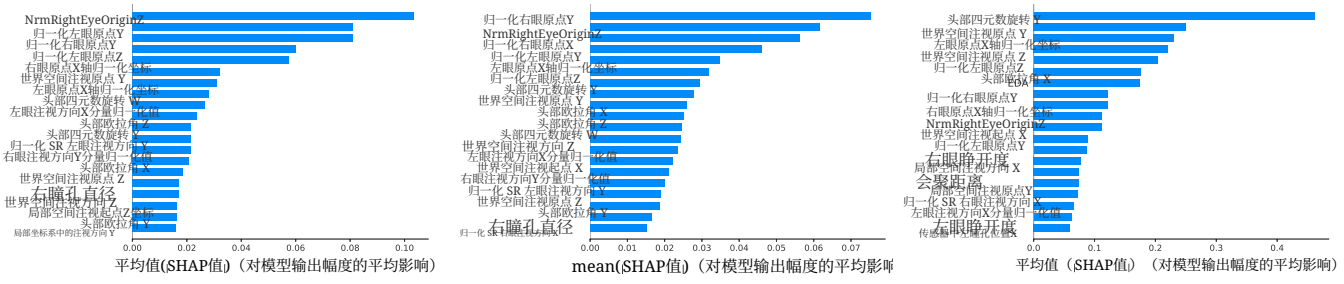
使用 SHAP 对长短期记忆网络模型进行局部解释的结果如图2所示。图2a展示了高网络晕动症严重程度分类。在该图中，红色线条表示该样本对应的网络晕动症概率，*x*轴表示模型以对数几率形式给出的输出（特征对预测的重要性概率），*y*轴列出模型的特征。可以看出，眼动追踪特征 归一化左眼原点*Y* 是高网络晕动症严重程度分类中影响最大的特征，其概率最高，接近 1。0。大多数

除*HeadQRotationY* 外，这些特征属于眼动追踪特征的结果概率最高；因此，对高晕动症严重程度进行分类是恰当的决策。例如，在图2b中，无晕动症严重程度类别中，眼动追踪特征的概率得分较低，其中*HeadQRotationW* 是影响最大的特征，概率为 0.4；而*NrmRightEyeOriginZ* 影响最小，概率为 0。因此，无晕动症严重程度类别得到了正确分类。类似地，图3展示了使用门控循环单元模型对分类后的晕动症所作的局部解释。在图3 a和图3 b中，我们观察到，大多数有助于判定为高晕动症严重程度类别的特征都对应于眼动追踪特征；而对于无晕动症严重程度类别，占主导地位的特征大多来自头部追踪特征，即头部欧拉角 *Z*、头部欧拉角 *Y*、*CLeftPupilPosInSensorY*、*Conver-˙distance*等。同样，图 4展示了使用多层感知机模型对分类后的晕动症所作的局部解释。从图4a可以观察到，最具影响力的特征是



(a) (b)

图 4：基于多层感知机的晕动症分类的局部解释：**(a)** 网络晕动症严重程度较高时的解释，**(b)** 网络晕动症严重程度较低时的解释。



(a) (b) (c)

图 5：使用 SHAP 全局解释得出的总体特征重要性：（**a**）长短期记忆网络模型（**b**）门控循环单元模型（**c**）多层感知机模型

高晕动症严重程度类别对应的眼动追踪特征，例如 *GazeOriginWrldSpc Y*、左眼原点*X*轴归一化坐标等。然而，诸如电皮肤活动、*HeadQRotationY*、头部欧拉角 *X*等特征，在高晕动症严重程度分类中也具有重要作用。这一观察结果容易造成误导，因为我们知道眼动追踪特征对晕动症预测更为关键。不过，这一解释也说明了为什么使用多层感知机模型进行晕动症分类的准确率低于长短期记忆网络和门控循环单元模型。因此，图4b显示，对正向结果（无晕动症严重程度类别）产生影响的大多数特征，同样属于头部追踪特征和生理信号特征。头部欧拉角 *Z*、头部欧拉角 *Z*、头部四元数旋转 *W*、心率等，是预测无晕动症严重程度类别时最重要的特征。识别这类特征有助于深入了解分类结果，并增强对模型输出的信任，从而做出恰当的决策。

5.3 晕动症检测的模型缩减

本节介绍特征选择和网络晕动症模型简化的详细过程。利用上一节所述的（全局）解释结果，我们首先根据算法 1 中的 SHAP 解释结果确定排名前 1/3 的特征。选定特征后，我们基于这些特征重新训练深度学习模型。我们选择 1/3 的主导特征，从基础模型开发精简模型，因为使用这一比例的特征能够实现相近的检测准确率。需要注意的是，对于不同的数据集，这一比例可能有所不同。还值得一提的是，我们使用门控循环单元和长短期记忆网络模型的解释结果，因为与多层感知机模型相比，它们具有更高的准确率。表 7 展示了通过 SHAP 解释结果选定的特征，这些特征用于重新训练长短期记忆网络、门控循环单元和多层感知机模型。表 8、表 9 和表 10 展示了重新训练的长短期记忆网络、门控循环单元和多层感知机模型的网络架构。我们观察到，基于解释的模型简化显著减小了模型规模。例如，简化后，长短期记忆网络模型共有 19,108 个可训练参数，相较于其未简化版本减少了 6.26 倍。同样，门控循环单元和多层感知机模型在模型简化后分别具有 24,388 个和 4,020 个可训练参数，相较于表 2 和表 3 中各自的未简化版本，分别减少了 4 倍和 4.25 倍。精简模型的训练和推理时间见表 4。我们观察到，精简多层感知机、长短期记忆网络和门控循环单元模型的训练时间分别为 21.25、376.37 和 774.29 秒，相较于各自的未简化版本，分别快 2.4 倍、5.6 倍和 4.2 倍。同样，简化后的推理时间也显著缩短。例如，精简多层感知机、长短期记忆网络和门控循环单元模型的推理时间分别为 0.09 、 1.07 和 1.32 秒，相较于各自的未简化版本，分别快 1.66 倍、3.78 倍和 3.72 倍。在下面的小节中，我们将介绍精简网络晕动症深度学习模型的训练和推理性能。

5.3.1 使用简化的深度学习模型进行晕动症分类

表 11 总结了使用降阶长短期记忆网络、门控循环单元和多层感知机模型进行网络晕动症分类的准确率、精确率、召回率和 F1分数。例如，使用精简长短期记忆网络进行网络晕动症分类的准确率为 94%，略高于未精简长短期记忆网络模型的准确率（见表 5）。此外，与各自未精简的版本相比，精简多层感知机和门控循环单元模型的网络晕动症分类准确率分别提高了 11.5% 和 9.7%。此外，与未精简的模型相比，精简长短期记忆网络、门控循环单元和多层感知机模型在无、低、中和高网络晕动症类别上的精确率、召回率和 F1分数等其他性能指标也略有提高

表 7：通过 SHAP 解释重新训练 LSTM、门控循环单元和多层感知机模型所选定的特征列表

测量类型	通过 SHAP 解释选定的特征
眼动追踪数据	NrmRightEyeOriginZ；归一化左眼原点Y；右眼原点X轴归一化坐标；归一化右眼原点Y；归一化左眼原点Z；左眼原点X轴归一化坐标；左眼注视方向X分量归一化值；归一化 SR 左眼注视方向 Y；右眼注视方向Y分量归一化值；GazeOriginWrldSpc Y；GazeOriginWrldSpc Z；世界空间注视方向Z分量；右瞳孔直径；局部空间注视原点Z坐标
头部追踪数据	头部四元数旋转 W；头部四元数旋转 Y；头部欧拉角 X；头部欧拉角 Z

表 8：精简长短期记忆网络架构（使用精简特征）

层	Type	输出形状	参数数量	丢弃	激活函数
1	LSTM	64	16896	0.15	-
2	丢弃	64	0	0.15	
5	全连接层	32	2080	-	ReLU
6	全连接层	4	132	-	Softmax
参数总数：			**19,108**		

表 9：精简GRU模型架构（含精简特征）

层	Type	输出形状	参数数量	丢弃	激活函数
1	GRU	32	3360	0.2	-
2	丢弃	32	0	0.2	
3	GRU	64	18816	0.2	-
4	丢弃	64	0	0.2	-
5	全连接层	32	2080	-	ReLU
6	全连接层	4	132	-	Softmax
参数总数：			**24,388**		

版本。例如，精简多层感知机在无、低、中和高网络晕动症类别上的精确率分别为 97%、88%、90% 和 92%，略优于未精简长短期记忆网络模型。同样，精简多层感知机模型在无、低、中和高网络晕动症类别上的召回率分别为 93%、91%、79% 和 85%，约为未精简多层感知机模型的 1.1X、1.06X、1.09X 和 1.15X。这些结果表明，在所有情况下，使用精简模型都显著提升了网络晕动症分类效果。

5.3.2 使用降维深度学习模型进行网络晕动症回归

表12展示了使用降阶长短期记忆网络、门控循环单元和多层感知机模型进行网络晕动症回归的性能。具体而言，降阶长短期记忆网络模型的均方误差、均方根误差、平均绝对误差和（R^2）分别为0.18、0.30、0.17和0.92；降阶门控循环单元模型的相应指标分别为0.24、0.36、0.20和0.88。对于降阶多层感知机，我们观察到均方误差、均方根误差、R^2, 和平均绝对误差均有显著改善。这是因为多层感知机在特征集较小时效果更好。例如，与未降阶版本相比，降阶多层感知机的均方根误差、均方误差和平均绝对误差分别降低了约14%、17%和18%，而R^2 则提高了7.6%。在所有用于网络晕动症回归的降阶深度学习模型中，都可以看到这类改进。

6 讨论

基于 SHAP 的全局解释揭示了正常眼睛原点、注视原点、瞳孔直径等特征是导致网络晕动症的最重要特征。相反，针对特定预测结果提供的基于 SHAP 的局部解释为每个样本提供了有用的洞见，有助于解释误分类情形。因此，我们利用 SHAP 根据全局解释对特征的重要性进行了排序，并使用重要特征重新训练深度学习模型。结果表明，在不牺牲准确率的情况下，SHAP 引导的降维深度学习模型能够显著缩短训练和推理时间。例如，与未降维版本相比，使用仅 1/3 特征进行训练的降维 LSTM 晕动症模型，以 94% 的准确率对晕动症严重程度进行了分类。同样，在对晕动症进行回归时，降维 LSTM 获得了均方根误差（RMSE）

表 10：精简多层感知机模型架构（采用精简特征）

层	Type	输出形状	参数数量	丢弃	激活函数
1	全连接层	64	1344	0.25	ReLU
2	丢弃	64	0	0.25	
3	全连接层	32	2080	0.25	ReLU
4	丢弃	32		0.25	-
5	全连接层	16	528	-	ReLU
6	全连接层	4	68	-	Softmax
参数总数：			**4,020**		

值 0.30，比未降维版本低 16.8%。LSTM 和门控循环单元模型在晕动症检测和回归方面均表现出色，而多层感知机表现较差。多层感知机在分类和回归中的表现不佳并不令人意外，因为与 LSTM 和门控循环单元相比，多层感知机无法记忆过去的序列。我们的 LSTM 模型在基于深度学习的晕动症检测方面的准确率优于多项最先进水平的研究。例如，Islam 等人 [24] 采用深度融合模型，以 87.7% 的准确率对晕动症严重程度进行分类，并使用眼动和头部追踪数据报告了 0.51 的均方根误差。在 [23] 中，同一作者使用深度时序卷积网络（DeepTCN），基于眼动追踪、心率和皮肤电反应数据，以 0.49 的均方根误差预测晕动症 FMS 评分（范围为 0–10）。我们的 LSTM 模型在分类和回归准确率方面均优于这些研究。此外，还有其他与我们工作相关的研究。例如，Qu 等人 [58], 、Islam 等人 [25], 、Kim 等人 [35], 和 Jeong 等人 [28] 分别使用生理信号和脑电/心电信号，报告了 96.85%、97.44%、89.16% 和 94.02% 的晕动症检测准确率。

尽管关于网络晕动症检测方法的研究已相当丰富，但迄今只有少数研究探讨了网络晕动症的成因 [21,24,35,52]。然而，据我们所知，目前尚无研究将可解释人工智能应用于解释网络晕动症深度学习模型并简化其维度。事实上，基于可解释人工智能的解释有助于研究人员理解晕动症分类正确与错误的原因，并可进一步用于开发有效的网络晕动症减少方法。另一方面，使用复杂的深度学习模型进行实时晕动症检测需要大量计算资源，可能不适合集成到能源受限的独立式虚拟现实头戴显示器中 [5]。因此，我们认为，所提出的基于可解释人工智能的晕动症模型简化方法有助于将这些模型部署到独立式头戴显示器上。

7 局限性与未来工作

尽管所提出的基于可解释人工智能和深度学习的网络晕动症检测与缓解方法（具体而言是长短期记忆网络）优于以往最先进水平的晕动症检测模型，但我们的方法仍存在一些局限。例如，尽管我们证明了所提出的基于可解释人工智能的模型简化方法有效，且训练和推理时间较短，但并未将其部署到实际的虚拟现实头显上。我们计划在未来解决这一局限。此外，网络晕动症可能会因个人特征、虚拟现实环境和性别而对不同的人产生不同影响 [69]。因此，未来我们计划面向人口统计学背景更加多样、且性别比例均衡的人群开展进一步研究，

表 11：使用降维深度学习模型（精简特征）进行晕动症严重程度分类的 10折交叉验证性能

模型	特征数量		准确率%	精确率%				召回率%				F1分数			
	原始	精简		None	Low	中	High	None	Low	中	High	None	Low	中	High
LSTM	43	18	94	97	88	90	92	94	95	88	96	94	93	87	91
GRU	43	18	93	96	91	85	89	96	92	84	90	95	93	83	88
MLP	43	18	89	91	84	82	79	93	91	79	85	92	88	80	84

表12：使用精简特征的降维深度学习模型进行晕动症回归

回归模型	特征数量		MSE	RMSE	R^2	MAE
LSTM	43	18	0.18	0.30	0.92	0.17
GRU	43	18	0.24	0.36	0.88	0.20
MLP	43	18	0.35	0.49	0.79	0.33

此外，本研究仅使用眼动追踪、头部追踪和生理信号来检测网络晕动症。未来，我们计划研究立体图像和立体视频数据对结合可解释性进行网络晕动症检测的作用。最后，我们承认目前无法获取暴露时间超过 7 分钟的数据集。未来，我们计划开发一个长暴露时间的新数据集，以进一步验证我们的方法。

8 结论

本文提出了 LiteVR 框架，这是一种基于可解释人工智能的框架，利用轻量级深度学习模型及其解释来检测网络晕动症。具体而言，我们开发了三种用于网络晕动症检测的深度学习模型：多层感知机、长短期记忆网络和门控循环单元，并应用可解释人工智能对其进行解释和精简。我们使用包含眼动追踪、头部追踪和生理信号的集成传感器数据集，展示了所提方法的有效性。全局解释结果表明，眼动追踪特征对网络晕动症的诱发影响最大；借助局部解释，我们识别出了误分类实例。此外，基于可解释人工智能的特征排序，我们大幅缩减了模型规模（最高达 4.3 倍），缩短了训练时间（最高达 5.6 倍）和推理时间（最高达 3.8 倍）。例如，我们的精简长短期记忆网络晕动症模型能够以 94% 的准确率和 0.30 的均方根误差预测网络晕动症，性能优于最先进的研究。我们相信，这项研究将有助于未来致力于网络晕动症检测与缓解，以及将这些模型部署到消费级虚拟现实头显上的研究人员。

致谢

本研究由美国国家科学基金会资助（CNS 2114035、IIS 2007041、IIS 2211785）。本出版物中表达的任何观点、发现、结论或建议均属于作者个人意见，不一定反映美国国家科学基金会的观点。

参考文献

[1] 用于对虚拟现实沉浸引发的网络晕动症进行分类的机器–深度–集成学习模型。《网络心理学、行为与社交网络》，24(11):729–736，2021年。[2] M. Alca˜niz、I. A. Chicchi-Giglioli、L. A. Carrasco-Ribelles、J. Mar´ın-Morales、M. E. Minissi、G. Teruel-Garc´ıa、M. Sirera 和 L. Abad。利用虚拟现实与机器学习，将眼动作为识别自闭症谱系障碍的生物标志物：用于诊断的概念验证。《自闭症研究》，15(1):131–145，2022年。[3] M. H. Alkinani 和 M. R. El-Sakka。用于图像去噪的基于图像块的模型与算法：针对加性噪声抑制的基于图像块的图像去噪方法比较综述。《*EURASIP* 图像与视频处理期刊》，2017(1):1–27，2017年。[4] Y. Bengio 和 Y. Grandvalet。不存在 k 折交叉验证方差的无偏估计量。神经信息处理系统进展，16，2003年。

[5] L. Castaneda 和 M. Pacampara. 课堂中的虚拟现实——对硬件、管理、内容和教学法的探索. 载于教育信息技术与教师教育国际会议，第 527–534 页. 教育计算促进协会（AACE），2016. [6] T. T. Chan、Y. Wang、R. H. Y. So 和 J. Jia. 预测在对 VR 场景产生前庭-眼反射期间，VR 头显镜头畸变引起的主观不适.《*IEEE*可视化与计算机图形学汇刊》，第 1–1 页，2022. DOI: 10.1109/TVCG. 2022.3168190[7] T. T. Chan、Y. Wang、R. H. Y. So 和 J. Jia. 预测在对 VR 场景产生前庭-眼反射期间，VR 头显镜头畸变引起的主观不适.《*IEEE*可视化与计算机图形学汇刊》，2022. [8] E. Chang、H. T. Kim 和 B. Yoo. 基于用户在基于头戴式显示器的虚拟现实中的注视行为预测晕动症.《计算设计与工程杂志》，8(2):728–739，2021. [9] U. A. Chattha、U. I. Janjua、F. Anwar、T. M. Madni、M. F. Cheema 和 S. I. Janjua. 虚拟现实中的晕动症：一项实证评估.《*IEEE Access*》，8:130486–130499，2020.[10] D. Chicco、M. J. Warrens 和 G. Jurman. 在回归分析评估中，决定系数 r 平方比 SMAPE、平均绝对误差、MAPE、均方误差和均方根误差更具信息量.《*PeerJ* 计算机科学》，7:e623，2021. [11] M. S. Dennison 、 A. Z. Wisti 和 M. D'Zmura. 使用生理信号预测网络晕动症.《显示器》，44:42–52，2016. [12] J. Dong、K. Ota 和 M. Dong. VR 游戏为何会引发晕动症？一项关于采集和分析 VR 游戏头部运动数据集的实证研究.《*IEEE MultiMedia*》，第

1 1, 2022 10 1109 2022 3176142 – [13]

页 . DOI: . /MMUL. . J. Dong 、 K. Ota 和 M. Dong. VR 游戏为何会引发晕动症？一项关于采集和分析 VR 游戏头部运动数据集的实证研究.《*IEEE MultiMedia*》，2022.[14] L. E. Garrido、M. Fr´ ıas-Hiciano、M. Moreno-Jim´enez、G. N. Cruz、Z. E. Garc´ ıa-Batista、K. Guerra-Pe˜na 和 L. A. Medrano. 聚焦网络晕动症：普遍性、潜在轨迹、易感性及其对虚拟现实体验的影响.《虚拟现实》，第 1–25 页，2022. [15] C. Goutte 和 E. Gaussier. 对精确率、召回率和 F 值的概率解释及其对评估的影响. 载于欧洲信息检索会议，第 345–359 页. Springer，2005年. [16] J. G. Grandi、M. Ogren 和 R. Kopper. 为公共安全组织设计新一代用户界面的方法. 载于*2019*年*IEEE*虚拟现实与三维用户界面会议（*VR*），第 944–945 页. IEEE，2019. [17] A. Gulhane、A. Vyas、R. Mitra、R. Oruche、G. Hoefer、S. Valluripally、P. Calyam 和 K. A. Hoque. 虚拟现实学习环境应用的安全性、隐私和安全风险评估. 载于*2019*年第*16*届*IEEE*年度消费者通信与网络会议（*CCNC*），第 1–9 页. IEEE，2019. [18] A. Hadadi、C. Guillet、J.-R. Chardonnet、M. Langovoy、Y. Wang 和 J. Ovtcharova. 使用拓扑数据分析和机器学习预测虚拟环境中的网络晕动症.《虚拟现实前沿》，3，2022. [19] A. Halbig、S. Babu、S. Gatter、M. Latoschik、K. Brukamp 和 S. von Mammen. 虚拟现实在医疗保健中的机遇与挑战——领域专家的调查. front. *Virtual Real. 3: 837616. doi: 10.3389/frvir*，2022. [20] M. C. Howard 和 E. C. Van Zandt. 虚拟现实问题的元分析：个体差异导致虚拟现实晕动症的影响各不相同.《虚拟现实》，25(4):1221–1246，2021. [21] R. Islam. 一种基于深度学习的框架，用于检测和减轻网络晕动症的发作. 载于*2020*年*IEEE*虚拟现实与三维用户界面会议摘要与研讨会（*VRW*），第 559–560 页.

IEEE，2020. [22] R. Islam、S. Ang 和 J. Quarles. Cybersense：用于检测网络晕动症严重程度并自适应应用减轻技术的闭环框架. 载于*2021*年*IEEE*虚拟现实与三维用户界面会议摘要与研讨会（*VRW*），第 148–155 页. IEEE，2021年. [23] R. Islam、K. Desai 和 J. Quarles. 通过多模态深度融合网络融合生理、头部追踪和眼动追踪数据，以预测网络晕动症的发作. [24] R. Islam、K. Desai 和 J. Quarles. 基于集成头戴式显示器传感器的网络晕动症预测：采用眼动追踪和头部追踪数据的多模态深度融合方法. 载于*2021*年*IEEE*混合与增强现实国际研讨会（*ISMAR*），第 31–40 页. IEEE，2021年.[25] R. Islam、Y. Lee、M. Jaloli、I. Muhammad、D. Zhu、P. Rad、Y. Huang 和 J. Quarles. 使用深度神经网络从用户的生理信号中自动检测和预测网络晕动症严重程度. 载于*2020*年*IEEE*混合与增强现实国际研讨会（*ISMAR*），第 400–411 页. IEEE，2020. [26] D. Jeong 和 K. Han. 利用多模态感官信息预测网络晕动症. 载于第*28*届*ACM*虚拟现实软件与技术研讨会论文集，第 1–2 页，2022. [27] D. Jeong、S. Yoo 和 J. Yun. 使用深度学习算法进行基于脑电图的网络晕动症分析. 载于*2019*年*IEEE*虚拟现实与三维用户界面会议（*VR*），第 827–835 页. IEEE，2019. [28] D. Jeong、S. Yoo 和 J. Yun. 使用深度学习算法进行基于脑电图的网络晕动症分析. 载于*2019*年*IEEE*虚拟现实与三维用户界面会议（*VR*），第 827–835 页，2019年3月. DOI: 10.1109/VR.2019.8798334[29] W. Jia、M. Sun、J. Lian 和 S. Hou. 特征降维：综述.《复杂与智能系统》，第 1–31 页，2022. [30] W. Jin、J. Fan、D. Gromala 和 P. Pasquier. 自动预测虚拟现实游戏中的网络晕动症. 载于*2018*年*IEEE*游戏、娱乐与媒体会议（*GEM*），第 1–9 页. IEEE，2018. [31] R. S. Kennedy 、 N. E. Lane 、 K. S. Berbaum 和 M. G. Lilienthal. 模拟器晕动症问卷：量化模拟器晕动症的改进方法.《国际航空心理学杂志》，3(3):203–220，1993. [32] B. Keshavarz 和 H. Hecht. 验证一种量化晕动症的高效方法.《人为因素》，53(4):415–426，2011. [33] B. Keshavarz、K. Peck、S. Rezaei 和 B. Taati. 结合生理测量指标和机器学习技术检测并预测视觉诱发晕动症.《国际心理生理学杂志》，2022. [34] H. K. Kim、J. Park、Y. Choi 和 M. Choe. 虚拟现实晕动症问卷（VRSQ）：虚拟现实环境中的晕动症测量指标.《应用人体工程学》，69:66–73，2018. [35] J. Kim、W. Kim、H. Oh、S. Lee 和 S. Lee. 一种基于脑信号分析、用于预测虚拟现实内容所致晕动症的深度预测器. 载于*IEEE/CVF*国际计算机视觉会议论文集，第 10580–10589 页，2019. [36] Y. Y. Kim、E. N. Kim、M. J. Park、K. S. Park、H. D. Ko 和 H. T. Kim. 应用生物信号反馈减轻暴露于虚拟环境所致的网络晕动症.《临场感：远程操作与虚拟环境》，17(1):1–16，2008. [37] R. Kottaimalai、M. P. Rajasekaran、V. Selvam 和 B. Kannapiran. 在脑机接口应用中，使用主成分分析和神经网络进行脑电信号分类. 载于*2013*年*IEEE*新兴计算、通信与纳米技术趋势国际会议（*ICECCN*），第 227–231 页. IEEE，2013. [38] R. K. Kundu、R. Islam、P. Calyam 和 K. A. Hoque. TruVR：使用可解释机器学习进行可信的晕动症检测. 载于*2022*年*IEEE*混合与增强现实国际研讨会（*ISMAR*），第 777–786 页. IEEE，2022. [39] J. J. LaViola Jr. 关于虚拟环境中网络晕动症的讨论.《*ACM SIGCHI* 通讯》，32(1):47–56，2000. [40] T. M. Lee 、 J.-C. Yoon 和 I.-K. Lee. 使用三维卷积神经网络预测立体视频中的晕动症.《*IEEE* 可视化与计算机图形学汇刊》，25 1927，2019 5 1919( ): – . [41] C.-T. Lin、S.-F. Tsai 和 L.-W. Ko. 基于脑电图的学习系统，用于动态车辆环境中的在线运动病级别估计.《IEEE神经网络与学习系统汇刊》，24(10):1689–1700，2013年。[42] Y.-T. Lin、Y.-Y. Chien、H.-H. Wang、F.-C. Lin 和 Y.-P. Huang。65-3：利用脑电图和心电图对虚拟现实头戴式显示器的晕动症程度进行量化。载于《SID技术论文研讨会文摘》，第49卷，第862–865页。Wiley Online Library，2018年。[43] P. Lopes、N. Tian 和 R. Boulic。眼睛看起来不太对劲！探索用于虚拟现实晕动症检测的眼动行为。载于《运动、交互与游戏》，第1–10页。2020年。[44] S. M. Lundberg 和 S.-I. Lee。《解释模型预测的统一方法》。《神经信息处理系统进展》，第30卷，2017年。[45] N. Martin、N. Mathieu、N. Pallamin、M. Ragot 和 J.-M. Diverrez。虚拟现实晕动症检测：一种基于生理信号和机器学习的方法。载于《2020年IEEE混合与增强现实国际研讨会》（ISMAR），第387–399页。IEEE，2020年。[46] M. A. Mawalid、A. Z. Khoirunnisa、M. H. Purnomo 和 A. D. Wibawa。利用时域特征提取和朴素贝叶斯对脑电信号进行分类以检测网络晕动症。载于《2018年计算机工程、网络与智能多媒体国际会议》（CENIM），第29–34页。IEEE，2018年。[47] M. Melo、J. Vasconcelos-Raposo 和 M. Bessa。沉浸式内容中的临场感与网络晕动症：内容类型、暴露时间和性别的影响。《计算机与图形学》，71:159–165，2018年。[48] A. Munawar、Z. Li、P. Kunjam、N. Nagururu、A. S. Ding、P. Kazanzides、T. Looi、F. X. Creighton、R. H. Taylor 和 M. Unberath。用于协同外科训练和数据生成的虚拟现实。《生物力学与生物医学工程中的计算方法：成像与可视化》，10(4):366–374，2022年。[49] Y. Nam、U. Hong、H. Chung 和 S. R. Noh。反映网络晕动症的眼动模式：来自虚拟现实游戏不同体验模式的证据。《网络心理学、行为与社交网络》，25(2):135–139，2022年。[50] H. Oh 和 W. Son。虚拟现实内容引发的网络晕动症及其严重程度：一项综合研究。《传感器》，22(4):1314，2022年。[51] S. Ooi、T. Tanimoto 和 M. Sano。用于提升灾害意识的虚拟现实火灾灾害训练系统。载于《2019年第8届教育与信息技术国际会议论文集》，第301–307页，2019年。[52] N. Padmanaban、T. Ruban、V. Sitzmann、A. M. Norcia 和 G. Wet-zstein。探索基于机器学习的360度立体视频晕动症预测方法。《IEEE可视化与计算机图形学汇刊》，24(4):1594–1603，2018年。[53] A. Panahi、R. Askari Moghadam、B. Tarvirdizadeh 和 K. Madani。简化U-Net：用于青光眼检测的深度学习智能医疗辅助工具。《进化智能》，第1–12页，2022年。[54] E. S. Pane、A. Zuhriya Khoirunnisaa、A. D. Wibawa 和 M. H. Purnomo。利用脑电信号和CN2规则归纳算法识别网络晕动症严重程度。载于《2018年智能信息学与生物医学科学国际会议》（ICIIBMS），第3卷，第 170 176, 2018 10 1109 2018 8549968 –页。DOI: . / [55] ICIIBMS。. F. Pedregosa、G. Varoquaux、A. Gramfort、V. Michel、B. Thirion、O. Grisel、M. Blondel、P. Prettenhofer、R. Weiss、V. Dubourg 等。Scikit-Learn：Python中的机器学习。《机器学习研究杂志》，12:2825–2830，2011年。[56] A. S. Pillai 和 P. S. Mathew。虚拟现实对医疗保健的影响：综述。《心理健康治疗中的虚拟现实与增强现实》，第17–31页，2019年。[57] T. Porcino、E. O. Rodrigues、F. Bernardini、D. Trevisan 和 E. Clua。利用符号机器学习算法识别虚拟现实游戏中的网络晕动症成因。《娱乐计算》，第100473页，2021年。[58] C. Qu、X. Che、S. Ma 和 S. Zhu。基于生物生理信号的虚拟现实晕动症检测。《CCF普适计算与交互汇刊》，第1–17页，2022年。[59] L. Rebenitsch 和 C. Owen。应用程序和视觉显示器中的网络晕动症综述。《虚拟现实》，20(2):101–125，2016年。[60] M. Recenti、C. Ricciardi、R. Aubonnet、I. Picone、D. Jacob、H. ´A. Svansson、S. Agnarsd´ottir、G. H. Karlsson、V. Baeringsd´ottir、H. Pe-

tersen 等。利用虚拟现实和移动平台评估大脑、肌肉与心脏信号，以期预测晕动症。《机器人学与人工智能前沿》，9:132，2021年。[61] G. Reddy、C. A. Spencer、K. Durkee、B. Cox、O. Fox Cotton、S. Gal-breath、S. Meyer、M. Natali、T. Seech、G. Severe-Valsaint 等。通过非侵入式生理传感器估计虚拟现实模拟中飞行员的认知负荷和网络晕动症。载于国际人机交互会议，第251–269页。Springer，2022年。[62] P. Reyero Lobo 和 P. Perez。利用心率变异性进行非侵入式网络晕动症检测。载于*ACM互动媒体体验国际会议*，第221–228页，2022年。[63] M. Satu、S. Rahman、M. Khan、M. Z. Abedin、M. S. Kaiser、M. Mah-mud 等。利用双向多层长短期记忆神经网络提升认知表现检测。载于国际脑信息学会议，第297–306页。Springer，2020年。[64] E. J. Schmitt 和 H. Jula。线性模型在旅行时间预测方面的局限性。载于*2007年IEEE智能交通系统会议*，第830–835页。IEEE，2007年。[65] A. Sergeev 和 M. Del Balso。Horovod：在TensorFlow中快速、简便地进行分布式深度学习。*arXiv*预印本 *arXiv:1802.05799*，2018年。[66] V. Sevinc 和 M. I. Berkman。对模拟器晕动症问卷及其变体进行心理测量学评估，以衡量消费级虚拟环境中的网络晕动症。《应用人体工程学》，82 102958，2020：。[67] A. Shrikumar、P. Greenside、A. Shcherbina 和 A. Kundaje。不只是黑箱：通过传播激活差异学习重要特征。*arXiv*预印本 *arXiv:1605.01713*，2016年。[68] A. Singla、S. G¨oring、D. Keller、R. R. Ramachandra Rao、S. Fremerey 和 A. Raake。全向视频模拟器晕动症问卷评估。载于*2021年IEEE虚拟现实与三维用户界面（VR）*，第198–206页，2021年。DOI：10.1109/VR50410.2021 .00041[69] K. Stanney、C. Fidopiastis 和 L. Foster。虚拟现实存在性别歧视，但并非不可改变。《机器人学与人工智能前沿》，7:4，2020年。[70] W. B. Stone III。模拟器晕动症问卷作为网络晕动症衡量工具的心理测量学评估。博士论文，爱荷华州立大学，2017年。[71] N. Tian、P. Lopes 和 R. Boulic。头戴式显示器中的网络晕动症综述：关注个体易感性。《虚拟现实》，第1–33页，2022年。[72] E. Tjoa 和 C. Guan。可解释人工智能（XAI）综述：迈向医疗可解释人工智能。*IEEE神经网络与学习系统汇刊*，32(11):4793–4813，2021年。DOI：10.1109/TNNLS . 2020.3027314[73] E. Toreini、M. Aitken、K. Coopamootoo、K. Elliott、C. G. Zelaya 和 A. Van Moorsel。人工智能信任与可信机器学习技术之间的关系。载于*2020年公平性、问责制与透明度会议论文集*，第272–283页，2020年。[74] S. Valluripally、B. Frailey、B. Kruse、B. Palipatana、R. Oruche、A. Gul-hane、K. A. Hoque 和 P. Calyam。检测破坏用户在虚拟现实学习环境中沉浸式体验的安全与隐私攻击。*IEEE服务计算汇刊*，2022年。[75] S. Valluripally、A. Gulhane、K. A. Hoque 和 P. Calyam。诱发网络晕动症的社交虚拟现实攻击建模与防御。*IEEE可信与安全计算汇刊*，2021年。[76] A. van Kasteren、K. Brunnstr¨om、J. Hedlund 和 C. Snijders。360度视频的体验质量——对编码和冻结失真的主观评估与眼动追踪评估。《多媒体工具与应用》，81(7):9771–9802，2022年。[77] H. Ved 和 C. Yildirim。利用脑电频谱数据检测虚拟现实中的心理工作负荷：一种深度学习方法。载于*2021年IEEE人工智能与虚拟现实国际会议（AIVR）*，第173–178页。IEEE，2021年。[78] J. Wang、R. Shi、Z. Xiao、X. Qin 和 H.-N. Liang。虚拟现实游戏中分辨率权衡对游戏体验、表现和模拟器晕动症的影响。载于*2022年IEEE虚拟现实与三维用户界面会议摘要与研讨会（VRW）*，第542–543页，2022年。DOI：10.1109/VRW55335.2022.00122[79] S. Wang、T. Zhou 和 J. Bilmes。Jumpout：改进的深度学习丢弃方法带有 ReLU 的神经网络。载于国际机器学习会议，第6668–6676页。PMLR，2019年。[80] Y. Wang、J.-R. Chardonnet、F. Merienne和J. Ovtcharova。运用模糊逻辑纳入个体差异，以预测虚拟现实导航过程中的网络晕动症。载于*IEEE虚拟现实与三维用户界面（VR）*，第373–381页。IEEE，2021年。[81] A. H. X. Yang、N. Kasabov和Y. O. Cakmak。用于研究网络晕动症的机器学习方法：系统综述。*Brain Informatics*，9(1):1–25，2022年。[82] H. Yi、J. Hong、H. Kim和W. Lee。Dexcontroller：设计一款具备抓握识别功能的虚拟现实控制器，丰富自然游戏体验。载于第*25*届*ACM虚拟现实软件与技术研讨会*，第1–11页，2019年。[83] Y. Yokota、M. Aoki、K. Mizuta、Y. Ito和N. Isu。健康受试者中与视觉诱发的姿势不稳定及心脏自主神经反应相关的晕动症易感性。*Acta oto-laryngologica*，125(3):280–285，2005年。[84] D. Zielasko。受试者001——一份关于虚拟现实诱发晕动症的详细自我报告。载于*2021年IEEE虚拟现实与三维用户界面会议摘要与研讨会（VRW）*，第165–168页，2021年。DOI：10.1109/VRW52623.2021.00038[85] D. Zielasko、A. Meißner、S. Freitag、B. Weyers和T. W. Kuhlen。基于头戴式显示器的数据分析任务中，与主观晕动症测量指标相关的动态视野缩减。