Comunicar, nº 76, v. XXXI, 2023 | Revista Científica de Educomunicación | ISSN: 1134-3478; e-ISSN: 1988-3293
www.revistacomunicar.com | www.comunicarjournal.com

Hacia una metodología de evaluación del rendimiento del alumno en entornos de aprendizaje iVR utilizando eye-tracking y aprendizaje automático

Towards learner performance evaluation in iVR learning environments using eye-tracking and machine-learning

Dra. Ana Serrano-Mamolar. Investigadora Postdoctoral, Departamento de Ingeniería Informática, Universidad de Burgos (España) (asmamolar@ubu.es) (https://orcid.org/0000-0002-0027-7128)
Ines Miguel-Alonso. Personal Investigador en Formación, Departamento de Ingeniería Informática, Universidad de Burgos (España) (imalonso@ubu.es) (https://orcid.org/0000-0001-8882-7587)
Dr. David Checa. Profesor Ayudante Doctor, Departamento de Ingeniería Informática, Universidad de Burgos (España) (dcheca@ubu.es) (https://orcid.org/0000-0001-6623-3614)
Dr. Carlos Pardo-Aguilar. Profesor Titular, Departamento de Ingeniería Informática, Universidad de Burgos (España) (cpardo@ubu.es) (https://orcid.org/0000-0003-1424-1318)

RESUMEN

Actualmente, el uso de los datos del seguimiento de la mirada en entornos de aprendizaje de Realidad Virtual inmersiva (iVR) está destinado a ser una herramienta fundamental para maximizar los resultados de aprendizaje, dada la naturaleza poco intrusiva del eye-tracking y su integración en las gafas comerciales de Realidad Virtual. Pero, antes de que se pueda generalizar el uso del eye-tracking en entornos de aprendizaje, se deben identificar las tecnologías más adecuadas para el procesamiento de datos. Esta investigación propone el uso de técnicas de aprendizaje automático para este fin, evaluando sus capacidades para clasificar la calidad del entorno de aprendizaje y predecir el rendimiento de aprendizaje del usuario. Para ello, se ha desarrollado una experiencia docente en iVR para aprender el manejo de un puente-grúa. Con esta experiencia se ha evaluado el rendimiento de 63 estudiantes, tanto en condiciones óptimas de aprendizaje como en condiciones con factores estresores. El conjunto de datos final incluye 25 características, siendo la mayoría series temporales con un tamaño de conjunto de datos superior a 50 millones de puntos. Los resultados muestran que la aplicación de diferentes clasificadores como KNN, SVM o Random Forest tienen una alta precisión a la hora de predecir alteraciones en el aprendizaje, mientras que la predicción del rendimiento del aprendizaje del usuario aún está lejos de ser óptima, lo que abre una nueva línea de investigación futura. Este estudio tiene como objetivo servir como línea de base para futuras mejoras en la precisión de los modelos mediante el uso de técnicas de aprendizaje automático más complejas.

ABSTRACT

At present, the use of eye-tracking data in immersive Virtual Reality (iVR) learning environments is set to become a powerful tool for maximizing learning outcomes, due to the low-intrusiveness of eye-tracking technology and its integration in commercial iVR Head Mounted Displays. However, the most suitable technologies for data processing should first be identified before their use in learning environments can be generalized. In this research, the use of machine-learning techniques is proposed for that purpose, evaluating their capabilities to classify the quality of the learning environment and to predict user learning performance. To do so, an iVR learning experience simulating the operation of a bridge crane was developed. Through this experience, the performance of 63 students was evaluated, both under optimum learning conditions and under stressful conditions. The final dataset included 25 features, mostly temporal series, with a dataset size of up to 50M data points. The results showed that different classifiers (KNN, SVM and Random Forest) provided the highest accuracy when predicting learning performance variations, while the accuracy of user learning performance was still far from optimized, opening a new line of future research. This study has the objective of serving as a baseline for future improvements to model accuracy using complex machine-learning techniques.

PALABRAS CLAVE | KEYWORDS

Entorno virtual, aprendizaje basado en juegos, aprendizaje automático, registro de mirada, extracción de características, neuroeducación.
Virtual environment, game-based learning, machinelearning, eye-tracking, feature extraction, neuroeducation.

Recibido: 2022-12-26 | Revisado: 2023-01-25 | Aceptado: 2023-02-23 | OnlineFirst: 2023-05-23 | Publicado: 2023-07-01
DOI https://doi.org/10.3916/C76-2023-01 | Páginas: 9-20

1. Introducción y estado del arte

Durante la última década, el abaratamiento de los sensores neurológicos y la simplificación de las técnicas de análisis y adquisición de datos en diferentes sectores han ampliado el alcance de muchas aplicaciones finales. Los sistemas de eye-tracking, por ejemplo, incorporan muchas de esas técnicas. Las soluciones costosas y personalizadas de investigación médica avanzada e incluso publicitarias (Duchowski, 2002) se han convertido en soluciones comerciales estables contando con computadoras portátiles de alta gama y visores de realidad virtual a precios razonables (Shadiev & Li, 2022). En comparación con otros neurosensores, el eye-tracking proporciona señales estables que describen el comportamiento de la mirada y que son una de las vías para analizar el comportamiento humano en educación y psicología (Rodero & Larrea, 2022), por mencionar algunos. Además, el eye-tracking tiene una poderosa ventaja en términos de aceptación del usuario final: su baja intrusividad. Por ejemplo, el usuario puede realizar libremente diversas tareas usando solo un par de gafas livianas equipadas con tecnología de eye-tracking. Este dispositivo neurosensorial también tiene un inconveniente: solo registra datos sobre la fijación ocular, la dilatación y constricción de la pupila. En otras palabras, no se monitorean las respuestas cerebrales a los objetos visuales externos que puedan causar que el ojo reaccione de una forma u otra.

Dos campos prometedores para la aplicación del eye-tracking son la educación (García Carrasco et al., 2015) y el entrenamiento (Gardony et al., 2020). El eye-tracking puede ayudar a responder muchas preguntas. ¿Cómo miramos los materiales de aprendizaje dependiendo de su presentación multimedia? ¿Con qué facilidad nos distraemos? ¿En qué actividades centramos más nuestra atención? ¿Durante cuánto tiempo podemos estar concentrados en un tema determinado, etc. (Farran et al., 2016; Glennon et al., 2020)? Las respuestas a estas preguntas pueden ayudar a los profesores y formadores a comprender mejor cómo aprendemos y a optimizar la experiencia de aprendizaje y entrenamiento para maximizar los resultados en ella. Para facilitar la resolución a estas cuestiones, se encuentra el eye-tracking presente tanto en entornos 2D: pantallas (Añaños-Carrasco, 2015), como en entornos 3D: mundo real y la realidad virtual inmersiva (iVR, según sus siglas en inglés).

Los entornos iVR presentan algunas ventajas desafiantes para el aprendizaje y el entrenamiento (Checa & Bustillo, 2020). En primer lugar, ofrecen aprendizaje práctico: experiencias interactivas centradas en el alumno en lugar de dirigidas por el profesor. En segundo lugar, los estudiantes aprenden de manera autónoma a su propio ritmo, a diferencia de las experiencias de aprendizaje estandarizadas que suelen reducir los resultados del aprendizaje. En tercer lugar, las dificultades de la vida real se pueden simular tanto para estudiantes como para trabajadores: desde reorientar la atención y el tiempo de permanencia en entornos urbanos (Lapborisuth et al., 2021) hasta la conciencia, prevención y detección de ansiedad o depresión en estudiantes (Martínez et al., 2021). Finalmente, los usuarios de entornos iVR no tienen la sensación de estar siendo observados: a medida que aumenta la inmersión de la experiencia dentro del entorno iVR después de un par de minutos, la sensación de ser observado disminuye, lo que provoca un comportamiento natural.

Como la experiencia iVR se puede grabar y monitorear de cerca, el rendimiento del usuario se evalúa con mayor precisión que, por ejemplo, en las experiencias de aprendizaje basadas en exámenes. El análisis de los comportamientos también se puede utilizar para evaluar el aprendizaje en iVR. Esta simulación de VR (Wismer et al., 2022) utilizada para la evaluación del cumplimiento y habilidades físicas de laboratorio predijo con precisión (77%) tanto el estatus de experto como el de novato del usuario. La recopilación de datos de comportamiento relevantes en VR, por ejemplo, el seguimiento del movimiento de la cabeza y los ojos, y los datos de behaviometría, proporcionarán resultados más precisos. El eye-tracking y los entornos iVR son tecnologías nuevas para el aprendizaje y la capacitación con un futuro desafiante tanto para el público general como para los especialistas. Los nuevos dispositivos de visualización montados en la cabeza (HMDs, según sus siglas en inglés) para experiencias iVR de alta calidad registran datos de eye-tracking de forma no intrusiva.

Hasta ahora, el eye-tracking se ha utilizado para acciones básicas: movimientos dentro de entornos iVR cuando el espacio físico es limitado (Sun et al., 2018), interacción con manos libres dentro del entorno iVR, como escribir texto (Ma et al., 2018) y mover objetos virtuales (Tanaka et al., 2021). Algunos ejemplos de tareas complejas son la priorización de una escena según la mirada del usuario (Patney

et al., 2016) y la medición de la carga de trabajo cognitivo mediante el eye-tracking, que se investigó por primera vez para una tarea muy específica: entrenar a los cirujanos durante tareas de anastomosis vesico-uretral análogas (Cowan et al., 2021). Aprovechar la tecnología de eye-tracking dentro de VR presenta un enfoque novedoso para estudiar la atención y la motivación del estudiante, mientras que mejora potencialmente la eficacia de la enseñanza y sirve como una valiosa herramienta de evaluación (Rappa et al., 2022). Sin embargo, se deben superar algunos problemas importantes antes de que pueda implementarse completamente en entornos de aprendizaje que aplican el eye-tracking. En primer lugar, se debe demostrar el procesamiento eficiente de conjuntos de datos masivos de iVR de las experiencias de aprendizaje con eye-tracking. En segundo lugar, suponiendo que se pudiera identificar información útil en esos conjuntos de datos: ¿podríamos identificar la mejor forma de aprender según los contenidos iVR disponibles? En tercer lugar, se deben establecer las técnicas más precisas para extraer esta información oculta, considerando que el aprendizaje es un proceso cambiante y personalizado para cada ser humano. Todas estas preguntas deben responderse para el eye-tracking 3D, una tarea más compleja que el eye-tracking 2D tradicional, basado en pantalla (Gardony et al., 2020). La tecnología de eye-tracking tiene el potencial de complementar otras herramientas de recopilación de datos y proporcionar conjuntos de datos distintos que pueden mejorar el aprendizaje en entornos de realidad virtual. Para este propósito, las técnicas de aprendizaje automático podrían ser una de las soluciones más prometedoras para todas estas tareas y preguntas (Gardony et al., 2020).

El aprendizaje automático (ML, según sus siglas en inglés) implica técnicas basadas en datos que se utilizan para aprender de grandes conjuntos de datos que describen tareas complejas. La aplicación de técnicas de ML a conjuntos de datos de eye-tracking registrados en entornos de aprendizaje iVR puede ser para diferentes tareas (Gardony et al., 2020). En primer lugar, el ML puede realizar una tarea comúnmente conocida como extracción de características, que se utiliza para identificar las características principales de aquellos conjuntos de datos donde se concentra la información clave. Por ejemplo, el análisis de componentes discriminantes jerárquicos, una técnica ML utilizada con éxito para la extracción de características de conjuntos de datos de eye-tracking y electroencefalogramas (EEG) para la reorientación de la mirada y la atención en diferentes eventos de la mirada (Lapborisuth et al., 2021). En segundo lugar, el ML puede clasificar un ejercicio de atención del usuario y la calidad de un entorno de aprendizaje; además, sobre la misma base, puede predecir el rendimiento de aprendizaje del usuario en comparación con patrones anteriores. Asish et al. (2022) propusieron el uso de aprendizaje profundo (redes neuronales convolucionales) para clasificar la atención en tres ejercicios durante una experiencia de aprendizaje iVR basada en un conjunto de datos de eye-tracking etiquetado. En tercer lugar, el ML se puede utilizar en una arquitectura más compleja para adaptar el entorno iVR de aprendizaje a las necesidades y el ritmo específicos de cada usuario individual.

Sobre la base de las tres tareas mencionadas anteriormente, el ML puede ayudar al diseño de experiencias iVR con eye-tracking. En esta investigación, se aborda la segunda tarea. Se utilizaron diferentes técnicas de ML para clasificar la calidad del entorno de aprendizaje y tratar de predecir el rendimiento de aprendizaje del aprendiz. Luego, estos dos objetivos se probaron en un gran conjunto de datos (>50 millones de puntos de datos) de experiencias reales dentro de un escenario de aprendizaje realista donde 63 estudiantes repitieron una tarea definida y mejoraron su rendimiento.

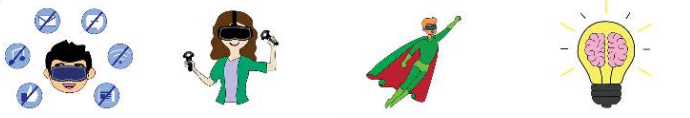
En comparación con un gran conjunto de datos anterior (Asish et al., 2022), compuesto de datos de eye-tracking etiquetado con 3,4 millones de puntos de datos, el de este estudio es 15 veces más grande y tiene una mayor diversidad de datos (diferentes niveles de experiencia y condiciones ambientales), lo que aumenta la complejidad de la tarea propuesta: desde la identificación del aprendiz hasta la clasificación de la calidad del aprendizaje y la predicción del rendimiento del aprendizaje del aprendiz. Finalmente, la pregunta a responder en esta investigación es si los conjuntos de datos basados en el eye-tracking de los entornos de aprendizaje iVR son adecuados para la evaluación de las condiciones de aprendizaje y rendimiento del aprendiz mediante el ML. Cabe señalar que esta investigación no pretende encontrar una solución fiable y robusta para estas tareas, sino un primer enfoque que proporcionará una línea de base para futuras mejoras en esta estrategia de investigación.

2. Material: Un entorno de aprendizaje iVR

En el desarrollo de una experiencia educativa iVR efectiva es recomendable seguir tres pasos (Figura 1): pre-diseño, diseño y evaluación (Checa & Bustillo, 2020). El primer paso, el pre-diseño, establece un escenario en el que se mejora el aprendizaje mediante la introducción de tecnologías iVR. En esta investigación se ha creado un entorno iVR para aprender a operar un puente-grúa. Esta maquinaria se utiliza en muchos procesos industriales y de transporte. La operación de control remoto significa que los simuladores de iVR pueden imitar de cerca las tareas industriales. La experiencia de entrenamiento de iVR adquiere datos de rendimiento del usuario durante los ejercicios para evaluar la experiencia y está diseñada para ser breve, fácil de aprender y repetible.

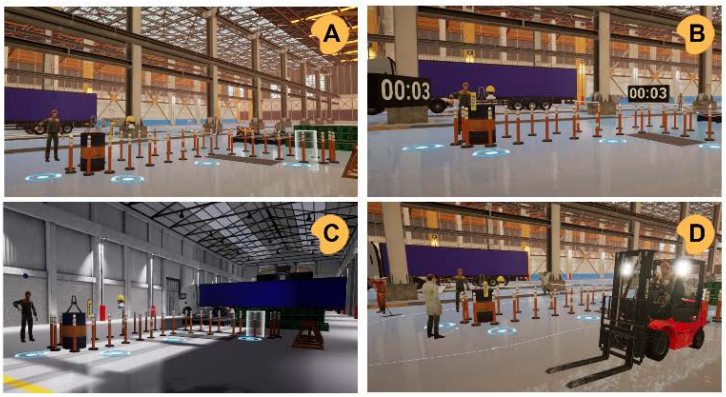
Figura 1. Etapas de desarrollo de la experiencia de aprendizaje iVR

(I) Pre-diseño: Definir un escenario en el que el aprendizaje se vea reforzado por la introducción de la tecnología iVR. Ventajas de la realidad virtual frente al aprendizaje tradicional (Checa & Bustillo 2022):



(II) Diseño: La experiencia está diseñada para lograr el máximo grado de inmersión del usuario. (A) Entorno fotorrealista para interactuar con el controlador del puente grúa. (B) Punto de teletransporte para el movimiento de fijación. (C) Actor virtual guiando al usuario. (D) Usuario realizando la experiencia con un HMD.

(III) Evaluación: (A) Analizar las habilidades de manejo de un puente grúa en usuarios noveles en condiciones ideales. (B) Entorno cronometrado (visual y auditivo). (C) Entorno con iluminación tenue (visual). (D) Entorno con alto tráfico de operarios (visual y auditivo).



Una vez fijados los objetivos de aprendizaje, es necesario aplicar un enfoque pedagógico y tener en cuenta las teorías de aprendizaje durante la fase de diseño. Las teorías del aprendizaje brindan pautas sobre las motivaciones de los aprendices, los procesos de aprendizaje y los resultados (Pritchard, 2017). Esta experiencia busca promover el aprendizaje vinculando iVR a una fusión de principios desde múltiples

perspectivas pedagógicas. Hay muchas teorías de aprendizaje desarrolladas para su uso en experiencias iVR o que pueden adaptarse fácilmente para su uso con estas nuevas tecnologías. Para esta investigación se consideraron cuatro teorías de aprendizaje.

En primer lugar, la Teoría del Aprendizaje Situado (Huang et al., 2010) que emplea un enfoque constructivista, en la medida en que los estudiantes aprenden habilidades participando activamente en experiencias iVR. En segundo lugar, la Perspectiva Tecnológica de los Entornos Virtuales de Aprendizaje 3D (Dalgarno & Lee, 2010), según los cuales los estudiantes aprenden a través de la interacción autónoma, el aprendizaje práctico y la resolución de problemas. En tercer lugar, el marco de la cognición incorporada (Wilson, 2002) donde plantea una conexión entre nuestros sentidos motor y visual; por lo tanto, cuanto más explícita sea la conexión, como en las experiencias iVR, más fácil se vuelve el aprendizaje. Finalmente, la Base Teórica del Cono de Experiencia de Dale (Dale, 1946) sostiene que los estudiantes aprenden mejor cuando pasan por una experiencia real o cuando la experiencia es simulada de manera realista. El entorno de aprendizaje iVR propuesto ofrece una experiencia realista en la que practicar estos principios y un entorno seguro donde corregir errores.

El segundo paso de esta metodología es la fase de diseño. La experiencia está diseñada para lograr el mayor grado de inmersión del usuario. La inmersión es la impresión subjetiva de participar en una experiencia realista e implica la suspensión voluntaria de la incredulidad. El diseño de experiencias de aprendizaje inmersivo que inducen esta incredulidad se basa en factores: 1) sensoriales, 2) orientados a la acción y 3) simbólicos (Dede, 2009). En relación con los factores sensoriales, el objetivo es reemplazar la información sensorial del mundo real con estímulos sintéticos, como imágenes visuales en 3D, sonido envolvente y respuestas táctiles (Bowman & McMahan, 2007). Relacionado con los factores de acción, la inmersión es una forma de proporcionar al participante una experiencia en la que se pueden iniciar acciones que replican las del mundo real. La experiencia está diseñada para permitir acciones intuitivas y naturales. Estas interacciones se desarrollaron con el apoyo de una plantilla creada previamente (Checa et al., 2020). Esta plantilla simplifica el proceso de desarrollo con funciones preprogramadas para su reutilización efectiva. El control remoto de una grúa puente es el principal medio de interacción entre el usuario y la aplicación. El usuario puede agarrar el controlador con cualquier mano y presionar los botones que controlan el movimiento del puente grúa con la otra mano como se muestra en el vídeo de presentación del simulador (Checa & Bustillo, 2022). Además, el usuario es capaz de moverse dentro del espacio disponible de su realidad actual, aproximadamente un espacio de 3x3 metros. Sin embargo, el usuario requería espacio adicional para realizar el ejercicio propuesto, por lo que se creó un sistema de movimiento basado en fijaciones. Se colocaron cuatro puntos de teletransportación como se muestra en la Figura 1 (II-B).

Finalmente, considerando los factores simbólicos, la activación de asociaciones semánticas y psicológicas es fundamental para la inmersión del participante en el contenido de la experiencia. Una situación real que se recrea en versión digital profundiza la experiencia inmersiva. En este caso, con el fin de fomentar estas asociaciones, el escenario, que se muestra en la Figura 1(II), fue diseñado para ser fotorrealista. Se utilizó Unreal Engine, un motor de juegos gráficos compatible con las gafas iVR (HMD, por sus siglas en inglés) seleccionadas, para la creación de esta experiencia educativa iVR.

La evaluación es la última fase del desarrollo de esta experiencia educativa iVR. En esta investigación se analizaron las habilidades de aprendices novatos para operar un puente-grúa en condiciones ideales y con factores externos que afectan el rendimiento visual o auditivo. Para ello, se crearon diferentes entornos en los que la tarea a realizar era siempre la misma, incluyendo en ciertos casos factores externos para alterar el rendimiento del aprendiz. La tarea propuesta consistía en mover el gancho de un puente-grúa hacia un barril en una posición inicial, engancharlo y completar el recorrido propuesto en el menor tiempo posible, tratando de no derribar ningún cono. En la Figura 1 (III) se muestran los diferentes espacios de la fábrica donde se realizó la tarea tanto en condiciones ideales (Figura 1.III-A), como con cronómetro y dificultades visuales y auditivas (Figura 1.III-B), con iluminación tenue (Figura1.III-C) y con alto tráfico de operarios y ruido (Figura 1.III-D). Cabe mencionar que se diseñó una experiencia corta con objetivos simples donde diferentes factores no previstos podrían ser fácilmente introducidos como perturbaciones. Con esta estrategia, los aprendices pueden probar la experiencia más de una vez

en poco tiempo, registrando diferentes niveles de experiencia a medida que aprenden rápidamente por repetición y en diferentes condiciones de aprendizaje, mientras que aumenta la cantidad de perturbaciones. Los diferentes tipos de datos, presentados en la Sección 3.2, se recopilaron automáticamente para esta evaluación.

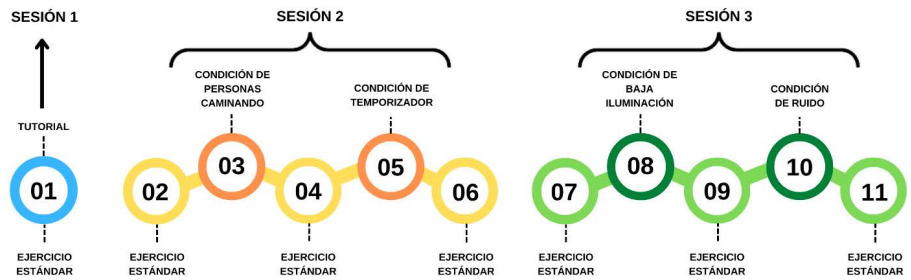
3. Experiencias de aprendizaje y conjunto de datos como método

En esta sección se describen los participantes, las experiencias de aprendizaje y los datos obtenidos de las mismas.

3.1. Experiencias de aprendizaje

Las experiencias de aprendizaje se dividieron en 3 sesiones realizadas en semanas consecutivas para la recopilación de datos. La estructura de la experiencia completa se muestra en la Figura 2.

Figura 2. Estructura de la experiencia dividida en sesiones y ejercicios



En la primera sesión (Sesión 1 en la Figura 2), los participantes realizaron un tutorial iVR para aprender a usar los controles básicos del puente-grúa y familiarizarse con el entorno iVR. A continuación, completaron el ejercicio estándar de la experiencia iVR educativa descrito en la Sección 2. En este ejercicio, los participantes debían operar el puente-grúa para que el barril fuera enganchado y transportado a través de un circuito entre conos sin que la carga cayera ni volcase cualquier cono. El ejercicio terminó cuando el aprendiz dejó la carga en el destino final del circuito. Este ejercicio estándar se repitió en los siguientes ejercicios para mejorar las habilidades de los participantes en el control del puente-grúa.

Una semana después, tuvo lugar la segunda sesión, que consistió en 5 ejercicios (Sesión 2 en la Figura 2), de los cuales el primero, el tercero y el quinto fueron ejercicios estándar. En el segundo, el aprendiz que controlaba el puente-grúa debía seguir procedimientos de seguridad cuando los operarios caminaban por la fábrica. En el cuarto ejercicio se incluyó el sonido de una campana de fábrica que podría ser estresante para el rendimiento del operario.

Finalmente, los últimos 5 ejercicios formaron la tercera sesión (Sesión 3 en la Figura 2). El ejercicio estándar se repitió en los ejercicios primero, tercero y quinto. En el segundo, las condiciones de iluminación empeoraron, lo que dificultó el manejo del puente-grúa. Finalmente, se agregaron ruidos de fondo potencialmente estresantes dentro de la fábrica durante el manejo del puente-grúa que podrían empeorar el rendimiento en el cuarto ejercicio. Además, para finalizar toda la experiencia, se invitó a todos los participantes a completar una encuesta de satisfacción. El propósito de recopilar esta información fue estudiar si los factores antes mencionados influían en los resultados de los participantes.

La muestra del experimento estaba formada por 63 estudiantes (56% mujeres), de tercer curso del Grado en Comunicación Audiovisual o de primer curso del Máster en Comunicación y Diseño Multimedia. La edad media de la muestra fue de 22,3 años (σ=2,15), y todos los participantes realizaron las tres sesiones en las mismas condiciones.

El equipo utilizado para las tres sesiones consistió en tres computadoras de escritorio equipadas con Intel Core i7-10710U, 32 GB de RAM y tarjetas gráficas NVIDIA GTX 2080 conectadas a HTC Vive Pro Eye HMD y sus controladores (Figura 1D). Todas las experiencias se realizaron siguiendo la normativa

española para evitar la transmisión de la COVID-19 y se siguió el protocolo aprobado por el Comité de Bioética de la Universidad de Burgos para la recogida de datos en cumplimiento de la protección de datos (Número de Referencia: UBU 01/2022).

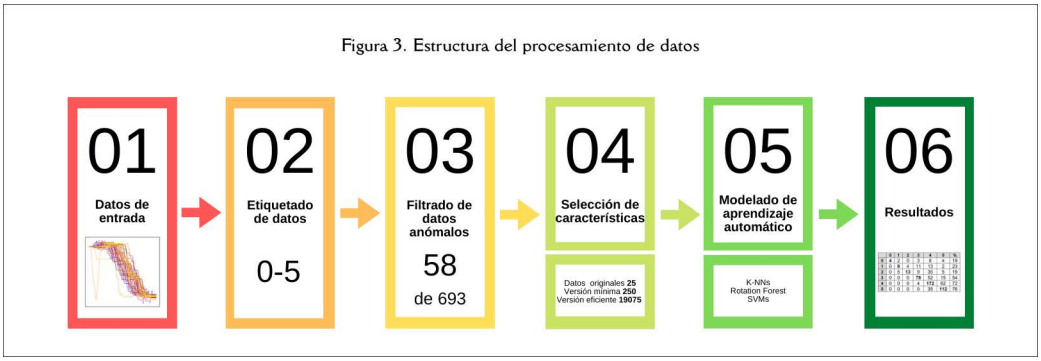
3.2. Descripción del conjunto de datos

Para recopilar la información de la experiencia descrita en la Sección 2, se creó un conjunto de datos. Incluía dos tipos de datos: 1) datos globales de cada ejercicio; y 2) datos de rendimiento del aprendiz. El conjunto de datos se resume en la Tabla 1 (http://bit.ly/3ZiA6cC). Para los datos globales, los atributos seleccionados fueron: identificador de aprendiz (ID); tiempo (T) dedicado a la tarea; fallos por colisión (F); y número de veces que se presionan dos botones simultáneamente en el control de la grúa (Pb). Los datos de rendimiento del aprendiz consistieron en 15 entradas o atributos dentro del entorno iVR relacionados con la posición y rotación de: la grúa ($Cp_{x,y,z}$ y $Cr_{y,z}$), la carga ($Lp_{x,y,z}$ y $L_{rx,y,z}$), y la cabeza del aprendiz ($Hp_{x,y,z}$ y Hr_x). Además, se extrajeron 10 entradas del sistema de eye-tracking: posición de enfoque de la mirada ($Fp_{x,y,z}$); distancia entre el aprendiz y el punto focal (D); apertura de los ojos (EL_o y ER_o); y posición de la pupila ($PLp_{x,y}$ y $PR_{px,y}$). Esas últimas 25 entradas fueron series temporales adquiridas a 120Hz. Las figuras en la columna derecha de la Tabla 1 muestran la evolución temporal de una entrada (Lp_x) para todos los aprendices para los ejercicios 2, 8 y 11, mostrando que no hay posibilidad de que el análisis de datos tradicional extraiga información inmediata de ellos. Los resultados del Número de experiencia (Xn) y el Rendimiento del aprendiz (P) también se recogen en la Tabla 1, variables que se considerarán como salidas o clases para los modelos de predicción, como se explicará en la Sección 4.1.

4. Análisis y resultados

Una vez registrados los datos de las experiencias de aprendizaje, hasta un total de 693 ejercicios, el modelado de ML se realizó en varias etapas. En primer lugar, se etiquetaron los datos. Luego, los datos se sometieron a preprocesamiento (codificación de datos, manejo de valores ausentes y detección de valores atípicos y normalización), visualización y selección de características, antes de introducirlos en los algoritmos de ML. Finalmente, se aplicaron diferentes técnicas de clasificación. La Figura 3 resume estas etapas.

Figura 3. Estructura del procesamiento de datos



4.1. Etiquetado de datos

En entornos del mundo real, la evaluación objetiva del rendimiento del que aprende puede ser difícil para determinar si la persona está lista para realizar una determinada tarea o si necesita más preparación. Sin embargo, las métricas, como el tiempo de finalización y la precisión, se pueden registrar de manera objetiva en un entorno virtual. Estas métricas se han utilizado en este estudio como medida de rendimiento del aprendiz.

La medida de rendimiento del aprendiz que se seleccionó y etiquetó como un valor entero entre 0 y 5, se basó en dos parámetros: 1) tiempo de finalización; y 2) fallos por colisión. Los fallos se calcularon a partir de la cantidad de conos derribados durante el ejercicio y la cantidad de veces que se presionaron varios botones del controlador físico al mismo tiempo. Ambos parámetros fueron calificados de 0 a 5, y el mínimo de ambos fue asignado como la métrica de desempeño final. En cuanto a las etiquetas utilizadas

para el contexto de aprendizaje, corresponden a las utilizadas para identificar cada ejercicio, descritas en la Figura 2.

Cada ejercicio explicado en la Sección 3.1 para cada aprendiz fue considerado una muestra única, por lo que se asignó una evaluación de rendimiento por cada aprendiz y ejercicio realizado. Cada participante realizó 11 sesiones y 63 aprendices participaron en el experimento, por lo que el conjunto de datos original estaba compuesto por 693 muestras. La distribución de cada etiqueta de conjunto de datos en la muestra fue la siguiente: 3% etiqueta 0; 5% etiqueta 1; 12% etiqueta 2; 26% etiqueta 3; 35% etiqueta 4 y 19% etiqueta 5. Se observó un claro desequilibrio, especialmente en las clases 0 y 1, resultado natural si se tiene en cuenta que los aprendices aprenden rápidamente la tarea propuesta, mejorando su puntuación después de las dos primeras sesiones.

4.2. Preparación de datos y extracción de características

Durante la etapa de captura de datos, pueden ocurrir errores que son difíciles de detectar mientras se realiza la experiencia, y es crucial filtrarlos para que no introduzcan ruido en el conjunto de datos. Los fallos de captura pueden venir de fallos en el software debidos a la saturación del búfer, del fallo momentáneo del sensor e incluso por circunstancias como reflejos o desalineación del HMDs. Estos fallos pueden ser detectados mediante la visualización de datos para luego ser filtrados.

Por lo tanto, se realizó un preprocesamiento de datos para filtrar datos anómalos antes de las tareas de ML. Para esta tarea se seleccionaron varias bibliotecas ampliamente utilizadas en el campo de la ciencia de datos. Por un lado, para la visualización de datos se seleccionó la biblioteca Pandas (McKinney, 2011) y el paquete tslearn (Tavenard et al., 2020), debido a su especial diseño para análisis de series temporales. Como resultado, se eliminaron 58 muestras del conjunto de datos original de 693 porque mostraban anomalías elevadas o comportamientos de aprendiz inusuales. Por otro lado, las muestras capturadas eran de diferente duración, ya que cada alumno completó los ejercicios en un tiempo diferente. La investigación sobre la clasificación de series temporales suele centrarse en el caso de series de longitud uniforme. Como este trabajo pretende proporcionar una línea de base para futuras investigaciones, las series temporales se normalizaron, en este caso a la longitud máxima (4326 puntos de datos), haciendo uso de la función TimeseriesResample de la biblioteca tslearn que realiza una interpolación lineal, de manera que se consiguió que todas tuvieran la misma longitud.

En segundo lugar, se realizó la selección de características de los datos sin procesar. El objetivo de esta tarea era explorar la cantidad de información útil oculta en cada conjunto de datos. Para ello se utilizó el algoritmo FRESH (Christ et al., 2016) del paquete tsfresh (Christ et al., 2018). Esta biblioteca incluye una amplia variedad de características que pueden ser extraídos de series temporales sin procesar. En este caso, se extrajeron 19.075 características para cada serie temporal. El conjunto de características puede incluir atributos estadísticos básicos (picos, máximos, mínimos, etc.), medidas de corrección y evolución de una serie temporal (ruido blanco, tendencia, estacionalidad, autocorrelación, etc.). Existen algunos diccionarios predefinidos de la biblioteca, dos de los cuáles se usaron en este estudio. Una versión más ligera, denominada «mínima», y otra más completa, denominada "eficiente". La extracción de características se realizó en ambos modos, obteniendo así dos nuevos conjuntos de datos: uno con la versión mínima y otro con todas las características (versión eficiente).

4.3. Aprendizaje y proceso de modelado

A continuación, se utilizaron diferentes algoritmos de ML para predecir el rendimiento del aprendiz y la calidad del entorno de aprendizaje. Se probaron los tres conjuntos de datos diferentes propuestos en la Sección 4.2: 1) los datos originales sin procesar, 2) la versión mínima de extracción de características y 3) la versión eficiente.

Para esta tarea se probaron tres técnicas de ML, cada una de naturaleza muy diferente: 1) k-vecinos más cercanos (kNN), un algoritmo de agrupamiento simple pero eficiente que usa la proximidad para hacer clasificaciones o predicciones sobre la agrupación de un punto de datos individual. El valor de k define cuántos vecinos se verificarán (en este caso, k se estableció en 1). 2) Máquinas de vectores de soporte (SVM), un algoritmo complejo y bien consolidado que define un hiperplano en un espacio N-

dimensional, con N como número de características, que separan claramente los puntos de datos dados. Y 3) Random Forest (RF), un algoritmo que pertenece a la familia de algoritmos ensemble. Crea árboles de decisión y combina los resultados de todos ellos utilizando la votación mayoritaria para la clasificación y los promedios para la regresión. El objetivo era evaluar cuál predice mejor el rendimiento y el conjunto de datos más adecuado para esa tarea de clasificación. Los tres algoritmos se evaluaron utilizando la herramienta WEKA (Hall et al., 2008).

Se usó una validación cruzada, dada la invariancia estadística para la selección de los subconjuntos, para dividir el conjunto de datos en instancias de entrenamiento e instancias de validación. Se seleccionó una validación cruzada con el 10% de los datos, debido al tamaño del conjunto de datos. El indicador de calidad seleccionado fue la precisión, que representa la proporción de observaciones correctamente clasificadas sobre el número total de instancias evaluadas.

4.4. Resultados

En la Tabla 2 se muestran los resultados obtenidos para cada uno de los experimentos mencionados. Los mejores resultados se destacan en negrita. El conjunto de datos de la versión mínima obtuvo mejores resultados que los otros dos conjuntos de datos, lo que demuestra la necesidad de la selección de características en conjuntos de datos de este tipo. En cuanto a los algoritmos, Random Forest fue el que claramente se desempeñó mejor en ambas tareas.

Tabla 2. Resultados de clasificación de la calidad del entorno de aprendizaje y el rendimiento del aprendiz

	Ejercicios (% de precisión)			Rendimiento del aprendiz (% de precisión)		
Algoritmo / Datos	RF	1-NN	SVM	RF	1-NN	SVM
Datos originales	**43,56**	26,17	35,41	**42,38**	35,24	40,74
Version mínima	**44,29**	31,12	42,34	**59,31**	48,20	51,98
Versión eficiente	**40,11**	25,84	40,72	**59,12**	48,05	51,72

Conviene señalar algunas cuestiones. En primer lugar, el bajo rendimiento de kNN demuestra que los algoritmos usados en la búsqueda de tareas previas que presenten gran similitud con la tarea que se quiere predecir no son adecuados para este tipo de tareas. Se trata de un desafío fascinante en que las técnicas de ML, especialmente diseñadas para estructuras de datos complejas, jugarán un papel central.

Tabla 3. Matriz de confusión para la clasificación de experiencias (izquierda) y rendimiento (derecha) - modelo RF

	1	2	3*	4	5*	6	7	8*	9	10*	11	%
1	**21**	11	0	3	0	0	4	0	0	0	0	54
2	12	**26**	5	2	4	3	7	0	0	0	0	44
3*	6	9	**39**	0	5	3	4	0	5	0	0	55
4	3	11	9	**10**	3	7	5	0	9	0	13	14
5*	0	0	9	0	**43**	0	0	0	0	0	0	83
6	3	7	2	15	10	**4**	3	0	7	0	17	6
7	5	8	7	5	3	2	**15**	0	10	0	5	25
8*	0	0	0	0	0	0	0	**65**	0	0	0	100
9	0	2	0	9	2	9	8	0	**7**	0	21	12
10*	0	0	4	0	5	2	3	0	2	**46**	0	74
11	0	2	0	7	2	9	11	0	7	0	**15**	28

	0	1	2	3	4	5	%
0	**4**	2	0	3	8	4	19
1	0	**9**	4	11	13	2	23
2	0	5	**13**	9	35	5	19
3	0	0	0	**78**	52	15	54
4	0	0	0	4	**172**	62	72
5	0	0	0	0	35	**112**	76

En segundo lugar, todas las técnicas de extracción de características y los algoritmos ML que se probaron proporcionaron un rendimiento de predicción medio-bajo, que raramente era muy preciso, lo que revela una futura línea de investigación para mejorar. Finalmente, en la Tabla 2 se muestran los valores de rendimiento promedio, sin mostrar el rendimiento de todas las clases (niveles de rendimiento o ejercicios). Para analizar esta cuestión en detalle, en la Tabla 3 se muestra la matriz de confusión para el mejor método, Random Forest, y para ambas tareas de clasificación. Se incluye el porcentaje de instancias predichas correctamente a la derecha de cada matriz de confusión.

La matriz de confusión para la clasificación de las experiencias (a la izquierda de la Tabla 3) mostró que las experiencias con algún tipo de limitación de aprendizaje (ruido, presión de tiempo…) alcanzaron altos niveles de precisión (78% de media frente al 26% de los ejercicios estándar); esos fueron los ejercicios 3, 5, 8 y 10, marcados con un asterisco en la Tabla 3. En cuanto al desempeño de los aprendices, aunque el modelo no obtuvo la clasificación correcta, mostró una tendencia a predecir clases cercanas a las correctas; por lo tanto, el sistema fue capaz de clasificar correctamente a los aprendices novatos y a los aprendices expertos. Las clasificaciones de los modelos fueron significativamente mejores para las clases 3, 4 y 5 (rendimiento medio-alto bueno) que para las clases 0, 1 y 2 (rendimiento bajo). Un resultado que también era previsible, dado el desequilibrio de las clases descrito en la Sección 4.1.

5. Discusión y conclusiones

Los sistemas iVR actuales generalmente utilizan métodos de aprendizaje estandarizados que no se adaptan a las características individuales de cada alumno. Esto conduce a altos niveles de desmotivación, actitudes pasivas, aburrimiento, bajo compromiso y frustración entre los alumnos. Los datos de eye-tracking pueden desempeñar un papel importante en la supervisión de estos entornos y como complemento de otras herramientas de recopilación de datos, por ejemplo, la conductimetría. El uso de técnicas de Inteligencia Artificial sobre esos conjuntos de datos extraídos de entornos de entrenamiento iVR puede ser la solución deseada, para adaptar los entornos de aprendizaje iVR a los diferentes antecedentes y características particulares de cada alumno. En este estudio se ha examinado la forma en que se pueden aplicar las técnicas básicas de ML para lograr ese objetivo, específicamente para evaluar las condiciones de aprendizaje y el rendimiento del aprendiz, dentro de áreas donde la bibliografía existente es especialmente limitada. Para ello se ha diseñado un entorno iVR y una experiencia de prueba, de forma que se esperaba que los aprendices repitieran una tarea corta y sencilla mientras eran expuestos a diferentes perturbaciones, aprendiendo rápidamente y generando un conjunto de datos con una gran diversidad de ejercicios para diferentes niveles de experiencia del aprendiz y bajo diferentes condiciones ambientales. Luego se probaron diferentes técnicas de ML para dos tareas: 1) clasificación de la calidad del entorno de aprendizaje; y 2) predicción del rendimiento del aprendiz. Se siguieron métodos de ciencia de datos bien establecidos para probar las siguientes técnicas: etiquetado de datos, filtrado de datos, extracción de características y modelado de ML bajo un esquema de validación cruzada. Entre los algoritmos que se probaron, Random Forest mostró la mejor precisión para ambas tareas. Si bien se logró una alta precisión para clasificar las condiciones de aprendizaje anómalas (78%), los resultados no fueron tan buenos para la predicción del rendimiento del aprendiz (59%). Cabe señalar que el objetivo de esta investigación no es encontrar una solución confiable y robusta para estas tareas, sino que se trata de una primera aproximación que proporcionará una línea base para futuras mejoras en el uso de ML en entornos de entrenamiento con iVR.

En comparación con la bibliografía existente, en este estudio se alcanzan niveles de precisión similares en la evaluación de la calidad del entorno. Así, mientras que en este estudio (Wismer et al., 2022) se consiguió predecir el estado de experto o novato del usuario con una precisión del 77% en una simulación de iVR para medir las habilidades de laboratorio y la evaluación y el cumplimiento del alumno mediante la conductimetría, en el presente estudio se lograron niveles de precisión del 78% al clasificar la calidad del entorno de aprendizaje. En comparación con la evaluación de atención o distracción (Asish et al., 2022), la precisión del modelo fue menor; diferencia que surge de la definición de clases en ambos trabajos: mientras que en esta investigación se utilizaron hasta 6 niveles, en Asish et al. (2022) utilizaron una clasificación binaria, que generalmente produce una mayor precisión. Finalmente, en comparación con la clasificación de calidad en la conducción (Deng et al., 2020), en este trabajo se han alcanzado algunas conclusiones comunes: la estabilidad y mayor precisión de las técnicas de ensembles, como Random Forest, sobre otros algoritmos clásicos, como kNN, o SVM. Una vez más, la alta precisión lograda en este trabajo (hasta el 89%) podría provenir de la selección de solo 3 clases y la gran diferencia de comportamiento entre los conductores en cada clase. Como también se describió en dichos trabajos anteriores, se requiere la extensión de los conjuntos de datos, en términos de aprendices y condiciones, para lograr una mayor precisión. Sin embargo, la idoneidad del ML para el desempeño de tales tareas ha sido confirmada

en esta investigación, en la medida en que uno de los conjuntos de datos más grandes, con más de 50 millones de puntos de datos, se procesó de manera mucho más eficiente que las técnicas convencionales de procesamiento de datos basadas en humanos.

Los estudios futuros podrían centrarse en mejorar la precisión de los modelos de predicción para la evaluación del aprendizaje en entornos iVR. Este objetivo podría lograrse ampliando el conjunto de datos para incluir experiencias de nuevos aprendices, mejorando las metodologías de etiquetado y utilizando técnicas de equilibrio para clases altamente desequilibradas (como el algoritmo SMOTE). Además, se podrían probar técnicas alternativas de ML, como los modelos ocultos de Markov con resultados probados para series temporales, a fin de capturar las tendencias dinámicas del rendimiento del aprendiz. Además, los resultados han motivado la necesidad de agregar información relacionada con la sesión al conjunto de datos, de modo que se puedan extraer los patrones de rendimiento de los aprendices durante y entre sesiones.

Contribución de Autores

Idea, C.P.A.; Revisión de literatura (estado del arte), D.C.; Metodología, I.M.A., D.C.; Análisis de datos, A.S.M., C.P.A.; Resultados, A.S.M.; Discusión y conclusiones, A.S.M., I.M.A.; Redacción (borrador original), C.P.A., I.M.A.; Revisiones finales, C.P.A., I.M.A.; Diseño del Proyecto y patrocinios, D.C., A.S.M.

Apoyos

Esta investigación ha sido financiada por el Proyecto ACIS (INVESTUN/21/BU/0002) de la Consejería de Empleo e Industria de la Junta de Castilla y León (España), el Proyecto Erasmus+ RISKREAL (2020-1-ES01-KA204-081847) de la Comisión Europea, el Proyecto HumanAid (TED2021-129485B-C43) del Ministerio Español de Ciencia e Innovación y el Programa Margarita Salas del Ministerio Español de Universidades financiado por NextGenerationEU.

Referencias

Añaños-Carrasco, E. (2015). Eyetracker technology in elderly people: How integrated television content is paid attention to and processed. [La tecnología del «EyeTracker» en adultos mayores: Cómo se atienden y procesan los contenidos integrados de televisión]. *Comunicar*, *45*, 75-83. https://doi.org/10.3916/C45-2015-08

Asish, S.M., Kulshreshth, A.K., & Borst, C.W. (2022). Detecting distracted students in educational VR environments using machine learning on eye gaze data. *Computers & Graphics*, *109*, 75-87. https://doi.org/10.1016/j.cag.2022.10.007

Bowman, D.A., & Mcmahan, R.P. (2007). Virtual reality: How much immersion is enough? *Computer*, *40*(7), 36-43. https://doi.org/10.1109/MC.2007.257

Checa, D., & Bustillo, A. (2020). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, *79*, 5501-5527. https://doi.org/10.1007/s11042-019-08348-9

Checa, D., & Bustillo, A. (2022). *Grua Rv*. http://3dubu.Es/En/Cranevr/

Checa, D., Gatto, C., Cisternino, D., De Paolis, L.T., & Bustillo, A. (2020). A Framework for Educational and Training Immersive Virtual Reality Experiences. In L. T. de Paolis, & P. Bourdot (Eds.), *Augmented reality, virtual reality, and computer graphics* (pp. 220-228). Springer International Publishing. https://doi.org/10.1007/978-3-030-58468-9_17

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A.W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh - A Python package). *Neurocomputing*, *307*, 72-77. https://doi.org/10.1016/j.neucom.2018.03.067

Christ, M., Kempa-Liehr, A., & Feindt, M. (2016). Distributed and parallel time series feature extraction for industrial big data applications. *ArXiv*, *1*. https://doi.org/10.48550/arXiv.1610.07717

Cowan, A., Chen, J., Mingo, S., Reddy, S.S., Ma, R., Marshall, S., Nguyen, J.H., & Hung, A.J. (2021). virtual reality vs dry laboratory models: Comparing automated performance metrics and cognitive workload during robotic simulation training. *Journal of Endourology*, *35*(10), 1571-1576. https://doi.org/10.1089/end.2020.1037

Dale, E. (1946). *Audiovisual methods in teaching*. Dryden Press. https://bit.ly/42aW03X

Dalgarno, B., & Lee, M.J.W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, (1), 41-41. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, *323*(5910), 66-69. https://doi.org/10.1126/science.1167311

Deng, Q., Wang, J., Hillebrand, K., Benjamin, C.R., & Soffker, D. (2020). Prediction performance of lane changing behaviors: A study of combining environmental and eye-tracking data in a driving simulator. *IEEE Transactions on Intelligent Transportation Systems*, *21*(8), 3561-3570. https://doi.org/10.1109/TITS.2019.2937287

Duchowski, A.T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 455-470. https://doi.org/10.3758/BF03195475

Farran, E., Formby, S., Daniyal, F., Holmes, T., & Herwegen, J. (2016). Route-learning strategies in typical and atypical development; eye-tracking reveals atypical landmark selection in Williams syndrome: Route-learning and eye-tracking. *Journal of Intellectual Disability Research*, *60*(10), 933-944. https://doi.org/10.1111/jir.12331

García-Carrasco, J., Hernández-Serrano, M.J., & Martín-García, A.V. (2015). Plasticity as a framing concept enabling transdisciplinary understanding and research in neuroscience and education. *Learning, Media and Technology*, *40*, 152-167.

https://doi.org/10.1080/17439884.2014.908907
Gardony, A.L., Lindeman, R.W., & Brunyé, T.T. (2020). Eye-tracking for human-centered mixed reality: Promises and challenges. *Proc.SPIE*, *11310*, 113100T. https://doi.org/10.1117/12.2542699
Glennon, J.M., Souza, H., Mason, L., Karmiloff-Smith, A., & Thomas, M.S.C. (2020). Visuo-attentional correlates of Autism Spectrum Disorder (ASD) in children with Down syndrome: A comparative study with children with idiopathic ASD. *Research in Developmental Disabilities*, *104*, 103678. https://doi.org/10.1016/j.ridd.2020.103678
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. (2008). The WEKA data mining software: An update. *SIGKDD Explor. Newsl*, *11*(1), 10-18. https://doi.org/10.1145/1656274.1656278
Huang, H.M., Rauch, U., & Liaw, S.S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers and Education*, *55*(3), 1171-1182. https://doi.org/10.1016/j.compedu.2010.05.014
Lapborisuth, P., Koorathota, S., Wang, Q., & Sajda, P. (2021). Integrating neural and ocular attention reorienting signals in virtual reality. *Journal of Neural Engineering*, *18*(6). https://doi.org/10.1088/1741-2552/ac4593
Ma, X., Yao, Z., Wang, Y., Pei, W., & Chen, H. (2018). Combining brain-computer interface and eye-tracking for high-speed text entry in virtual reality. In *IUI '18: 23rd International Conference on Intelligent User Interfaces* (pp. 263-267). https://doi.org/10.1145/3172944.3172988
Martinez, K., Menéndez-Menéndez, M.I., & Bustillo, A. (2021). Awareness, prevention, detection, and therapy applications for depression and anxiety in serious games for children and adolescents: Systematic review. *JMIR Serious Games*, *9*(4). https://doi.org/10.2196/30482
Mckinney, W. (2011). *pandas: A foundational Python library for data analysis and statistics*. Python High Performance Science Computer.
Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., & Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans. Graph*, *35*(6), 1-12. https://doi.org/10.1145/2980179.2980246
Pritchard, A. (2017). *Ways of learning: Learning theories for the classroom*. Routledge. https://doi.org/10.4324/9781315460611
Rappa, N.A., Ledger, S., Teo, T., Wong, K.W., Power, B., & Hilliard, B. (2022). The use of eye-tracking technology to explore learning and performance within virtual reality and mixed reality settings: A scoping review. *Interactive Learning Environments*, *30*(7), 1338-1350. https://doi.org/10.1080/10494820.2019.1702560
Rodero, E., & Larrea, O. (2022). Virtual reality with distractors to overcome public speaking anxiety in university students; [Realidad virtual con distractores para superar el miedo a hablar en público en universitarios]. *Comunicar*, *72*. https://doi.org/10.3916/C72-2022-07
Shadiev, R., & Li, D. (2022). A review study on eye-tracking technology usage in immersive virtual reality learning environments. *Computers & Education*. https://doi.org/10.1016/j.compedu.2022.104681
Sun, Q., Patney, A., Wei, L.Y., Shapira, O., Lu, J., Asente, P., Zhu, S., Mcguire, M., Luebke, D., & Kaufman, A. (2018). Towards virtual reality infinite walking: Dynamic saccadic redirection. *ACM Transactions on Graphics*, *37*(4), 1-13. https://doi.org/10.1145/3197517.3201294
Tanaka, Y., Kanari, K., & Sato, M. (2021). Interaction with virtual objects through eye-tracking. In *International Workshop on Advanced Imaging Technology (IWAIT) 2021* (pp. 1176624). SPIE. https://doi.org/10.1117/12.2590989
Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K., & Woods, E. (2020). Tslearn, A Machine-learning Toolkit for Time Series Data. *J. Mach. Learn. Res*, *21*, 1-6.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*, 625-636. https://doi.org/10.3758/BF03196322
Wismer, P., Soares, S.A., Einarson, K.A., & Sommer, M.O.A. (2022). Laboratory performance prediction using virtual reality behaviometrics. *PloS One*, *17*(12). https://doi.org/10.1371/journal.pone.0279320